Zero-Knowledge Against Quantum Attacks

John Watrous

Institute for Quantum Computing and School of Computer Science
University of Waterloo, Waterloo, Ontario, Canada

April 24, 2008
(with minor corrections on April 14, 2009)

Abstract

It is proved that several interactive proof systems are zero-knowledge against general quantum attacks. This includes the Goldreich–Micali–Wigderson classical zero-knowledge protocols for Graph Isomorphism and Graph 3-Coloring (assuming the existence of quantum computationally concealing commitment schemes in the second case). Also included is a quantum interactive proof system for a complete problem for the complexity class of problems having honest verifier quantum statistical zero-knowledge proofs, which therefore establishes that honest verifier and general quantum statistical zero-knowledge are equal: $\mathrm{QSZK} = \mathrm{QSZK}_{\mathrm{HV}}$. Previously no non-trivial interactive proof systems were known to be zero-knowledge against quantum attacks, except in restricted settings such as the honest-verifier and common reference string models. This paper therefore establishes for the first time that true zero-knowledge is indeed possible in the presence of quantum information and computation.

1 Introduction

The security of classical cryptographic systems against quantum computer attacks has the potential to become an issue of critical importance in cryptography in years to come. It is well-known, for instance, that Shor's algorithm [Sho97] allows for efficient quantum cryptanalysis of many public-key cryptosystems such as RSA [RSA78]. In the event that large-scale quantum computing becomes technologically feasible, these cryptosystems will therefore be rendered completely insecure. While quantum cryptography offers a secure alternative in the case of key-exchange [BB84, May01, SP00], it is not reasonable to expect that quantum cryptographic systems will become widely used in the near future—even the relatively low technological requirements of quantum key exchange are presently well beyond the means of most computer users. A more practical solution is to design classical cryptosystems that are secure against quantum computer attacks.

This paper investigates the security of zero-knowledge interactive proof systems against quantum computer attacks. Although both quantum and classical interactive proof systems are considered, the main focus of this paper is on the security of classical zero-knowledge proof systems against quantum attacks. This is the case of greatest practical importance, for the reasons suggested above. Independent of present-day progress toward large-scale quantum computing, this study is motivated by an obvious goal in cryptography: to prove security of cryptosystems against as wide a range of malicious attacks as possible, while at the same time minimizing the resource requirements of honest users.

The notion of zero-knowledge was introduced by Goldwasser, Micali and Rackoff [GMR89]. Informally speaking, an interactive proof system has the property of being zero-knowledge if arbitrary verifiers that interact with the honest prover of the system learn nothing from the interaction beyond the validity of the statement being proved. At first consideration this notion may seem to be paradoxical, but indeed several interesting computational problems that are not known to be polynomial-time computable admit zero-knowledge interactive proof systems in the classical setting. Examples include the Graph Isomorphism [GMW91] and Quadratic Residuosity [GMR89] problems, various lattice problems [GG00], and the Statistical Difference [SV03] and Entropy Difference [GV99] problems, which concern outputs of Boolean circuits with random inputs. The fact that the last three examples have interactive proof systems that are zero-knowledge relies on a fundamental result of Goldreich, Sahai and Vadhan [GSV98] equating zero-knowledge with *honest verifier* zero-knowledge in some settings. Under certain intractability assumptions, every language in NP has a zero-knowledge interactive proof system [GMW91]. A related notion is that of *interactive arguments*, which differ from interactive proof systems in that computational restrictions are placed on the prover as well as the verifier [BCC88]. In the interactive argument setting, zero-knowledge protocols exist that have somewhat different characteristics than protocols in the interactive proof system setting.

There are several variants of zero-knowledge that differ in the way that the notion of "learning nothing" is formalized. In each variant, it is viewed that a particular verifier learns nothing if there exists a polynomial-time *simulator* whose output is indistinguishable from the output of the verifier when the honest prover and verifier interact on any positive instance of the problem. The different variants concern the strength of this indistinguishability. In particular, *perfect* and *statistical* zero-knowledge refer to the situation where the simulator's output and the verifier's output are indistinguishable in an information-theoretic sense and *computational* zero-knowledge refers to the weaker restriction that the simulator's output and the verifier's output cannot be distinguished by any computationally efficient procedure.

The security of zero-knowledge proof systems against *quantum* verifiers has been a topic of study for several years [Gra97, Wat02, Kob03, DFS04]. Progress has been slow, however, for a simple reason—while it is straightforward to formulate natural quantum analogues of the classical definitions of zero-knowledge, no interactive proof systems or interactive arguments could be shown to satisfy these definitions prior to this paper. This has left open several possibilities, including the possibility that any "correct" definition of quantum zero-knowledge would necessarily be qualitatively different from the usual classical definitions, as well as the possibility that zero-knowledge is simply impossible in a quantum world. The fundamental difficulty that one encounters when trying to apply these quantum definitions was first discovered by van de Graaf [Gra97], and will be discussed shortly.

The main task involved in proving that a given interactive proof system is zero-knowledge is the construction of a simulator for every possible deviant polynomial-time verifier. The most typical method for doing this involves the simulator treating a given verifier as a *black box*: the simulator randomly generates transcripts, or parts of transcripts, of possible interactions between a prover and verifier, and feeds parts of these transcripts to the given verifier. If the verifier produces a message that is not consistent with the other parts of the transcript that were generated, the simulator *rewinds*, or backs up and tries again to randomly generate parts of the transcript. By storing intermediate results, and repeating different parts of this process until the given verifier's output is consistent with a randomly generated transcript, the simulation is eventually successful.

In the quantum setting, the rewinding technique faces two basic obstacles that are unique to quantum information. The first is the *no-cloning theorem* [WZ82], which states that unknown quantum states cannot be copied, and the second is the principle of *information gain versus state dis-*

turbance (see Fuchs and Peres [FP96], for instance). Quantum verifiers may store and manipulate quantum information, and may even begin a given protocol with *auxiliary* quantum information—perhaps resulting from a previous iteration of the protocol being considered. A simulator cannot make copies of the verifier's quantum state at any given time, and must therefore treat the quantum information stored by the verifier very carefully. This makes rewinding problematic, because if the simulation of given verifier is inconsistent with an actual interaction at some point, it is not obvious how to rewind the process and try again: measurements made by the simulator, such as ones providing information about the success of the simulation, will generally cause a disturbance in the quantum information it stores.

Other methods of constructing simulators for quantum verifiers have also been considered, in an attempt to circumvent the problematic rewinding issue. For example, Damgård, Fehr, and Salvail [DFS04] proved several interesting results concerning quantum zero-knowledge protocols in the *common reference string* model, wherein it is assumed that an honest third party samples a string from some specified distribution and provides both the prover and verifier with this string at the start of the interaction. Their results, based on what they call the *no quantum rewinding paradigm*, are mostly concerned with interactive arguments and rely on certain quantum complexity-theoretic intractability assumptions. Another weaker notion of zero-knowledge is *honest verifier* zero-knowledge, which only requires that a simulator outputs the honest verifier's view of an interaction with the honest prover P. A quantum variant of honest verifier statistical zero-knowledge was considered in [Wat02], wherein it was proved that the resulting complexity class $\mathrm{QSZK}_{\mathrm{HV}}$ shares many of the basic properties of its classical counterpart [SV03]. A non-interactive variant of this notion was studied by Kobayashi [Kob03]. The problematic issue regarding simulator constructions does not occur in honest verifier settings.

The present paper resolves, at least to a significant extent, the main difficulties previously associated with quantum analogues of zero-knowledge. This is done by establishing that the most natural quantum analogues of the classical definitions of zero-knowledge indeed can be applied to a large class of interactive proof systems. This includes some well-known classical interactive proof systems as well as quantum interactive proof systems for several problems, in particular the class of all problems admitting quantum interactive proof systems that are statistical zero-knowledge against honest verifiers. It is therefore proved unconditionally that zero-knowledge indeed is possible in the presence of quantum information and computation, and moreover that the notion of quantum zero-knowledge is correctly captured by the most natural and direct quantum analogues of the classical definitions.

The main technique that is introduced in this paper is algorithmic in nature: it is shown how to construct efficient quantum simulators for arbitrary quantum polynomial-time verifiers for several interactive proof systems. These simulators rely on a general *Quantum Rewinding Lemma* that establishes simple conditions under which the success probabilities of certain processes with quantum inputs and outputs can be amplified.

The remainder of this paper is organized as follows. Section 2 gives a summary of key concepts needed throughout the paper, including interactive proof systems, zero-knowledge, and various quantum information-theoretic notions; Section 3 states the definitions of quantum statistical and computational zero-knowledge that are studied in the sections that follow; and Section 4 contains a statement and proof of the Quantum Rewinding Lemma, which encapsulates the main technical tool that is used to prove the security of various zero-knowledge interactive proof systems against quantum attacks. The remaining sections are Section 5, which discusses quantum statistical zero-knowledge protocols, and Section 6, which discusses quantum computational zero-knowledge protocols. The paper concludes with Section 7.

2 Preliminaries

This section is intended to summarize some of the notation, conventions, and known facts concerning interactive proof systems, zero-knowledge, and quantum information and computation that are used throughout the paper. The reader is assumed to be familiar with basic computational complexity as well as quantum information and computation. These topics are independently covered in several books [AB06, Pap94, NC00, KSV02], to which the reader is referred for further details. Most of the complexity-theoretic facts proved in this paper are expressed in terms of *promise problems*, which are discussed in Refs. [ESY84, Gol05]. Further information on interactive proof systems and zero-knowledge can be found in [Gol01], and information on quantum variants of interactive proof systems can be found in [KW00].

For the rest of the paper let us fix the alphabet $\Sigma = \{0,1\}$, and only consider strings, promise problems, and complexity classes over this alphabet. When we say that a function g having the form $g : \{0,1,2,\ldots\} \to \{1,2,3,\ldots\}$ is a *polynomially bounded function*, we mean that there exists a deterministic Turing machine M_g such that (i) on input 1^n, M_g outputs $1^{g(n)}$, for every non-negative integer n, and (ii) the running time of M_g is bounded by some polynomial p. A function having the form $f : \{0,1,2,\ldots\} \to (0,\infty)$ is said to be *negligible* if, for every polynomially bounded function g, it holds that $f(n) < 1/g(n)$ for all but finitely many values of n.

2.1 Interactive proof systems

Interactive proof systems will be specified by pairs (V,P) representing an honest verifier and honest prover. The soundness property of such an interactive proof system concerns interactions between pairs (V,P') and the zero-knowledge property concerns interactions between pairs (V',P), where P' and V' deviate arbitrarily from P and V, respectively. It may be the case that a given prover/verifier pair is such that both are classical, both are quantum, or one is classical and the other is quantum. When either or both of the parties is classical, all communication between them is (naturally) assumed to be classical—only two quantum parties are permitted to transmit quantum information to one another. It will always be assumed that verifiers are represented by polynomial-time (quantum or classical) computations. Depending on the setting of interest, the honest prover P may either be computationally unrestricted or may be represented by a polynomial-time (quantum or classical) computation augmented by specific information about the input string, such as a witness for an NP problem. Deviant provers will be assumed to be computationally unrestricted. (The results of this paper are applicable to interactive arguments but none are specific to them, and so for simplicity they are not considered further.)

For a given promise problem $A = (A_{\mathrm{yes}}, A_{\mathrm{no}})$, we say that a pair (V,P) is an interactive proof system for A having completeness error ε and soundness error δ if (i) for every input $x \in A_{\mathrm{yes}}$, the interaction between P and V causes V to accept with probability at least $1-\varepsilon$, and (ii) for every input $x \in A_{\mathrm{no}}$ and every prover P', the interaction between P' and V causes V to accept with probability at most δ. It may be the case that ε and δ are constant or are functions of the length of the input string x. When they are functions, it is assumed that they can be computed deterministically in polynomial time. It is generally desired that ε and δ are exponentially small; but because sequential repetition followed by majority vote, or unanimous vote in case $\varepsilon = 0$, reduces these errors exponentially quickly, it is usually sufficient that $1-\varepsilon-\delta$ is lower-bounded by the reciprocal of a polynomial. A similar statement holds for parallel repetition, but the zero-knowledge property to be discussed shortly will generally be lost in this case.

2.2 Classical zero-knowledge

There are different notions of what it means for an interactive proof system (V, P) for a promise problem A to be zero-knowledge. At this point we will just discuss the completely classical case, meaning that only classical verifiers are considered.

An arbitrary verifier V' takes two strings as input: a string x representing the common input to both the verifier and prover, as well as a string w called an *auxiliary input*, which is not known to the prover and which may influence the verifier's behavior during the interaction. Based on the interaction with P, the verifier V' produces a string as output. For a given input string x, let $n = q(|x|)$ denote the length of the auxiliary input string and let $m = r(|x|)$ denote the length of the output string, for polynomially bounded functions q and r. Because there may be randomness used by either or both of P and V', the verifier's output will in general be random. The (string-valued) random variable representing the verifier's output will be written $(V'(w), P)(x)$. For the honest verifier V, we may view that $n = 0$ and $m = 1$ for all $x \in \Sigma^*$, because there is no auxiliary input and the output is a single bit that indicates whether the verifier accepts or rejects. By a (classical) simulator for a verifier V', we mean a polynomial-time randomized algorithm $S_{V'}$ that takes the strings w and x as input and produces some output string of length m. Such a simulator's output is a random variable denoted $S_{V'}(w, x)$. Note that the simulator does not interact with the prover.

Now, for a given promise problem A, we say that an interactive proof system (V, P) for A is *zero-knowledge* if, for every verifier V', there exists a simulator $S_{V'}$ such that $(V'(w), P)(x)$ and $S_{V'}(w, x)$ are indistinguishable for every choice of strings $x \in A_{\text{yes}}$ and $w \in \Sigma^n$. The specific formalization of the word "indistinguishable" gives rise to different variants of zero-knowledge. *Statistical* zero-knowledge refers to the situation in which $(V(w), P)(x)$ and $S_{V'}(w, x)$ have negligible statistical difference, and *computational* zero-knowledge refers to the situation in which no Boolean circuit with size polynomial in $|x|$ can distinguish $(V'(w), P)(x)$ and $S_{V'}(w, x)$ with a non-negligible advantage over randomly guessing. *Perfect* zero-knowledge is slightly stronger than statistical zero-knowledge in that it essentially requires a zero-error simulation: the simulator may report failure with some small probability, but conditioned on the simulator not reporting failure the output $S_{V'}(w, x)$ of the simulator is distributed identically to $(V'(w), P)(x)$. We will only consider statistical and computational zero-knowledge once we move to the quantum case.

Two points concerning the definitions just discussed should be mentioned. The first point concerns the auxiliary input, which actually was not included in the definitions given in the very first papers on zero-knowledge (but which already appeared in the 1989 journal version of [GMR89]). The inclusion of an auxiliary input in the definition is critical: it is necessary for the closure of zero-knowledge interactive proof systems under sequential composition [GK96]. Informally speaking, the inclusion of auxiliary inputs in the definition captures the notion that a given zero-knowledge interactive proof system cannot be used to *increase* knowledge, as opposed to prohibiting one from gaining knowledge starting from none. The second point concerns the order of quantification between V' and $S_{V'}$. Specifically, the definition states that a zero-knowledge interactive proof system is one such that for all V' there exists a simulator $S_{V'}$ that satisfies the requisite properties. There is an argument to be made for reversing these quantifiers by requiring that for a given interactive proof system (V, P) there should exist a single simulator S that interfaces in some uniform way with any given V' to produce an output that is indistinguishable from that verifier's output. Typical simulator constructions, as well as the ones that will be discussed in this paper in the quantum setting, do indeed satisfy this stronger requirement.

2.3 Quantum information and computation

By a *quantum register*, we simply mean a collection of qubits that we wish to view as a single unit and to which we give some name. Names of registers will always be uppercase letters in *sans serif* font, such as X, Y, and Z. The finite dimensional Hilbert spaces associated with registers will be denoted by capital script letters such as $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$, and it will generally be convenient to use the same letter in the two different fonts to denote a quantum register and its corresponding space. Dirac notation is used to express vectors in Hilbert spaces and linear mappings between them in a standard way. The norm of a vector $|\psi\rangle$ is written $\||\psi\rangle\|$, and the all-zero standard basis vector in a given space $\mathcal{X}$ is denoted $|0_{\mathcal{X}}\rangle$.

For given spaces $\mathcal{X}$ and $\mathcal{Y}$, the following notation is used to represent various sets of linear mappings from $\mathcal{X}$ to $\mathcal{Y}$. We let $\mathrm{L}(\mathcal{X},\mathcal{Y})$ denote the set of all linear mappings (or *operators*) from $\mathcal{X}$ to $\mathcal{Y}$, and write $\mathrm{L}(\mathcal{X})$ as shorthand for $\mathrm{L}(\mathcal{X},\mathcal{X})$. The set $\mathrm{Pos}(\mathcal{X})$ consists of all positive semidefinite operators acting on $\mathcal{X}$, and $\mathrm{D}(\mathcal{X})$ denotes the set of unit trace, positive semidefinite operators (i.e., density operators) acting on $\mathcal{X}$. Finally, we denote the set of unitary operators on $\mathcal{X}$ by $\mathrm{U}(\mathcal{X})$. The identity element of $\mathrm{L}(\mathcal{X})$ is denoted $\mathbb{1}_{\mathcal{X}}$.

A linear super-operator $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ is said to be *admissible* if it is completely positive and preserves trace. Admissible super-operators represent mappings from density operators to density operators that are physically realizable (in an idealized sense).

2.3.1 Measures of similarity and distance

The concept of quantum zero-knowledge requires formal notions of distance between quantum states and between admissible super-operators. The specific notions we will make use of are summarized in this section.

The *operator norm* of an operator $X \in \mathrm{L}(\mathcal{X},\mathcal{Y})$ is defined as

$$\|X\| \stackrel{\text{def}}{=} \max\{\|X|\psi\rangle\| \,:\, |\psi\rangle \in \mathcal{X},\ \||\psi\rangle\| = 1\},$$

and the *trace norm* is defined as

$$\|X\|_1 \stackrel{\text{def}}{=} \mathrm{Tr}\sqrt{X^*X}.$$

The significance of the trace norm in quantum information is that it relates very closely to the optimal probability with which two density operators can be distinguished by means of some measurement. In essence, the trace norm functions in a similar way to the 1-norm for differences between probability distributions.

For positive semidefinite operators $X, Y \in \mathrm{Pos}(\mathcal{X})$, the *fidelity* between X and Y is defined as

$$\mathrm{F}(X,Y) \stackrel{\text{def}}{=} \left\|\sqrt{X}\sqrt{Y}\right\|_1,$$

and the *squared-fidelity* is simply this quantity squared. For $|\phi\rangle \in \mathcal{X}$ and $X \in \mathrm{Pos}(\mathcal{X})$, the squared-fidelity of X with $|\phi\rangle\langle\phi|$ is $\langle\phi|X|\phi\rangle$. If $|\phi\rangle, |\psi\rangle \in \mathcal{X}\otimes\mathcal{Y}$ are vectors that *purify* X and Y, respectively, meaning that $\mathrm{Tr}_{\mathcal{Y}}|\phi\rangle\langle\phi| = X$ and $\mathrm{Tr}_{\mathcal{Y}}|\psi\rangle\langle\psi| = Y$, then it holds that

$$\mathrm{F}(X,Y) = \max\{|\langle\phi|\mathbb{1}\otimes U|\psi\rangle| \,:\, U \in \mathrm{U}(\mathcal{Y})\}.$$

The fidelity and the trace norm are related by the *Fuchs–van de Graaf Inequalities* [FvdG99]:

$$1 - \mathrm{F}(\rho,\xi) \leq \frac{1}{2}\|\rho - \xi\|_1 \leq \sqrt{1 - \mathrm{F}(\rho,\xi)^2}$$

for any choice of density operators ρ and ξ.

The notion of distance between admissible super-operators that we will use is given by *Kitaev's super-operator norm* [Kit97, KSV02, AKN98], which is commonly known as the *diamond norm*. For any super-operator $\Phi : \mathrm{L}(\mathcal{X}) \to \mathrm{L}(\mathcal{Y})$ the value of this norm is defined as

$$\|\Phi\|_\diamond \stackrel{\text{def}}{=} \max\left\{ \left\| (\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(X) \right\|_1 \,:\, X \in \mathrm{L}(\mathcal{X} \otimes \mathcal{W}) \,,\, \|X\|_1 \leq 1 \right\},$$

where $\mathcal{W}$ is any space with dimension equal to that of $\mathcal{X}$. (The value is the same for any choice of $\mathcal{W}$, provided its dimension is at least that of $\mathcal{X}$.) The diamond norm of the difference between two admissible super-operators satisfies

$$\|\Phi_0 - \Phi_1\|_\diamond = \max\left\{ \left\| (\Phi_0 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho) - (\Phi_1 \otimes \mathbb{1}_{\mathrm{L}(\mathcal{W})})(\rho) \right\|_1 \,:\, \rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W}) \right\},$$

meaning that the maximum in the definition occurs for some density operator. Because the trace norm is convex and super-operators are linear, the maximum is always achieved for some pure state $\rho = |\psi\rangle\langle\psi|$.

The intuition behind the diamond norm is as follows. Suppose that admissible super-operators Φ_0 and Φ_1 are given, both mapping $\mathrm{L}(\mathcal{X})$ to $\mathrm{L}(\mathcal{Y})$. If we fix some bipartite state $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{W})$ and apply either Φ_0 or Φ_1 to the part of this state corresponding to the space $\mathcal{X}$, then $\|\Phi_0 - \Phi_1\|_\diamond$ is the maximum trace-norm distance between the two possible outputs. By making use of the triangle inequality, one may observe the following important fact: if Φ_0 and Φ_1 are admissible super-operators for which $\|\Phi_0 - \Phi_1\|_\diamond$ is negligible, then no physical process that makes a polynomial number of evaluations of Φ_0 or Φ_1 can distinguish between the two mappings with non-negligible bias.

2.3.2 Quantum circuits

We will make reference to two types of quantum circuits in this paper: *unitary* quantum circuits and *general* quantum circuits. By unitary quantum circuits we mean circuits composed of unitary gates, chosen from some finite, universal set. It is not important for this paper that we choose a particular universal set, but we will make the simplifying assumption that this set is capable of performing reversible computations and phase-flips without error. General quantum circuits are composed of gates that may perform general admissible operations as opposed to just unitary operations. The inclusion of gates that perform general admissible operations does not change the computational power of quantum circuits [AKN98], but it is nevertheless convenient to refer to such circuits.

When we refer to a *purification* of a general quantum circuit, we mean a unitary circuit that simulates the general quantum circuit. As described in [AKN98], such a simulation is always possible by allowing the unitary circuit to act on the input qubits of the general circuit together will some number of additional *ancillary* qubits, which are initialized to the all-zero state before the unitary circuit is applied. In addition to the output qubits of the general circuit, the unitary circuit also produces some number of *residual* (or *garbage*) qubits that may be traced-out to yield the output of the general circuit. This process of circuit purification can always be done efficiently, in a gate-by-gate manner.

Sometimes it will be convenient to consider quantum circuits that implement measurements. A *measurement circuit* refers to any general quantum circuit, followed by a measurement of all of its output qubits with respect to the standard basis. It is of course straightforward to simulate intermediate measurements with measurements that are delayed until after the circuit has been

applied, so there is no loss of generality in defining measurement circuits in this way. We say that a measurement circuit is an *n-qubit measurement circuit* when it is helpful to refer explicitly to the number n of input qubits it takes. If Q is a measurement circuit that is applied to a collection of qubits in the state ρ, then $Q(\rho)$ is interpreted as a string-valued random variable describing the resulting measurement.

The *size* of a quantum circuit is defined to be the number of gates in the circuit plus the number of qubits on which it acts. This notion of size forbids the possibility that a tiny circuit acts on a large number of qubits. We assume that some reasonable scheme for encoding quantum circuits as binary strings has been fixed, where the length of such an encoding is always polynomially related to the circuit's size. A collection $\{Q_x : x \in \Sigma^*\}$ of quantum circuits is said to be *polynomial-time generated* if there exists a deterministic polynomial-time Turing machine that, on input $x \in \Sigma^*$, outputs an encoding of Q_x. Our assumptions about encoding schemes imply that if $\{Q_x : x \in \Sigma^*\}$ is a polynomial-time generated collection, then Q_x must have size polynomial in $|x|$.

3 Definitions of quantum zero-knowledge

This section presents the formal definitions of quantum zero-knowledge that are the main focus of this paper. These definitions concern the zero-knowledge property of both quantum and classical interactive proof systems against attacks by polynomial-time quantum verifiers. Definitions of quantum statistical zero-knowledge and quantum computational zero-knowledge are given. The definition of quantum computational zero-knowledge requires a formal notion of quantum computational indistinguishability, which is therefore also discussed in this section.

3.1 General notions

Let (V, P) be a quantum or classical interactive proof system for a promise problem A. We assume for simplicity that the structure of the interaction is completely determined by the length of the input x, meaning that the number, order, and length (but not contents) of the messages is a function of $|x|$ alone and not on random events or measurement outcomes performed by P or V.

An arbitrary (possibly cheating) verifier V' is any quantum computational process that interacts with P according to the message structure determined by $|x|$. Similar to the classical case, a verifier V' will take, in addition to the input string x, an auxiliary input, and produce some output. The most general situation allowed by quantum information theory is that both the auxiliary input and the output are quantum, meaning that the verifier operates on quantum registers whose initial state is arbitrary and may be entangled with some external system. Also similar to the classical case, we will assume that for any given polynomial-time verifier V' there exist polynomially bounded functions q and r that determine the number of auxiliary input qubits and output qubits of V'. To say that V' is a polynomial-time verifier means that the entire action of V' must be described by some polynomial-time generated family of quantum circuits.

The interaction of V' with P on input x is a physical process, and therefore induces some admissible super-operator from the verifier's $n = q(|x|)$ auxiliary input qubits to $m = r(|x|)$ output qubits. We will let $\mathcal{W}$ denote the vector space corresponding to the auxiliary input qubits, let $\mathcal{Z}$ denote the space corresponding to the output qubits, and let $\Phi_x : \mathrm{L}(\mathcal{W}) \to \mathrm{L}(\mathcal{Z})$ denote the resulting admissible super-operator induced by the interaction of V' with P on input x. It should be stated explicitly that the mapping Φ_x is completely determined for any choice of x, V', and P, assuming that V' and P agree on the same message structure on input x. In the situation that P is classical and V' tries to send quantum information to P, we assume that the qubits decohere

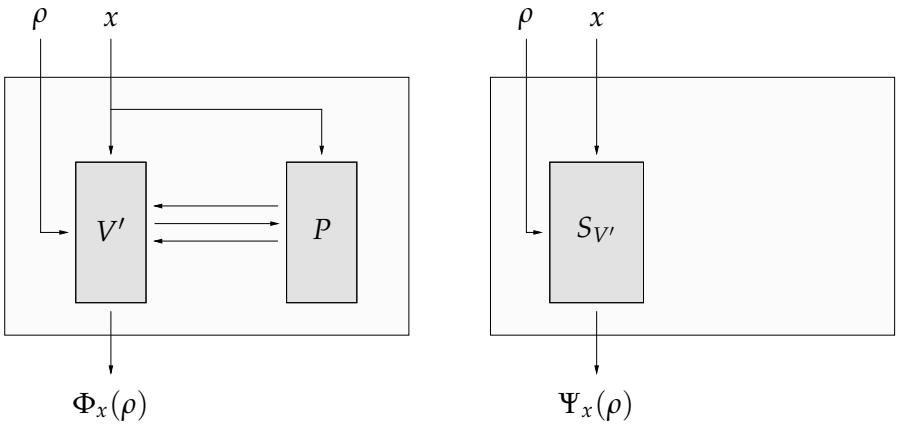

Figure 1: The mapping Φ_x is induced by the actual interaction between P and V', and the mapping Ψ_x is induced by the simulator alone.

immediately upon being touched by P. In other words, a classical prover P effectively measures immediately all qubits received from V' with respect to the standard basis.

A simulator $S_{V'}$ for a given verifier V' is described by a polynomial-time generated family of general quantum circuits that agrees with V' on the numbers n and m representing the number of auxiliary input qubits and output qubits respectively. Such a simulator does not interact with P, but simply induces an admissible operation that we will denote by $\Psi_x : \mathrm{L}(\mathcal{W}) \rightarrow \mathrm{L}(\mathcal{Z})$ on each input x.

Figure 1 illustrates the two mappings Φ_x and Ψ_x. Informally speaking, (V, P) is a quantum zero-knowledge interactive proof system for A if the two mappings Φ_x and Ψ_x are indistinguishable for every choice of $x \in A_{\text{yes}}$. As in the classical case, different notions of indistinguishability give rise to different variants of zero-knowledge. We will consider quantum *statistical* zero-knowledge and quantum *computational* zero-knowledge in the subsections that follow.

3.2 Quantum statistical zero-knowledge

The simpler variant of quantum zero-knowledge is quantum statistical zero-knowledge, which is defined as follows.

Definition 1. An interactive proof system (V, P) for a promise problem A is *quantum statistical zero-knowledge* if it holds that, for every polynomial-time generated quantum verifier V', there exists a polynomial-time generated quantum simulator $S_{V'}$ that satisfies the following requirements.

1. The verifier V' and simulator $S_{V'}$ agree on the polynomially bounded functions q and r that specify the number of auxiliary input qubits and output qubits, respectively.
2. Let Φ_x be the admissible super-operator that results from the interaction between V' and P on input x, and let Ψ_x be the admissible super-operator induced by the simulator $S_{V'}$ on input x, both as described above. Then there exists a negligible function δ such that $\|\Phi_x - \Psi_x\|_\diamond < \delta(|x|)$ for all $x \in A_{\text{yes}}$.

This definition is intended to be completely analogous to the classical definition. By referring to the diamond norm, it implicitly takes into account the fact that the input to Φ_x and Ψ_x could

be entangled to an external system (which happens to not be a concern in the classical setting). In operational terms, the definition implies that if (i) a pair of quantum registers (W, Y), with Y an arbitrary external register and W the register containing the auxiliary input to Φ_x or Ψ_x, is initialized to any state ρ, and (ii) one of the mappings Φ_x or Ψ_x is applied to W, yielding a new register Z, then the two possible resulting states of (Z, Y) will have negligible trace distance. Consequently, any measurement of the registers would then result in distributions having negligible statistical difference. By the triangle inequality, a similar statement holds for any procedure that is permitted to apply either Φ_x or Ψ_x a polynomial number of times, interleaved with arbitrary admissible operations. In short, the definition means that no physical process can distinguish the two boxes shown in Figure 1 with a non-negligible bias without making a super-polynomial number of queries.

It is necessary to include the possibility of an external system Y if one desires a cryptographically sound definition. For instance, it is possible to define admissible super-operators Φ and Ψ such that $\|\Phi(\rho) - \Psi(\rho)\|_1$ is strictly smaller than $\|\Phi - \Psi\|_\diamond$ for all $\rho \in \mathrm{D}(\mathcal{W})$ [KSV02]. Indeed, it is possible to define admissible super-operators Φ and Ψ such that $\|\Phi(\rho) - \Psi(\rho)\|_1$ is exponentially small (in the number of input and output qubits) for all states $\rho \in \mathrm{D}(\mathcal{W})$, but for which $\|\Phi - \Psi\|_\diamond = 2$, implying that an external system Y that is initially entangled with W allows the two mappings to be perfectly distinguished. One can imagine natural situations in which potential attacks on zero-knowledge proofs could be based on this principle.

3.3 Quantum computational indistinguishability and zero-knowledge

A quantum variant of computational zero-knowledge requires a formal notion of quantum computational indistinguishability of admissible super-operators. We discuss this notion in this section, as well as the closely related notion of quantum computational indistinguishability of states. This second notion will also be important later in the context of quantum computationally concealing commitment schemes.

The reader should be alerted to the fact that our definitions of quantum computational indistinguishability will be very strict, in that they allow polynomial-size quantum circuits to rely on an arbitrary auxiliary quantum state to aid them in distinguishing between two possibilities. This is required for the standard proofs of security that we adapt to the quantum setting. The true importance of the distinction between this notion of indistinguishability and one not allowing for such auxiliary states is an interesting topic, but is not considered further in this paper.

3.3.1 Distinguishability of states

We begin with a definition that quantifies computational distinguishability between states.

Definition 2. Let ρ and ξ be m-qubit mixed states. Then ρ and ξ are said to be (s, k, ε)-*distinguishable* if there exists a mixed state σ on k qubits and an $(m + k)$-qubit quantum measurement circuit Q of size s, such that

$$|\Pr[Q(\rho \otimes \sigma) = 1] - \Pr[Q(\xi \otimes \sigma) = 1]| \geq \varepsilon.$$

If ρ and ξ are not (s, k, ε)-*distinguishable* then they are (s, k, ε)-*indistinguishable*.

Notice that this definition gives a strong quantum analogue to the typical non-uniform notion of classical polynomial indistinguishability. It is strong because the non-uniformity includes an *arbitrary* quantum state σ that may aid some circuit Q in the task of distinguishing ρ from ξ. The inclusion of the arbitrary state σ is important in the situation we will consider in the context

of zero-knowledge, where indistinguishability of quantum states must hold in the presence of auxiliary quantum information.

Based on the previous definition, we now specify what it means for two ensembles of states to be quantum computationally indistinguishable.

Definition 3. Assume that $S \subseteq \Sigma^*$ is an infinite set of strings, r is a polynomially bounded function, and ρ_x and ξ_x are mixed states on $r(|x|)$ qubits for each $x \in S$. Then the ensembles $\{\rho_x : x \in S\}$ and $\{\xi_x : x \in S\}$ are *polynomially quantum indistinguishable* if, for every choice of polynomially bounded functions p, s, and k, it holds that ρ_x and σ_x are $(s(|x|), k(|x|), 1/p(|x|))$-indistinguishable for all but finitely many $x \in S$.

If $\{\rho_n : n \in \mathbb{N}\}$ and $\{\xi_n : n \in \mathbb{N}\}$ are ensembles indexed by the natural numbers, we identify S with 1^*, interpreting each n with its unary representation. Let us also note explicitly that the above definition includes the situation that classical probabilistic ensembles are to be distinguished. This corresponds to the case where the collections $\{\rho_x : x \in S\}$ and $\{\xi_x : x \in S\}$ are diagonal with respect to the standard basis.

It is convenient at this point to state and prove a simple fact concerning the distinguishability of states. It will not be used until later in the paper.

Proposition 4. *Suppose that $\rho_1, \ldots, \rho_n$ and $\xi_1, \ldots, \xi_n$ are m-qubit states such that $\rho_1 \otimes \cdots \otimes \rho_n$ and $\xi_1 \otimes \cdots \otimes \xi_n$ are (s, k, ε)-distinguishable. Then there exists at least one choice of $j \in \{1, \ldots, n\}$ for which ρ_j and ξ_j are $(s, (n-1)m + k, \varepsilon/n)$-distinguishable.*

Proof. The proposition is trivial if $n = 1$, so let us assume $n \geq 2$. Let Q be an $(nm + k)$-qubit measurement circuit of size s, and let σ be a k-qubit state, such that

$$|\Pr[Q(\rho_1 \otimes \cdots \otimes \rho_n \otimes \sigma) = 1] - \Pr[Q(\xi_1 \otimes \cdots \otimes \xi_n \otimes \sigma) = 1]| \geq \varepsilon.$$

For each $j \in \{1, \ldots, n\}$, define

$$\begin{aligned}\delta_j = \Big|&\Pr[Q(\rho_1 \otimes \cdots \otimes \rho_j \otimes \xi_{j+1} \otimes \cdots \otimes \xi_n \otimes \sigma) = 1] \\ &- \Pr[Q(\rho_1 \otimes \cdots \otimes \rho_{j-1} \otimes \xi_j \otimes \cdots \otimes \xi_n \otimes \sigma) = 1]\Big|.\end{aligned}$$

It is clear that Q is a measurement circuit that δ_j-distinguishes ρ_j and ξ_j by means of the auxiliary state

$$\rho_1 \otimes \cdots \otimes \rho_{j-1} \otimes \xi_{j+1} \otimes \cdots \otimes \xi_n \otimes \sigma$$

for each choice of j. By the triangle inequality we have

$$\sum_{j=1}^{n} \delta_j \geq \varepsilon$$

and thus $\delta_j \geq \varepsilon/n$ for at least one choice of $j \in \{1, \ldots, n\}$. □

3.3.2 Distinguishability of admissible super-operators

Next we will extend the notion of quantum computational indistinguishability to admissible super-operators.

Definition 5. Let Φ and Ψ be admissible super-operators from n qubits to m qubits. These super-operators are said to be (s, k, ε)-*distinguishable* if there exists a mixed state σ on $n + k$ qubits and an $(m + k)$-qubit measurement circuit Q of size s such that

$$|\Pr[Q((\Phi \otimes \mathbb{1}_k)(\sigma)) = 1] - \Pr[Q((\Psi \otimes \mathbb{1}_k)(\sigma)) = 1]| \geq \varepsilon.$$

Here, $\mathbb{1}_k$ denotes the identity super-operator on k qubits. If Φ and Ψ are not (s, k, ε)-*distinguishable* then they are said to be (s, k, ε)-*indistinguishable*.

In this definition, it should be viewed that the state σ plays multiple roles: it represents the input to the admissible super-operators, allows the possibility of auxiliary qubits that may be entangled with the qubits input to Φ or Ψ, and may include additional qubits that aid the measurement circuit in distinguishing the outputs as in Definition 2.

Again, this definition leads to a notion of quantum computational indistinguishability of admissible super-operators as in the following definition.

Definition 6. Assume that $S \subseteq \Sigma^*$ is an infinite set of strings, q and r are polynomially bounded functions, and Φ_x and Ψ_x are admissible super-operators from $q(|x|)$ qubits to $r(|x|)$ qubits for each $x \in S$. Then the ensembles $\{\Phi_x : x \in S\}$ and $\{\Psi_x : x \in S\}$ are *polynomially quantum indistinguishable* if, for every choice of polynomially bounded functions p, s, and k it holds that Φ_x and Ψ_x are $(s(|x|), k(|x|), 1/p(|x|))$-indistinguishable for all but finitely many $x \in S$.

3.3.3 Definition of quantum computational zero-knowledge

Now we are prepared to state a definition for quantum computational zero-knowledge.

Definition 7. An interactive proof system (V, P) for a promise problem A is *quantum computational zero-knowledge* if, for every polynomial-time generated quantum verifier V', there exists a polynomial-time generated quantum simulator $S_{V'}$ that satisfies the following requirements.

1. The verifier V' and simulator $S_{V'}$ agree on the polynomially bounded functions q and r that specify the number of auxiliary input qubits and output qubits, respectively.
2. Let Φ_x be the admissible super-operator that results from the interaction between V' and P on input x, and let Ψ_x be the admissible super-operator induced by the simulator $S_{V'}$ on input x, both as described above. Then the ensembles $\{\Phi_x : x \in A_{\text{yes}}\}$ and $\{\Psi_x : x \in A_{\text{yes}}\}$ are polynomially quantum indistinguishable.

3.4 A brief note on closure properties

Similar to the classical case, a sequential composition of quantum zero-knowledge protocols results in a zero-knowledge protocol—a property that again relies on the inclusion of the auxiliary input to the verifier. Specifically, the auxiliary input provides a means by which a cheating verifier's memory of previous interactions may be considered during an execution of a given protocol. Along similar lines, the presence of the auxiliary input implies that the classes of promise problems that are quantum statistical zero-knowledge and quantum computational zero-knowledge are closed under Karp reductions. This fact is discussed in [GMW91] for the classical case, and the quantum and classical settings do not differ in this respect.

In the above definitions we have assumed that the input x is classical. They therefore do not address, for instance, the situation in which one quantum zero-knowledge protocol is run in superposition inside of another.

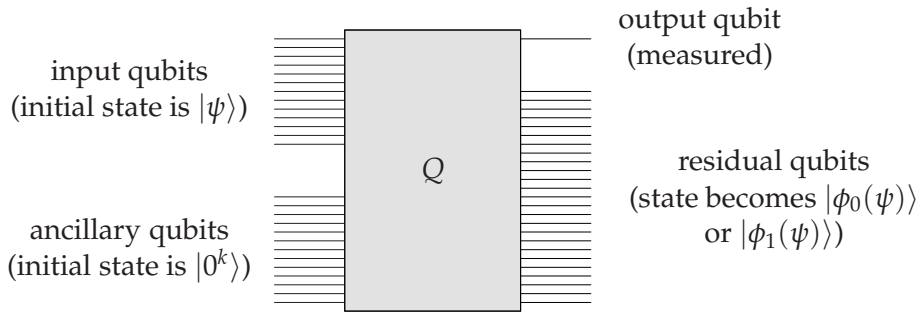

Figure 2: Given circuit Q for Quantum Rewinding Lemma

4 The Quantum Rewinding Lemma

The polynomial-time quantum simulator constructions for the various protocols considered in this paper rely on the *Quantum Rewinding Lemma* proved in this section. More precisely, two variants of the Quantum Rewinding Lemma are proved: an exact version (Lemma 8) and a version that allows for small perturbations (Lemma 9). Both variants concern an abstract computational problem that will now be discussed.

Let Q be a unitary quantum circuit acting on $n + k$ qubits. For an arbitrary pure quantum state $|\psi\rangle$ on n qubits, consider the process in which the circuit Q is applied to the state $|\psi\rangle |0^k\rangle$, after which the first qubit is measured with respect to the standard basis. Let $p(\psi)$ denote the probability that this measurement outcome is 0, and let us assume that $p(\psi)$ is neither 0 nor 1. Under this assumption, there is a unique choice of unit vectors $|\phi_0(\psi)\rangle$ and $|\phi_1(\psi)\rangle$ such that

$$Q |\psi\rangle |0^k\rangle = \sqrt{p(\psi)}\, |0\rangle |\phi_0(\psi)\rangle + \sqrt{1 - p(\psi)}\, |1\rangle |\phi_1(\psi)\rangle .$$

We write $|\phi_0(\psi)\rangle$ and $|\phi_1(\psi)\rangle$ to stress the dependence of these vectors on $|\psi\rangle$. These vectors represent the residual state of the $n + k - 1$ qubits aside from the first after the measurement takes place, respective to the measurement outcome. This situation is illustrated in Figure 2.

Now, let us imagine that it is our goal to construct from Q a procedure that takes as input an arbitrary state $|\psi\rangle$ and outputs a state that is as close as possible to $|\phi_0(\psi)\rangle$. The notion of closeness that we will focus on is the squared-fidelity between the output state and $|\phi_0(\psi)\rangle$. Note that by simply running Q on the state $|\psi\rangle |0^k\rangle$ and discarding the first qubit, we of course succeed in producing an output that has squared-fidelity at least $p(\psi)$ with $|\phi_0(\psi)\rangle$; and without further assumptions on the circuit Q it may not be possible to do significantly better.

The Quantum Rewinding Lemma establishes a condition on Q that allows for the state $|\phi_0(\psi)\rangle$ to be output with any desired fidelity. Specifically, under the assumption that the probability $p(\psi)$ is independent of the state $|\psi\rangle$, there exists an efficient procedure that outputs a state having very high fidelity with $|\phi_0(\psi)\rangle$. To achieve fidelity-squared at least $1 - \varepsilon$, the procedure requires

$$O\left(\frac{\log(1/\varepsilon)}{p(1-p)}\right)$$

executions of Q and Q^*, interleaved with simple unitary gates and measurements. The Quantum Rewinding Lemma with small perturbations concerns precisely the same procedure under slightly weaker assumptions.

Before stating and proving the Quantum Rewinding Lemma more formally, let us briefly discuss the connection between the above problem and the construction of simulators for zero-knowledge protocols. The circuit Q will represent a "reasonable attempt" to construct a simulator for some cheating verifier for some protocol, and the input state $|\psi\rangle$ will represent the auxiliary input of this verifier. It is very important that Q is unitary, so it is perhaps more accurate to view Q as being a *purification* of a "reasonable attempt" to simulate the given verifier. The measurement of the first qubit will indicate the success or failure of the simulation, with output 0 meaning that the simulation was successful and output 1 indicating that the simulation has failed, so rewinding is necessary. Supposing that the probability of a successful simulation is small but non-negligible, we therefore have that the Quantum Rewinding Lemma establishes a condition under which rewinding is possible: if the success probability of our "reasonable attempt" at a simulator is independent, or nearly independent, of the auxiliary input, then it is possible to generate the output corresponding to a successful simulation with very high fidelity.

This property of independence, or near-independence, of the simulator's success probability from the auxiliary input could potentially represent an obstacle to applying our method to some protocols. For the protocols considered in this paper, however, it does not—the most straightforward "reasonable attempts" to construct simulators will easily be shown to have this independence or near-independence property.

4.1 The exact case

We will begin with the Quantum Rewinding Lemma in the exact setting, where it is assumed that the measurement outcome in the process described above is completely independent of the input state $|\psi\rangle$.

For clarity, let us give a name to the type of circuit discussed above; we define an (n,k)-*quantum circuit* to be any unitary quantum circuit that acts on $n+k$ qubits, where the first n qubits may take an arbitrary quantum state $|\psi\rangle$ as input and the remaining k qubits are initially set to the state $|0^k\rangle$. Given such a circuit, the probability $p(\psi)$ and the residual states $|\phi_0(\psi)\rangle$ and $|\phi_1(\psi)\rangle$ are defined as above.

Lemma 8 (Quantum Rewinding Lemma, exact case)**.** *Let Q be an (n,k)-quantum circuit, and assume that $p = p(\psi)$ is constant over all choices of the input $|\psi\rangle$ and satisfies $p \in (0,1)$. Then for every $\varepsilon > 0$ there is a general quantum circuit R, with*

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

such that for every input $|\psi\rangle$, the output $\rho(\psi)$ of R satisfies

$$\langle\phi_0(\psi)|\rho(\psi)|\phi_0(\psi)\rangle \geq 1-\varepsilon. \tag{1}$$

Proof. For a given circuit Q and error bound ε, let R be a quantum circuit implementing the procedure described in Figure 3. The size of R can be seen to be as claimed. It therefore remains to prove that the output $\rho(\psi)$ of R on any input $|\psi\rangle$ satisfies the required bound (1).

To this end, let us define three projections, each acting on $n+k$ qubits:

$$\Pi_0 = |0\rangle\langle 0| \otimes \mathbb{1}, \quad \Pi_1 = |1\rangle\langle 1| \otimes \mathbb{1}, \quad \Delta = \mathbb{1} \otimes |0^k\rangle\langle 0^k|.$$

The measurement of the qubit B that is performed in the procedure may be viewed as a measurement with respect to the projections $\{\Pi_0, \Pi_1\}$, while the phase flip performed in case the

Initial conditions:

The register W contains an n-qubit quantum input $|\psi\rangle$.
The register X is initialized to the state $|0^k\rangle$.

The procedure:

Set $t = 0$.

Apply the circuit Q to the pair (W, X) obtaining (B, Y). (The register B represents the output qubit of Q and the register Y represents the remaining $n + k - 1$ residual qubits.)

Repeat:

Measure B with respect to the computational basis.

If the outcome of the measurement is 1:

Apply Q^* to (B, Y), obtaining (W, X).

Perform a phase flip in case any of the qubits of X is set to 1. (Equivalently, apply the unitary operation $2|0^k\rangle\langle 0^k| - \mathbb{1}$ to X.)

Apply Q to the pair (W, X) obtaining (B, Y).

Set $t = t + 1$.

Until the measurement outcome is 0 or $t = \lceil \log(1/\varepsilon)/(4p(1-p)) \rceil$.

Output the register Y.

Figure 3: Quantum Rewinding Procedure

measurement result is 1 may be written $2\Delta - \mathbb{1}$. Let us also define positive semidefinite operators P_0 and P_1 as follows:

$$P_0 = (\mathbb{1} \otimes \langle 0^k|)Q^*\Pi_0 Q(\mathbb{1} \otimes |0^k\rangle),$$
$$P_1 = (\mathbb{1} \otimes \langle 0^k|)Q^*\Pi_1 Q(\mathbb{1} \otimes |0^k\rangle).$$

The n-qubit measurement that is effectively performed on the quantum input $|\psi\rangle$ when the circuit Q is applied and the output qubit is measured is described as a POVM-type measurement by $\{P_0, P_1\}$.

The assumptions of the lemma imply that for every input state $|\psi\rangle$ we have $\langle\psi|P_1|\psi\rangle = 1 - p$. There is only one possibility for the operator P_1 given this fact: it must be that $P_1 = (1-p)\mathbb{1}$. This is because P_1, similar to any other linear operator, is uniquely determined by the function $|\psi\rangle \mapsto \langle\psi|P_1|\psi\rangle$ defined on the unit sphere. Consequently, we have

$$\Delta Q^*\Pi_1 Q\Delta = (\mathbb{1} \otimes |0^k\rangle)P_1(\mathbb{1} \otimes \langle 0^k|) = (1-p)\Delta. \tag{2}$$

Let us now fix an arbitrary n-qubit quantum input $|\psi\rangle$, so that $|\psi\rangle\,|0^k\rangle$ is the initial state of the pair (W, X) when the procedure described by R is run. After applying Q, we obtain the state

$$Q\,|\psi\rangle\,|0^k\rangle = \sqrt{p}\,|0\rangle\,|\phi_0(\psi)\rangle + \sqrt{1-p}\,|1\rangle\,|\phi_1(\psi)\rangle$$

in registers (B, Y). A measurement of B with respect to the standard basis now occurs. If the measurement results in outcome 0, then the residual state of register Y becomes $|\phi_0(\psi)\rangle$ and the

procedure is terminated, giving the desired output. If, however, the measurement outcome is 1, then the state of the pair (B, Y) becomes $|1\rangle\,|\phi_1(\psi)\rangle$. The operations that are performed in this case transform the state of the pair (B, Y) to

$$Q(2\Delta - \mathbb{1})Q^*\,|1\rangle\,|\phi_1(\psi)\rangle\,.$$

Using the above equation (2) along with the observation $\Delta\,|\psi\rangle\,|0^k\rangle = |\psi\rangle\,|0^k\rangle$, we may now calculate:

$$\begin{aligned}
Q(2\Delta - \mathbb{1})Q^*\,|1\rangle\,|\phi_1(\psi)\rangle &= \frac{1}{\sqrt{1-p}}Q(2\Delta - \mathbb{1})Q^*\Pi_1 Q\Delta\,|\psi\rangle\,|0^k\rangle \\
&= 2\sqrt{1-p}\,Q\,|\psi\rangle\,|0^k\rangle - \frac{1}{\sqrt{1-p}}\Pi_1 Q\,|\psi\rangle\,|0^k\rangle \\
&= 2\sqrt{p(1-p)}\,|0\rangle\,|\phi_0(\psi)\rangle + (1-2p)\,|1\rangle\,|\phi_1(\psi)\rangle\,.
\end{aligned}$$

This equation, which establishes that the vector

$$Q(2\Delta - \mathbb{1})Q^*\,|1\rangle\,|\phi_1(\psi)\rangle$$

lies in the two-dimensional space spanned by $|0\rangle\,|\phi_0(\psi)\rangle$ and $|1\rangle\,|\phi_1(\psi)\rangle$, is the key to the proof. Figure 4 illustrates the relationship among the relevant vectors. It should be stressed that this fact relies critically on the assumption that $p(\psi)$ is constant over all choices of $|\psi\rangle$.

We now see that a measurement of B at this point results in outcome 0 and corresponding residual state $|\phi_0(\psi)\rangle$ with probability $4p(1-p)$, and outcome 1 and residual state $|\phi_1(\psi)\rangle$ with probability $(1-2p)^2$. For each subsequent iteration of the loop, which is only performed in case the measurement outcome was 1, the pattern is identical. Consequently, whenever the measurement outcome is 0, the output of the procedure is $|\phi_0(\psi)\rangle$, and the probability that the measurement outcome is 0 within t iterations is

$$1 - (1-p)(1-2p)^{2t}.$$

The probability that $|\phi_0(\psi)\rangle$ is output by the procedure is therefore greater than $1-\varepsilon$ if at least

$$\frac{\log(1/\varepsilon)}{4p(1-p)}$$

iterations of the loop are permitted, which implies the required bound (1). □

A preliminary version of the present paper contained a somewhat less direct proof of the above lemma, based on the QMA error reduction technique presented in Marriott and Watrous [MW05]. Another proof has been suggested by Oded Regev [Reg06], based on the notion of angles between subspaces developed by Jordan [Jor75]. (Section VII.1 of Bhatia [Bha97] includes an extensive discussion of this notion.)

4.1.1 Relationship to Grover's Algorithm

It will be clear to some readers that the Quantum Rewinding Procedure described in Figure 3 has some resemblance to Grover's Algorithm [Gro96, Gro97, BBHT98] and the process known as Amplitude Amplification [BHMT02]. Specifically, if the measurement of B were replaced with a

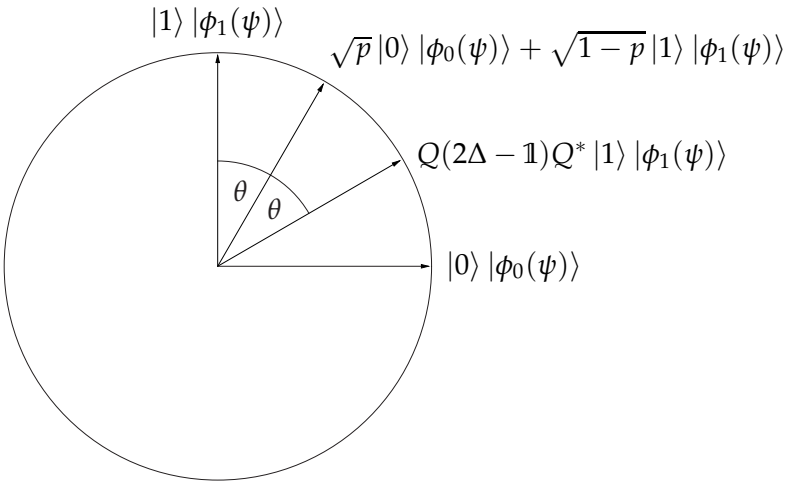

Figure 4: The action of $Q(2\Delta - \mathbb{1})Q^*$ on $|1\rangle\,|\phi_1(\psi)\rangle$. The angle between the vectors $|1\rangle\,|\phi_1(\psi)\rangle$ and $Q(2\Delta - \mathbb{1})Q^*\,|1\rangle\,|\phi_1(\psi)\rangle$ is twice the angle $\theta = \sin^{-1}(\sqrt{p})$.

phase-flip $2\Pi_0 - \mathbb{1}$, the resulting algorithm could reasonably be described as performing Amplitude Amplification with a quantum state input. Naturally, if there are no measurements within the loop, then the number of iterations of the loop that are performed must be determined by some other means. This approach has been investigated within the context of zero-knowledge by Matsumoto [Mat06].

It should be noted, however, that there are two fundamental differences between the Quantum Rewinding Procedure and Grover's Algorithm/Amplitude Amplification. The first major difference is of course that the Quantum Rewinding Procedure takes a quantum state input whereas, to the knowledge of the author, this has not been considered in the case of Grover's Algorithm or Amplitude Amplification. The requirement that $p(\psi)$ is constant over all input states $|\psi\rangle$, which has paramount importance in the above analysis, is obviously vacuous when there is no quantum state input.

The second difference is that, in the interest of simplicity, the Quantum Rewinding Procedure completely sacrifices the speed-up that represents the key feature of Grover's Algorithm and Amplitude Amplification. It has been argued above that the output of the procedure satisfies the required bound (1) and runs in polynomial time, even for exponentially small ε. When we apply the Quantum Rewinding Lemma to the construction of simulators for zero-knowledge protocols, it will be necessary that ε can be chosen to be negligible. If, however, we were to replace the measurement of B within the loop with a phase-flip in an attempt to mimic the behavior of Amplitude Amplification, it would be necessary in the general case to make further modifications to the procedure to guarantee the required bound (1). This could presumably be done using the techniques of [BHMT02], but the procedure would likely be significantly more complicated than the one described above.

A study of Amplitude Amplification with a quantum state input has the potential to be both interesting and useful. However, in the context of constructing zero-knowledge simulators this method has limited appeal. This is because a simulator for a zero-knowledge protocol is doing something that is *useless by assumption*—it acts only as a reduction, establishing that a given verifier must fail to extract knowledge from the honest prover. It therefore does not seem worthwhile to

optimize simulators, or to sacrifice their simplicity in the interest of the quadratic speed-up that could be offered by Amplitude Amplification.

4.2 Quantum rewinding with small perturbations

The assumptions of Lemma 8 require that p is independent of $|\psi\rangle$. Here we note that this assumption may be relaxed slightly if one is willing to accept a small perturbation in the output of the procedure.

Lemma 9 (Quantum Rewinding Lemma with small perturbations)**.** *Let Q be an (n,k)-quantum circuit and let $p_0, q \in (0,1)$ and $\varepsilon \in (0,1/2)$ be real numbers such that*

1. $|p(\psi) - q| < \varepsilon$,
2. $p_0(1-p_0) \leq q(1-q)$, *and*
3. $p_0 \leq p(\psi)$

for all n-qubit states $|\psi\rangle$. Then there exists a general quantum circuit R with

$$\mathrm{size}(R) = O\left(\frac{\log(1/\varepsilon)\,\mathrm{size}(Q)}{p_0(1-p_0)}\right)$$

such that, for every n-qubit state $|\psi\rangle$, the output $\rho(\psi)$ of R satisfies

$$\langle\phi_0(\psi)|\rho(\psi)|\phi_0(\psi)\rangle \geq 1 - 16\varepsilon\,\frac{\log^2(1/\varepsilon)}{p_0^2(1-p_0)^2}. \tag{3}$$

Readers not interested in the technical details of the proof of this lemma may be satisfied by the following brief summary of the proof, whose main idea is very simple. A given circuit Q may not satisfy the requirements of Lemma 8, but if it satisfies the weaker assumptions of Lemma 9, then Q must induce a unitary operation that is close (in operator norm) to a unitary operator U that does satisfy the requirements of Lemma 8. We then consider precisely the same circuit R that is described in the proof of Lemma 8, except with p_0 replacing p. The bound given by Lemma 9 is obtained from the one of Lemma 8 together with an analysis of the possible difference that could result by substituting U for Q in the circuit R.

Proof. The circuit construction for R is identical to the Quantum Rewinding Procedure given in the proof of Lemma 8, except that p_0 is substituted for p. The size of R is therefore as claimed, so it remains to prove that the stated bound (3) holds.

Let $\{|\psi_1\rangle, \ldots, |\psi_{2^n}\rangle\}$ be an orthonormal basis consisting of eigenvectors of the operator

$$(\mathbb{1} \otimes \langle 0^k|)Q^*\Pi_0 Q(\mathbb{1} \otimes |0^k\rangle).$$

We will assume hereafter that a general input state $|\psi\rangle$ is given by

$$|\psi\rangle = \sum_{i=1}^{2^n} \alpha_i |\psi_i\rangle .$$

The equation

$$Q|\psi\rangle|0^k\rangle = \sqrt{p(\psi)}\,|0\rangle|\phi_0(\psi)\rangle + \sqrt{1-p(\psi)}\,|1\rangle|\phi_1(\psi)\rangle$$

therefore implies that

$$|\phi_0(\psi)\rangle = \sum_{i=1}^{2^n} \alpha_i \sqrt{\frac{p(\psi_i)}{p(\psi)}} \, |\phi_0(\psi_i)\rangle \quad \text{and} \quad |\phi_1(\psi)\rangle = \sum_{i=1}^{2^n} \alpha_i \sqrt{\frac{1-p(\psi_i)}{1-p(\psi)}} \, |\phi_1(\psi_i)\rangle \,.$$

Now, given that

$$\Pi_0 Q(\mathbb{1} \otimes |0^k\rangle) \, |\psi_i\rangle = \sqrt{p(\psi_i)} \, |0\rangle \, |\phi_0(\psi_i)\rangle$$

for $i = 1, \ldots, 2^n$, we have that

$$0 = \langle \psi_i | \, (\mathbb{1} \otimes \langle 0^k|) Q^* \Pi_0 Q (\mathbb{1} \otimes |0^k\rangle) \, |\psi_j\rangle = \sqrt{p(\psi_i) p(\psi_j)} \, \langle \phi_0(\psi_i) | \phi_0(\psi_j) \rangle$$

for $i \neq j$. It follows that the set $\{|\phi_0(\psi_1)\rangle, \ldots, |\phi_0(\psi_{2^n})\rangle\}$ is orthonormal. By similar reasoning $\{|\phi_1(\psi_1)\rangle, \ldots, |\phi_1(\psi_{2^n})\rangle\}$ is orthonormal as well.

These orthonormality relations allow us to define a unitary operator U that is close to Q and satisfies the requirements of Lemma 8. Specifically let us define $U = VQ$, where V is the uniquely determined unitary operator that satisfies the equations

$$V \left(\sqrt{p(\psi_i)} \, |0\rangle \, |\phi_0(\psi_i)\rangle + \sqrt{1-p(\psi_i)} \, |1\rangle \, |\phi_1(\psi_i)\rangle \right) = \sqrt{q} \, |0\rangle \, |\phi_0(\psi_i)\rangle + \sqrt{1-q} \, |1\rangle \, |\phi_1(\psi_i)\rangle \,,$$
$$V \left(\sqrt{1-p(\psi_i)} \, |0\rangle \, |\phi_0(\psi_i)\rangle - \sqrt{p(\psi_i)} \, |1\rangle \, |\phi_1(\psi_i)\rangle \right) = \sqrt{1-q} \, |0\rangle \, |\phi_0(\psi_i)\rangle - \sqrt{q} \, |1\rangle \, |\phi_1(\psi_i)\rangle \,,$$

for $i = 1, \ldots, 2^n$, and acts trivially on the orthogonal complement to the subspace spanned by

$$\{|0\rangle \, |\phi_0(\psi_1)\rangle, \ldots, |0\rangle \, |\phi_0(\psi_{2^n})\rangle, |1\rangle \, |\phi_1(\psi_1)\rangle, \ldots, |1\rangle \, |\phi_1(\psi_{2^n})\rangle\} \,.$$

It is clear that

$$\|Q - U\| \leq \max_{1 \leq i \leq 2^n} \sqrt{\left(\sqrt{p(\psi_i)} - \sqrt{q}\right)^2 + \left(\sqrt{1-p(\psi_i)} - \sqrt{1-q}\right)^2} < \sqrt{2\varepsilon}.$$

Next, consider the effect of replacing Q by U in the circuit R. Let us denote by $\xi(\psi)$ the output of R when this replacement is made on input $|\psi\rangle$. We have that

$$U \, |\psi_i\rangle \, |0^k\rangle = \sqrt{q} \, |0\rangle \, |\phi_0(\psi_i)\rangle + \sqrt{1-q} \, |1\rangle \, |\phi_1(\psi_i)\rangle$$

for all choices of i, and therefore

$$U \, |\psi\rangle \, |0^k\rangle = \sqrt{q} \, |0\rangle \, |\delta_0(\psi)\rangle + \sqrt{1-q} \, |1\rangle \, |\delta_1(\psi)\rangle$$

for

$$|\delta_0(\psi)\rangle = \sum_{i=1}^{2^n} \alpha_i \, |\phi_0(\psi_i)\rangle \quad \text{and} \quad |\delta_1(\psi)\rangle = \sum_{i=1}^{2^n} \alpha_i \, |\phi_1(\psi_i)\rangle \,.$$

Given that $p_0(1-p_0) \leq q(1-q)$, we conclude from Lemma 8 that

$$\langle \delta_0(\psi) | \xi(\psi) | \delta_0(\psi) \rangle \geq 1 - \varepsilon. \tag{4}$$

It is convenient at this point to make use of a notion of distance known as the *fidelity distance*. This distance is defined by

$$d_{\mathrm{F}}(\sigma, \tau) \stackrel{\text{def}}{=} \min\{\| \, |\eta\rangle - |\mu\rangle \, \| \, : \, |\eta\rangle \text{ and } |\mu\rangle \text{ purify } \sigma \text{ and } \tau, \text{ respectively}\} = \sqrt{2 - 2\,\mathrm{F}(\sigma, \tau)}.$$

This is a proper notion of distance: it is symmetric, positive definite, and the triangle inequality holds [KSV02]. We will prove bounds on three quantities:

$$d_{\mathrm{F}}(\rho(\psi), \xi(\psi)),$$
$$d_{\mathrm{F}}(\xi(\psi), |\delta_0(\psi)\rangle \langle\delta_0(\psi)|),$$

and

$$d_{\mathrm{F}}(|\delta_0(\psi)\rangle \langle\delta_0(\psi)|, |\phi_0(\psi)\rangle \langle\phi_0(\psi)|).$$

The triangle inequality will therefore imply that the quantity $d_{\mathrm{F}}(\rho(\psi), |\phi_0(\psi)\rangle \langle\phi_0(\psi)|)$ is at most the sum of these three quantities. The equation $\mathrm{F}(\sigma, \tau)^2 \geq 1 - d_{\mathrm{F}}(\sigma, \tau)^2$, which is easily shown to hold for any choice of density operators σ and τ, will then yield the required bound (3).

Let us begin with a bound on $d_{\mathrm{F}}(\rho(\psi), \xi(\psi))$. The fidelity distance between these two density operators is no larger than the Euclidean distance between two purifications of our choice—so we may choose purifications that arise from running a purification of the circuit R. Given that the total number of times that either of Q or Q^* is applied during the execution of R is $2t + 1$, where

$$t = \left\lceil \frac{\log(1/\varepsilon)}{4p_0(1 - p_0)} \right\rceil,$$

the distance between the two purifications in question is at most $(2t + 1) \|Q - U\|$. Thus we have

$$d_{\mathrm{F}}(\rho(\psi), \xi(\psi)) \leq (2t + 1) \|Q - U\| \leq (2t + 1)\sqrt{2\varepsilon}. \tag{5}$$

Next, we wish to bound the quantity $d_{\mathrm{F}}(\xi(\psi), |\delta_0(\psi)\rangle \langle\delta_0(\psi)|)$. It follows from the above equation (4) that

$$d_{\mathrm{F}}(\xi(\psi), |\delta_0(\psi)\rangle \langle\delta_0(\psi)|) \leq \sqrt{2\varepsilon}. \tag{6}$$

Finally, the quantity $d_{\mathrm{F}}(|\delta_0(\psi)\rangle \langle\delta_0(\psi)|, |\phi_0(\psi)\rangle \langle\phi_0(\psi)|)$ is equal to the Euclidean distance between the vectors $|\delta_0(\psi)\rangle$ and $|\phi_0(\psi)\rangle$. As

$$\| |\delta_0(\psi)\rangle - |\phi_0(\psi)\rangle \|^2 = \sum_{i=1}^{2^n} |\alpha_i|^2 \left(1 - \sqrt{\frac{p(\psi_i)}{p(\psi)}}\right)^2 \leq \frac{2\varepsilon}{p_0},$$

we have

$$d_{\mathrm{F}}(|\delta_0(\psi)\rangle \langle\delta_0(\psi)|, |\phi_0(\psi)\rangle \langle\phi_0(\psi)|) \leq \sqrt{\frac{2\varepsilon}{p_0}}. \tag{7}$$

By combining the above equations (5), (6), and (7) by means of the triangle inequality, we have

$$d_{\mathrm{F}}(\rho(\psi), |\phi_0(\psi)\rangle \langle\phi_0(\psi)|) \leq \left(2t + 2 + \frac{1}{\sqrt{p_0}}\right)\sqrt{2\varepsilon} \leq \frac{4\log(1/\varepsilon)}{p_0(1 - p_0)}\sqrt{\varepsilon}.$$

Therefore

$$\langle\phi_0(\psi)|\rho(\psi)|\phi_0(\psi)\rangle \geq 1 - 16\varepsilon\frac{\log^2(1/\varepsilon)}{p_0^2(1 - p_0)^2}$$

as claimed. □

The bound given in the statement of the above lemma is clearly not tight—some precision is sacrificed in order to obtain a simpler expression. The lemma will later be used in the situation that ε is negligible, but where $p_0(1 - p_0)$ is not. The conclusion in this case is that the state produced by R has fidelity 1 minus some negligible quantity with the desired state $|\phi_0(\psi)\rangle$.

Zero-Knowledge Protocol for Graph Isomorphism

The input is a pair (G_0, G_1) of simple, undirected n-vertex graphs. It is assumed that the prover knows a permutation $\sigma \in S_n$ that satisfies $\sigma(G_1) = G_0$ if G_0 and G_1 are isomorphic.

Prover's step 1: Choose $\pi \in S_n$ uniformly at random and send $H = \pi(G_0)$ to the verifier.

Verifier's step 1: Choose $a \in \Sigma$ uniformly at random and send a to the prover. (Implicitly, the verifier is challenging the prover to exhibit an isomorphism between G_a and H.)

Prover's step 2: Set $\tau = \pi\sigma^a$ and send τ to the verifier. (If $\sigma(G_1) = G_0$, then $\tau(G_a) = H$.)

Verifier's step 2: Accept if $\tau(G_a) = H$, reject otherwise.

Figure 5: The Goldreich–Micali–Wigderson Graph Isomorphism protocol.

5 Quantum statistical zero-knowledge proof systems

The Goldreich–Micali–Wigderson Graph Isomorphism protocol is a simple and well-known example of an interactive proof system that is perfect (and therefore statistical) zero-knowledge against classical polynomial-time verifiers. The question of whether this protocol remains zero-knowledge against quantum verifiers was the starting point of the research in the present paper, and so it is fitting to illustrate our technique by first considering this protocol. Section 5.1 discusses this protocol and contains a proof that it indeed remains zero-knowledge against quantum verifiers.

The Goldreich–Micali–Wigderson protocol for Graph Isomorphism has a simple form:

1. The prover sends a message to the verifier,
2. the verifier flips a single coin and sends the result to the prover, and
3. the prover responds with a second message. The verifier then decides to accept or reject based on the three messages exchanged.

Protocols of this form are amenable to analysis in the quantum setting by means of the Quantum Rewinding Lemma of the previous section, provided certain assumptions are met. These assumptions translate to the independence of p from $|\psi\rangle$ in Lemma 8. In the quantum setting, protocols of this simple form are universal for *honest-verifier* quantum statistical zero-knowledge [Wat02], meaning that every problem having a quantum interactive proof system that is statistical zero-knowledge with respect to an honest verifier also has a proof system of the above form. This leads to a proof, discussed in Section 5.2, that honest verifier and general quantum statistical zero-knowledge interactive proof systems are equivalent with respect to computational power.

5.1 Graph isomorphism

Figure 5 describes the Goldreich–Micali–Wigderson Graph Isomorphism protocol [GMW91].

The corresponding interactive proof system has perfect completeness and soundness error $1/2$; if $G_0 \cong G_1$, then the verifier V will accept with probability 1 when interacting with the honest

prover P, while if $G_0 \ncong G_1$ then no prover P' can convince V to accept with probability greater than $1/2$ (essentially because the graph H sent by the prover in the first message cannot be isomorphic to both G_0 and G_1 when $G_0 \ncong G_1$).

For any choice of $\sigma \in S_n$ that satisfies the required property $\sigma(G_1) = G_0$ for $G_0 \cong G_1$, the interactive proof system (V, P) is perfect zero-knowledge with respect to any classical polynomial-time verifier V'. Sequential repetition followed by a unanimous vote can be used to decrease the soundness error to an exponentially small quantity while preserving the perfect completeness and classical zero-knowledge properties.

Our goal to prove this protocol is zero-knowledge with respect to polynomial-time *quantum* verifiers. It will be sufficient to consider a restricted type of verifier as follows:

1. In addition to (G_0, G_1), the verifier takes a quantum register W as input, representing the auxiliary quantum input. The verifier will use two additional quantum registers that function as work space: V, which is an arbitrary (polynomial-size) register, and A, which is a single qubit register. The registers V and A are initialized to their all-zero states before the protocol begins.

2. In the first message, the prover P sends an n-vertex graph H to the verifier. For each graph H there corresponds a unitary operator V'_H that the verifier applies to the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$. After applying the appropriate operation V'_H, the verifier measures the register A with respect to the standard basis, and sends the resulting bit a to the prover.

3. After the prover responds with some permutation $\tau \in S_n$, the verifier simply outputs the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$, along with the classical messages H and τ sent by the prover during the protocol.

Any polynomial-time quantum verifier can be modeled as a verifier of this restricted form followed by some polynomial-time post-processing of the restricted verifier's output. The same post-processing can be applied to the output of the simulator that will be constructed for the given restricted verifier. Notice that a verifier of this restricted form is completely determined by the collection $\{V'_H\}$.

Now let us consider the admissible super-operator induced by an interaction of a verifier of the above type with the prover P in the case that $G_0 \cong G_1$. Although the messages sent from the prover to the verifier are classical messages, it will simplify matters to view them as being stored in quantum registers denoted P_1 and P_2, respectively. Later, when we consider simulations of the interaction, we will need quantum registers to store these messages anyway, and it is helpful to have the registers used in the actual protocol and in the simulation share the same names.

Let us write $\mathcal{G}_n$ to denote the set of all simple, undirected graphs having vertex set $\{1, \ldots, n\}$. For each $H \in \mathcal{G}_n$ and each $a \in \Sigma$, define a linear mapping

$$M_{H,a} = (\mathbb{1}_{\mathcal{W}\otimes\mathcal{V}} \otimes \langle a|)\, V'_H (\mathbb{1}_{\mathcal{W}} \otimes |0_{\mathcal{V}\otimes\mathcal{A}}\rangle)$$

from $\mathcal{W}$ to $\mathcal{W} \otimes \mathcal{V}$. If the initial state of the register W is a pure state $|\psi\rangle \in \mathcal{W}$, then the state of the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ after the verifier applies V'_H is

$$(M_{H,0}\,|\psi\rangle)\,|0\rangle + (M_{H,1}\,|\psi\rangle)\,|1\rangle\,,$$

and therefore the state of the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ after the verifier applies V'_H and measures A with respect to the standard basis is

$$\sum_{a\in\Sigma} M_{H,a}\,|\psi\rangle\,\langle\psi|\,M^*_{H,a} \otimes |a\rangle\,\langle a|\,.$$

The admissible super-operator that results from the interaction is now easily described by incorporating the description of P. It is given by

$$\Phi(X) = \frac{1}{n!} \sum_{\pi \in S_n} \sum_{a \in \Sigma} M_{\pi(G_0),a} X M^*_{\pi(G_0),a} \otimes |a\rangle \langle a| \otimes |\pi(G_0)\rangle \langle \pi(G_0)| \otimes |\pi\sigma^a\rangle \langle \pi\sigma^a| \tag{8}$$

for all $X \in \mathrm{L}(\mathcal{W})$.

In order to define a simulator for a given quantum verifier V', it is helpful to consider the classical case. A classical simulation for a classical verifier in the above protocol may be obtained as follows. The simulator randomly chooses a permutation π and a bit b, and feeds $\pi(G_b)$ to the verifier. This verifier chooses a bit a for its message back to the prover. If $a = b$, the simulator can easily complete the simulation, otherwise it rewinds and tries a new choice of π and b. With very high probability, the simulator will succeed after no more than a polynomial number of steps, given that the event $a = b$ must happen with probability exactly $1/2$ (regardless of the verifier's actions). In case of success, the output of the simulator and the verifier will be identically distributed.

Our simulator for a quantum verifier proceeds along similar lines, except that we must invoke the Quantum Rewinding Lemma instead of ordinary classical rewinding. Our procedure will require two registers B and R in addition to W, V, A, P_1, and P_2. The register R may be viewed as a quantum register whose basis states correspond to the possible random choices that a typical classical simulator would use. In the present case this means a random permutation together with a random bit. The register B will represent the simulator's "guess" for the verifier's message. For convenience, let us define $\mathcal{E} = \mathcal{V} \otimes \mathcal{A} \otimes \mathcal{Y} \otimes \mathcal{B} \otimes \mathcal{Z} \otimes \mathcal{R}$, which is the space corresponding to the collection of all registers aside from W.

The procedure will involve a composition of a few operations that we now describe. First, let T be any unitary operator acting on registers $(\mathsf{P}_1, \mathsf{B}, \mathsf{P}_2, \mathsf{R})$ that maps the initial all-zero state of these four registers to the state

$$\frac{1}{\sqrt{2n!}} \sum_{b \in \Sigma} \sum_{\pi \in S_n} |\pi(G_b)\rangle |b\rangle |\pi\rangle |\pi, b\rangle .$$

If the register R is traced out, the state of registers $(\mathsf{P}_1, \mathsf{B}, \mathsf{P}_2)$ corresponds to a probability distribution over triples $(\pi(G_b), b, \pi)$ for b and π chosen uniformly. In essence, T produces a purification of a uniform distribution of possible *transcripts* of an interaction between a prover and verifier.

Next, define a unitary operator V' acting on registers $(\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{P}_1)$ that effectively simulates (unitarily) the verifier V'. Specifically, V' uses P_1 as a control register, and applies V'_H to registers $(\mathsf{W}, \mathsf{V}, \mathsf{A})$ for each possible graph $H \in \mathcal{G}_n$ representing a standard basis state of P_1. More compactly,

$$V' = \sum_{H \in \mathcal{G}_n} V'_H \otimes |H\rangle \langle H| .$$

The operators T and V' are each tensored with the identity on the remaining spaces when we wish to view them both as operators on $\mathcal{W} \otimes \mathcal{E}$.

Now consider the quantum circuit Q acting on all of the above registers that is obtained by first applying T, then applying V', and finally performing a controlled-NOT operation on the pair (A, B) with A acting as the control. Suppose that Q is applied to $|\psi\rangle |0_{\mathcal{E}}\rangle$, and the register B is then measured with respect to the standard basis. The probability that the outcome is 0 is necessarily equal to $1/2$, independent of the behavior of the verifier V' and of the auxiliary input $|\psi\rangle$. This follows from similar reasoning to the classical case: there can be no correlation between

the verifier's choice of a and the simulator's guess b for a because each graph H is equally likely to be derived from G_0 as G_1. If we condition on the measurement outcome being 0, and trace out the register R, we obtain precisely the admissible super-operator Φ given in Eq. (8) describing the actual interaction between V' and P. In other words, conditioned on the measurement outcome being 0, the circuit Q correctly simulates the interaction between V' and P given auxiliary input $|\psi\rangle$.

Given that the measurement outcome is 0 with probability $1/2$, which is independent of $|\psi\rangle$, we may apply Lemma 8 in order to obtain a circuit R. This circuit R, followed by the partial trace over R, represents the final simulation procedure.

Notice that because we are in the special case where $p = 1/2$, the simulation procedure in fact works perfectly after either zero or one iterations of the loop in the Quantum Rewinding Procedure. This establishes that the outcome of the simulation procedure is *precisely* $\Phi(|\psi\rangle\langle\psi|)$ in case the initial state of W was $|\psi\rangle$. This may be viewed as an improvement over the classical case, where it is not known if a perfect simulation is possible in worst-case polynomial time.

Because the set $\{|\psi\rangle\langle\psi| : |\psi\rangle \in \mathcal{W}, \||\psi\rangle\| = 1\}$ spans all of $\mathrm{L}(\mathcal{W})$, and the super-operator induced by the simulation procedure is necessarily admissible (and therefore linear), it holds that this map is precisely Φ. In other words, because admissible super-operators are uniquely determined by their action on pure states, the super-operator induced by the simulation procedure must be Φ; the simulation procedure implements *exactly* the same admissible super-operator as the actual interaction between V and P.

Each of the operations constituting the circuit Q can be performed by polynomial-size circuits, and therefore the simulator has polynomial size (in the worst case).

5.2 Honest versus general verifier QSZK

We now consider a quantum interactive proof system for a complete promise problem for the class $\mathrm{QSZK}_{\mathrm{HV}}$. This is the class of problems having *honest verifier* quantum statistical zero-knowledge interactive proof systems. This interactive proof system has the same three-message form as the Goldreich–Micali–Wigderson Graph Isomorphism protocol, wherein the verifier's message is a single *classical* coin-flip. In this case, the prover's messages will be quantum.

Of course it is not the case that every protocol of the above form is zero-knowledge. However, the similarities between the Graph Isomorphism protocol and the protocol for $\mathrm{QSZK}_{\mathrm{HV}}$ to be considered are sufficient to admit a similar analysis in terms of quantum attacks. This fact allows us to conclude that honest and general verifier quantum statistical zero-knowledge are computationally equivalent.

5.2.1 Definition of honest verifier quantum statistical zero-knowledge

Let us begin with some definitions that formalize the notion of an honest verifier. We will not require these definitions; they are only included for the sake of completeness. Further information about these definitions can be found in [Wat02].

Definition 10. Suppose that r is a polynomially bounded function and $\{\rho_x\}$ is a collection of quantum states, where each ρ_x is a state on $r(|x|)$ qubits. This collection is said to be *polynomial-time preparable* if there exists a polynomial-time uniformly generated family $\{Q_x\}$ of general quantum circuits such that each circuit Q_x takes no inputs, has $r(|x|)$ output qubits, and results in the state ρ_x when run.

Definition 11. Let (V, P) be a quantum interactive protocol for which V is described by a collection of *unitary* circuits. Define $\text{view}_{V,P}(x, j)$ to be the reduced state of the verifier and message qubits after j messages have been sent during an execution of the protocol on input x. The pair (V, P) is an *honest verifier quantum statistical zero-knowledge interactive proof system* for a promise problem A if:

1. (V, P) is a quantum interactive proof system for A, and
2. there exists a polynomial-time preparable set $\{\sigma_{x,j}\}$ and a negligible function δ such that

$$\|\sigma_{x,j} - \text{view}_{V,P}(x, j)\|_{\text{tr}} < \delta(|x|)$$

for every $x \in A_{\text{yes}}$ and each message number j.

We denote by QSZK_{HV} the class of all promise problems having honest verifier quantum statistical zero-knowledge interactive proof systems with completeness and soundness error at most $1/3$.

5.2.2 Equivalence of honest and general verifier quantum statistical zero-knowledge

Denote by QSZK the class of promise problems having quantum statistical zero-knowledge interactive proof systems with completeness and soundness error at most $1/3$. We will now prove that $\text{QSZK} = \text{QSZK}_{\text{HV}}$ by making use of some results proved in [Wat02].

First, define the ε-*Close Quantum States* problem as follows.

ε-Close Quantum States

Input:	General quantum circuits Q_0 and Q_1, both having no input qubits and m output qubits. Let ρ_0 and ρ_1 denote the states obtained by running Q_0 and Q_1, respectively.
Yes:	$\|\rho_0 - \rho_1\|_1 < \varepsilon$.
No:	$\|\rho_0 - \rho_1\|_1 > 2 - \varepsilon$.

One may consider ε to be a fixed constant or a function of the size of the description of Q_0 and Q_1, with each choice giving rise to a different promise.

The ε-*Close Quantum States* problem is complete for QSZK_{HV} for a fairly wide range of choices for ε. One specific fact that results and is well suited to our needs is as follows.

Theorem 12. *Every promise problem $A \in \text{QSZK}_{\text{HV}}$ Karp-reduces to an instance of the ε-Close Quantum States problem for which ε is a negligible function of the input size.*

Finally, let us recall that the protocol described in Figure 6 is an interactive proof system for ε-*Close Quantum States* with completeness error bounded by $\varepsilon/2$ and soundness error bounded by $1/2 + \sqrt{\varepsilon}/2$. Sequential repetition followed by a unanimous vote results in negligible bounds for both completeness and soundness errors.

Now, with these facts in hand, we are ready to prove the main result of this subsection.

Theorem 13. $\text{QSZK} = \text{QSZK}_{\text{HV}}$.

Zero-Knowledge Protocol for ε-Close Quantum States

Let R_0 and R_1 be unitary circuit purifications of Q_0 and Q_1, respectively. Assume that the circuits R_0 and R_1 both act on $m+k$ qubits, with one of the circuits padded with extra unused ancillary qubits if necessary.

Prover's step 1: Apply R_0 to $|0^{m+k}\rangle$ and send the first m qubits to the verifier.

Verifier's step 1: Choose $a \in \Sigma$ uniformly at random and send a to the prover.

Prover's step 2: Let U be a unitary operator on k qubits such that

$$\langle 0^{m+k}|R_1^*(\mathbb{1}\otimes U)R_0|0^{m+k}\rangle = \mathrm{F}(\rho_0,\rho_1).$$

If $a=1$, apply the unitary operation U to the residual qubits of R_0 that were not sent to the verifier in the first message, then send these qubits to the verifier. If $a=0$, send these qubits to the verifier without performing any operation on them.

Verifier's step 2: Apply R_a^* to all of the qubits received from the prover (in both the first and second message) and measure them in the standard basis: *accept* if the result is 0^{m+k}, and *reject* otherwise.

Figure 6: Protocol for ε-Close Quantum States

Proof. The fact that QSZK $\subseteq$ QSZK$_{\mathrm{HV}}$ follows easily from the definitions. Because QSZK is closed under Karp reductions, the reverse containment will follow from a proof that the *ε-Close Quantum States* problem is in QSZK when ε is negligible. This fact will be proved by means of the protocol (V,P) described in Figure 6. Because this protocol is already known to be a valid interactive proof system for the *ε-Close Quantum States* problem, it therefore remains to prove that this protocol is statistical zero-knowledge against any polynomial-time quantum verifier.

Similar to the analysis of the Graph Isomorphism protocol, it suffices to consider a restricted type of verifier V' as follows:

1. In addition to the descriptions of Q_0 and Q_1, the verifier takes a quantum register W as input, representing an auxiliary quantum input. The verifier will also use two additional quantum registers: V, which is an arbitrary (polynomial-size) register, and A, which is a single qubit register. The registers V and A are initialized to their all-zero states before the protocol begins.

2. The prover P sends an m-qubit quantum register P_1 to the verifier in the first message. The verifier applies a unitary operation, which we simply denote by V', to the registers $(\mathsf{W},\mathsf{V},\mathsf{A},\mathsf{P}_1)$, then sends A to the prover. (It will be irrelevant whether V' measures A first, so we will assume for simplicity that it does not.)

3. The prover responds with a k-qubit quantum register P_2. At this point the verifier simply outputs all of the registers in its possession: $\mathsf{W},\mathsf{V},\mathsf{P}_1$ and P_2.

Let us now describe the admissible operation Φ that is induced by an interaction between such a verifier V' and P for a given yes-instance (Q_0, Q_1) of *ε-Close Quantum States*. Because we will wish to bound the diamond norm of the difference between Φ and the super-operator induced by the simulator soon to be described, it will be helpful to include an arbitrary external quantum register in this description. Specifically, let Y be an external register of arbitrary size, and consider the situation that pair (W, Y) initially contains the pure state $|\psi\rangle \in \mathcal{W} \otimes \mathcal{Y}$. Let

$$V'_a = (\langle a| \otimes \mathbb{1}_{\mathcal{W} \otimes \mathcal{V} \otimes \mathcal{P}_1}) V'$$

for $a \in \Sigma$, and define

$$|\gamma(\psi)\rangle = \sum_{a \in \Sigma} |a\rangle \left(V'_a \otimes \mathbb{1}_{\mathcal{P}_2 \otimes \mathcal{Y}}\right) \left(|0_{\mathcal{V} \otimes \mathcal{A}}\rangle \left(\mathbb{1}_{\mathcal{P}_1} \otimes U^a\right) R_0 \, |0_{\mathcal{P}_1 \otimes \mathcal{P}_2}\rangle \, |\psi\rangle\right).$$

Then the result of applying $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}$ to $|\psi\rangle \langle\psi|$ is $\mathrm{Tr}_{\mathcal{A}} |\gamma(\psi)\rangle \langle\gamma(\psi)|$. The action of $\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})}$ on mixed states is determined by this expression.

Now we will construct a simulator for a given verifier V'. To this end, consider the following process that uses registers $\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{P}_1$ and P_2 as above, as well as an additional single qubit register B:

1. All registers are initialized to their all-zero state, aside from the register W that contains the auxiliary quantum input.
2. Transform the all-zero state of registers $(\mathsf{B}, \mathsf{P}_1, \mathsf{P}_2)$ to the state

   $$\frac{1}{\sqrt{2}} \sum_{b \in \Sigma} |b\rangle \, R_b \, |0^{m+k}\rangle .$$

   This is implemented by performing a Hadamard transform on B followed by implementations of R_0 and R_1 that are controlled by B appropriately.
3. Apply V' to registers $(\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{P}_1)$.
4. Perform a controlled-not operation with A as the control and B as the target. The register B now represents the measured output qubit of this process, while the remaining registers represent the residual qubits.

Let Q be a unitary quantum circuit that implements the above process.

Consider again the situation that the pair (W, Y) initially contains the pure state $|\psi\rangle$, where Y is an arbitrary external register as before. After the circuit Q is applied to W together with the initialized registers named above, the resulting state (of the registers $(\mathsf{B}, \mathsf{A}, \mathsf{V}, \mathsf{P}_1, \mathsf{P}_2, \mathsf{W}, \mathsf{Y})$) is

$$\frac{1}{\sqrt{2}} \sum_{a,b \in \Sigma} |a \oplus b\rangle \, |a\rangle \left(V'_a \otimes \mathbb{1}_{\mathcal{P}_2 \otimes \mathcal{Y}}\right) \left(|0_{\mathcal{V} \otimes \mathcal{A}}\rangle \, R_b \, |0_{\mathcal{P}_1 \otimes \mathcal{P}_2}\rangle \, |\psi\rangle\right).$$

Define

$$|\delta(\psi)\rangle = \sum_{a \in \Sigma} |a\rangle \left(V'_a \otimes \mathbb{1}_{\mathcal{P}_2 \otimes \mathcal{Y}}\right) \left(|0_{\mathcal{V} \otimes \mathcal{A}}\rangle \, R_a \, |0_{\mathcal{P}_1 \otimes \mathcal{P}_2}\rangle \, |\psi\rangle\right),$$

so that $\frac{1}{\sqrt{2}} |0\rangle \, |\delta(\psi)\rangle$ represents the projection of the above state produced by Q onto the space in which B contains 0.

It is not necessarily the case that $|\delta(\psi)\rangle$ is a unit vector, but it will be shown that it is close to a unit vector. The probability that a measurement of B yields 0 is

$$p(\psi) = \frac{1}{2} \| |\delta(\psi)\rangle \|^2 .$$

We have

$$\| R_1 |0_{\mathcal{P}_1 \otimes \mathcal{P}_2}\rangle - (\mathbb{1}_{\mathcal{P}_1} \otimes U) R_0 |0_{\mathcal{P}_1 \otimes \mathcal{P}_2}\rangle \| = \sqrt{2 - 2\,\mathrm{F}(\rho_0, \rho_1)},$$

and thus $\| |\gamma(\psi)\rangle - |\delta(\psi)\rangle \|$ is negligible, given that we have assumed (Q_0, Q_1) is a yes-instance of ε-*Close Quantum States*. We therefore have that $|p(\psi) - 1/2|$ is negligible.

Now consider the circuit R given by applying Lemma 9 to Q, where we take $p_0 = 1/4$, $q = 1/2$, and the error ε' to be a negligible function for which $|p(\psi) - 1/2| < \varepsilon'$. This circuit has polynomial size. Running this circuit on the pair of registers (W, Y) when in state $|\psi\rangle$ as above results in a mixed state $\rho(\psi)$ whose trace distance from

$$|\phi_0(\psi)\rangle \langle \phi_0(\psi)| = \frac{1}{2p(\psi)} |\delta(\psi)\rangle \langle \delta(\psi)|$$

is negligible. The trace distance between $\rho(\psi)$ and $|\gamma(\psi)\rangle \langle \gamma(\psi)|$ is therefore negligible by the triangle inequality.

Finally, we take our simulator to be given by the circuit R, followed by a partial trace operation on A. The resulting admissible super-operator Ψ is described by

$$(\Psi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Y})})(|\psi\rangle \langle \psi|) = \mathrm{Tr}_{\mathcal{A}}\, \rho(\psi).$$

We therefore have that $\| \Phi - \Psi \|_\diamond$ is negligible as required. □

Although the statement of this theorem may be viewed as a quantum analogue to the fact $\mathrm{SZK} = \mathrm{SZK}_{\mathrm{HV}}$ proved by Goldreich, Sahai, and Vadhan [GSV98], we hasten to add that there is no similarity in the proofs. The quantum case presented above is greatly simplified by the fact that every problem in $\mathrm{QSZK}_{\mathrm{HV}}$ has the very simple type of protocol represented by the one described in Figure 6.

Because $\mathrm{SZK} \subseteq \mathrm{QSZK}_{\mathrm{HV}}$, all problems in SZK have *quantum* interactive proof systems that are statistical zero-knowledge against quantum verifiers. The question of whether every problem in SZK has a *classical* interactive proof system that is zero-knowledge against quantum attacks is not answered in this paper.

6 Quantum computational zero-knowledge

The final protocol that will be discussed in this paper is the Goldreich–Micali–Wigderson Graph 3-Coloring protocol [GMW91]. This protocol is computational zero-knowledge against polynomial-time classical verifiers, assuming the existence of unconditionally binding and computationally concealing *commitment schemes*. In this section it is shown that this protocol is computational zero-knowledge against quantum verifiers, albeit with somewhat stronger intractability assumptions than are required in the classical case. This implies the existence of a quantum computational zero-knowledge interactive proof system for any problem in NP under the same assumptions, as a protocol for an arbitrary NP problem can begin with both parties computing a reduction to Graph 3-Coloring.

6.1 Quantum computationally concealing commitment schemes

The Goldreich–Micali–Wigderson Graph 3-Coloring protocol makes use of a commitment scheme. It is well-known that there cannot exist unconditionally binding and concealing bit commitments based on quantum information alone [LC97, May97], and therefore one must consider commitments for which either or both of the binding and concealing properties is based on a computational assumption. In the interactive proof system setting, where one requires soundness against arbitrary provers, the binding property of the commitments must be unconditional, and therefore the concealing property must be computationally-based.

Naturally, to be secure against quantum attacks, the commitment scheme that is used must in fact be *quantum* computationally concealing. The existence of such schemes does not follow from the existence of classical computationally concealing commitment schemes. For example, good candidates for classically secure schemes based on the computational difficulty of factoring or computing discrete logarithms become insecure in the quantum setting because of Shor's algorithm [Sho97].

Classical commitment schemes can, however, be based on arbitrary one-way functions [Nao91, HILL99], and there are candidates for such functions that may be difficult to invert even with efficient quantum algorithms. Functions based on lattice problems, error-correcting codes, and non-abelian group-theoretic problems represent possible candidates. By considering quantum one-way functions, the results of Refs. [Nao91, HILL99] may extend to the quantum setting, but to the knowledge of this author no such proof has yet been published. A quantum computationally concealing commitment scheme based on the existence of quantum one-way *permutations*, however, has been established by Adcock and Cleve [AC02]. Although the definitions in their paper differ somewhat from ours, in particular in that they do not consider the stronger form of non-uniformity allowing an auxiliary quantum state that we require, the result can be translated to our setting. This naturally requires a stronger notion of a permutation being one-way that forbids the possibility that a quantum circuit can invert a one-way permutation using an auxiliary input.

The following definition states the properties we require of such a commitment scheme.

Definition 14. Assume that Γ is a finite set with $|\Gamma| \geq 2$. An *unconditionally binding, quantum computationally concealing* Γ*-commitment scheme* consists of a deterministic polynomial-time computable function f with the following properties.

1. *(Uniform length.)* For some polynomially bounded function q we have that $|f(a,y)| = q(|y|)$ for every $a \in \Gamma$ and $y \in \Sigma^*$.
2. *(Binding property.)* For every choice of $a \neq b \in \Gamma$ and $y, z \in \Sigma^*$, we have $f(a,y) \neq f(b,z)$.
3. *(Concealing property.)* Let $F_N(a)$ be the density operator that results from the evaluation of $f(a,y)$ for a string $y \in \Sigma^N$ chosen uniformly at random. Then the ensembles

$$\{F_N(a) \,:\, N \in \mathbb{N}\} \qquad \text{and} \qquad \{F_N(b) \,:\, N \in \mathbb{N}\}$$

   are polynomially quantum indistinguishable for any choice of $a, b \in \Gamma$.

When such a scheme is used, it is assumed that some *security parameter* N is chosen. When one party (the prover in the 3-Coloring protocol) wishes to commit to a value $a \in \Gamma$, a string $y \in \Sigma^N$ is chosen uniformly at random and the string $f(a,y)$ is sent to the other party (the verifier in the 3-Coloring protocol). To reveal the commitment, the first party simply sends y along with the value a to the second party, who checks the validity of the commitment by computing $f(a,y)$.

Computational Zero-Knowledge Protocol for 3-Coloring

Assume the input is a graph G with n vertices and m edges. Let $\phi : \{1,\ldots,n\} \to \{1,2,3\}$ be any function that constitutes a valid 3-coloring of G if one exists. Also assume a quantum computationally concealing $\{1,2,3\}$-commitment scheme is given that is described by the function f. Repeat the following steps sequentially m^2 times:

Prover's step 1: Choose a permutation $\pi \in S_3$ of the colors $\{1,2,3\}$ and strings $y_1,\ldots,y_n \in \Sigma^N$ uniformly at random. Compute $c_i = f(\pi(\phi(i)), y_i)$ for each $i = 1,\ldots,n$, and send $c_1,\ldots,c_n$ to V. Informally: commit to the coloring $\pi \circ \phi$ of G for a random $\pi \in S_3$.

Verifier's step 1: Uniformly choose an edge $\{i,j\}$ of G and send this edge to P. (It is assumed that any dishonest verifier's message sent in this step decodes to a valid edge in G.)

Prover's step 2: Send the strings y_i and y_j to the verifier. Informally: reveal the committed colors for i and j.

Verifier's step 2: Check that there exist $a,b \in \{1,2,3\}$ such that $f(a,y_i) = c_i$, $f(b,y_j) = c_j$, and $a \neq b$, rejecting if not. Informally: check the validity of the commitments and that the committed colors a and b for i and j are different.

If the verifier has not rejected in any of the m^2 iterations, it accepts.

Figure 7: The Goldreich–Micali–Wigderson zero-knowledge protocol for 3-coloring.

6.2 The Goldreich–Micali–Wigderson Graph 3-Coloring protocol

The Goldreich–Micali–Wigderson Graph 3-Coloring protocol is described in Figure 7. In this protocol, one must specify a choice for the security parameter N. It will be sufficient (as it is classically) to set N to be equal to the number of vertices n of the input graph in order to prove the zero-knowledge property of the protocol.

Of course, one iteration of the loop in this protocol can be viewed as an interactive proof system for Graph 3-Coloring that has perfect completeness and a bound of roughly $1 - 1/m$ on the soundness error—by iterating the loop, we are simply performing sequential repetition. It will therefore be sufficient to prove that a single iteration of the loop is quantum computational zero-knowledge in order to prove the same for the entire protocol. Note that a single iteration of the loop has a very similar form to the protocols considered previously: the prover sends a message, the verifier uniformly chooses an edge in the graph (as opposed to flipping a coin as in the previous protocols), and the prover responds to the randomly chosen edge. This will allow us to use the Quantum Rewinding Procedure in a similar manner to the previously considered protocols.

Let us recall one way to construct a classical simulator for a given cheating verifier V'. The simulator uniformly chooses an edge $\{i,j\}$, and then selects some function $\mu : \{1,\ldots,n\} \to \{1,2,3\}$, subject to the constraint that $\mu(i)$ and $\mu(j)$ are uniformly random over the six possibilities with $\mu(i) \neq \mu(j)$. The function μ is a "fake coloring" that looks valid for the edge $\{i,j\}$. The simulator then computes commitments of the values $\mu(1),\ldots,\mu(n)$. These commitments are computation-

ally indistinguishable from commitments of $\pi(\phi(1)), \ldots, \pi(\phi(n))$ for a valid coloring ϕ when one exists. Given the commitments of $\mu(1), \ldots, \mu(n)$, along with whatever auxiliary input it may have been given, the verifier V' will choose some edge $\{i', j'\}$. In the idealized setting where one views the commitments as being *perfectly* concealing, the choice of $\{i', j'\}$ must agree with $\{i, j\}$ with probability $1/m$, independent of the actions of V'. This will not necessarily be the case when the commitments are only computationally concealing, but the probability of agreement must be nearly $1/m$ due to the fact that the commitments are computationally concealing. In case $\{i, j\} = \{i', j'\}$, the commitments of $\mu(i)$ and $\mu(j)$ are revealed, and the simulation can easily be completed. Otherwise, the simulator rewinds and the entire process is repeated. By repeating the process $O(m^2)$ times, say, the simulator is very likely to obtain an iteration in which $\{i, j\} = \{i', j'\}$, representing a successful simulation.

In the quantum case we will use a similar construction, except that we will of course use quantum rewinding. As in the QSZK case, we must use the Quantum Rewinding Lemma with small perturbations (Lemma 9).

The following assumptions are made on any cheating verifier V', and as before these assumptions do not affect the generality of the simulator construction:

1. In addition to the input graph G, the verifier takes an auxiliary quantum register W as input. The verifier will also use two additional quantum registers: V, which is an arbitrary (polynomial-size) register, and A, which has sufficiently many qubits to encode an edge of G. The registers V and A are initialized to their all-zero states before the protocol begins.

2. The prover P sends n registers $\mathsf{C}_1, \ldots, \mathsf{C}_n$, each of size $q(N)$ for the polynomially bounded function q representing the length of the commitments, to the verifier in the first message. As for the Graph Isomorphism protocol analysis, these registers are viewed as quantum registers for convenience—they contain classical information because P is classical. The verifier applies a unitary operation V' to the registers $(\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{C}_1, \ldots, \mathsf{C}_n)$, measures A with respect to the computational basis, and sends the resulting edge $\{i', j'\}$ to the prover.

3. The prover responds with two N-qubit registers $\mathsf{Y}_{i'}$ and $\mathsf{Y}_{j'}$, which again really contain classical information. We assume that the verifier simply outputs all of the registers in its possession: $\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{C}_1, \ldots, \mathsf{C}_n, \mathsf{Y}_{i'}$ and $\mathsf{Y}_{j'}$.

The admissible super-operator that results from an interaction between such a verifier V' and P is conceptually simple. A formal expression of this super-operator really only serves to obfuscate its simplicity, so we will avoid a formal expression.

Now let us define a simulator for a given V'. The simulator will need a collection of registers that includes the ones used by V' as well as some others. Specifically, the simulator uses registers

$$\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{C}_1, \ldots, \mathsf{C}_n, \mathsf{Y}_1, \ldots, \mathsf{Y}_n, \mathsf{B}, \text{ and } \mathsf{Z}.$$

The register B is the same size as A, while the register Z just stores any "garbage" qubits that are needed to implement the computation that will be described.

Consider first the following classical procedure that was described informally above:

1. Uniformly select an edge $\{i, j\}$ of G, and uniformly select colors $a_i, a_j \in \{1, 2, 3\}$ subject to $a_i \neq a_j$. Set $a_k = 1$ for $k \notin \{i, j\}$.

2. Prepare commitments of the colors $(a_1, \ldots, a_n)$ by uniformly choosing $y_k \in \Sigma^N$ and computing $f(a_k, y_k)$ for $k = 1, \ldots, n$.

This process can easily be performed by a polynomial-size general quantum circuit, and so we may consider a purification of this general circuit; which maps an all-zero state to a purification of the distribution that results from the above classical procedure. We will assume that the choice of the edge $\{i, j\}$ is stored in register B, the commitments and keys to $(a_1, \ldots, a_n)$ are stored in registers $(\mathsf{C}_k, \mathsf{Y}_k)$ for $k = 1, \ldots, n$, and all additional qubits collectively form the "garbage" register Z.

Now, let Q be a quantum circuit that first performs the quantum process just described, then applies V' to the registers W, V, A, and $\mathsf{C}_1, \ldots, \mathsf{C}_n$, and finally sets an output qubit to 0 if A and B contain the same edge, and 1 otherwise.

We wish to apply the Quantum Rewinding Lemma to this circuit. In order to do this, it must first be argued that the probability $p(\psi)$ that this circuit outputs 0 is such that $|p(\psi) - 1/m|$ is negligible, regardless of the auxiliary input $|\psi\rangle$. This follows from the fact that the commitments are computationally concealing, together with the fact that V' is described by a polynomial-size quantum circuit.

Let us argue this claim more precisely. The procedure performed by Q feeds into V' the n registers $\mathsf{C}_1, \ldots, \mathsf{C}_n$, along with the auxiliary input register W and the two initialized registers V and A. The reduced state of the registers $\mathsf{C}_1, \ldots, \mathsf{C}_n$ is given by

$$F_N(a_1) \otimes \cdots \otimes F_N(a_n),$$

averaged over some choice of n-tuples $(a_1, \ldots, a_n) \in \{1, 2, 3\}^n$, where $F_N(a)$ is as defined in Definition 14. The auxiliary register W is completely uncorrelated with $\mathsf{C}_1, \ldots, \mathsf{C}_n$, as it is not touched by Q prior to being fed into V'. Each possible edge is then output by V' with some probability depending on the distribution of n-tuples $(a_1, \ldots, a_n)$ and the initial state of W.

By Proposition 4, however, for sufficiently large n, any two choices of n-tuples $(a_1, \ldots, a_n)$ and $(a'_1, \ldots, a'_n)$ necessarily cause V' to produce output distributions having negligible statistical difference. This is because there are only polynomially many edges, and for any fixed choice of an edge the difference in the probability that V' outputs that edge for $(a_1, \ldots, a_n)$ and $(a'_1, \ldots, a'_n)$ must be negligible. If this were not so, then for at least one index i the commitments $F_N(a_i)$ and $F_N(a'_i)$ would be $(poly, poly, \varepsilon)$-distinguishable for non-negligible ε, contradicting the assumption that the commitment scheme is quantum computationally concealing.

Given that V' outputs each edge $\{i', j'\}$ with some probability that varies negligibly as a function of $(a_1, \ldots, a_n)$, we have that the output agrees with the random choice $\{i, j\}$ chosen by Q with probability varying negligibly from $1/m$ as claimed.

Now consider the state of the residual qubits of Q conditioned on a measurement of its output qubit being 0. The output state of the general quantum circuit R resulting from Lemma 9 will have negligible trace distance from this state. This state is over all of the registers discussed above, so the simulator must further process this state as follows:

1. Measure the register B, obtaining an edge $\{i, j\}$.
2. Output registers $\mathsf{W}, \mathsf{V}, \mathsf{A}, \mathsf{C}_1, \ldots, \mathsf{C}_n, \mathsf{Y}_i$ and Y_j. All remaining registers are traced out.

It remains to verify that the output state contained in these registers is computationally indistinguishable from the output of V' when interacting with P. This may be argued in manner similar to the above argument that Q outputs 0 with probability varying negligibly from $1/m$. For this purpose it is convenient to consider an intermediate process that functions exactly as the simulator constructed above, except that instead of setting $a_k = 1$ for $k \notin \{i, j\}$ it sets each such a_k in accordance with the appropriate permutation of the valid coloring ϕ. This intermediate process is computationally indistinguishable from the simulator, and statistically indistinguishable from the super-operator induced by the interaction between V' and P.

7 Conclusion

This paper has described a method by which some interactive proof systems can be proved to be zero-knowledge against quantum polynomial-time verifiers. A natural direction for further research is to better understand the applicability and limitations of this method.

Another interesting topic related to this paper concerns the existence of quantum one-way functions and permutations. In particular, the existence of quantum one-way permutations implies the existence of quantum computationally concealing commitment schemes [AC02], which were needed for our proof that the Goldreich–Micali–Wigderson Graph 3-Coloring protocol is zero-knowledge against quantum attacks. Are there good candidates for quantum one-way functions or permutations that can be efficiently computed in the forward direction by classical computers? One candidate has recently been proposed by Moore, Russell and Vazirani [MRV07], based on the difficulty of the Hidden Subgroup Problem.

Acknowledgements

I thank Claude Crépeau for sharing his thoughts and insight on zero-knowledge, and for getting me interested in the problem discussed in this paper. I also thank Gilles Brassard, Richard Cleve, Ivan Damgård, Simon-Pierre Desrosiers, Lance Fortnow, Dmitry Gavinsky, Dan Gottesman, Jordan Kerenidis, Hirotada Kobayashi, Keiji Matsumoto, Dieter van Melkebeek, Ashwin Nayak, Oded Regev, Amnon Ta-Shma, and Alain Tapp, among others, for helpful comments and general discussions of quantum zero-knowledge. Finally, I am indebted to the anonymous referees who provided many suggestions for improving this paper.

References

[AB06] S. Arora and B. Barak. *Complexity Theory: A Modern Approach (Preliminary web draft)*, 2006.

[AC02] M. Adcock and R. Cleve. A quantum Goldreich-Levin theorem with cryptographic applications. In *Proceedings of the 19th International Symposium on Theoretical Aspects of Computer Science*, volume 2285 of *Lecture Notes in Computer Science*, pages 323–334. Springer-Verlag, 2002.

[AKN98] D. Aharonov, A. Kitaev, and N. Nisan. Quantum circuits with mixed states. In *Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing*, pages 20–30, 1998.

[BB84] C. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing*, pages 175–179, 1984.

[BBHT98] M. Boyer, G. Brassard, P. Høyer, and A. Tapp. Tight bounds on quantum searching. *Fortschritte Der Physik*, 46(4–5):493–505, 1998.

[BCC88] G. Brassard, D. Chaum, and C. Crépeau. Minimum disclosure proofs of knowledge. *Journal of Computer and System Sciences*, 37:156–189, 1988.

[Bha97] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[BHMT02] G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *AMS Contemporary Mathematics Series*, pages 53–74. American Mathematical Society, 2002.

[DFS04] I. Damgård, S. Fehr, and L. Salvail. Zero-knowledge proofs and string commitments withstanding quantum attacks. In *Advances in Cryptology – CRYPTO 2004: 24th Annual International Cryptology Conference*, volume 3152 of *Lecture Notes in Computer Science*, pages 254–272. Springer-Verlag, 2004.

[ESY84] S. Even, A. Selman, and Y. Yacobi. The complexity of promise problems with applications to public-key cryptography. *Information and Control*, 61:159–173, 1984.

[FP96] C. Fuchs and A. Peres. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Physical Review A*, 53(4):2038–2045, 1996.

[FvdG99] C. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.

[GG00] O. Goldreich and S. Goldwasser. On the limits of nonapproximability of lattice problems. *Journal of Computer and System Sciences*, 60:540–563, 2000.

[GK96] O. Goldreich and H. Krawczyk. On the composition of zero-knowledge proof systems. *SIAM Journal on Computing*, 25(1):169–192, 1996.

[GMR89] S. Goldwasser, S. Micali, and C. Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989.

[GMW91] O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *Journal of the ACM*, 38(1):691–729, 1991.

[Gol01] O. Goldreich. *Foundations of Cryptography: Volume 1 – Basic Tools*. Cambridge University Press, 2001.

[Gol05] O. Goldreich. On promise problems (a survey in memory of Shimon Even [1935–2004]). Electronic Colloquium on Computational Complexity, Report TR05-018, 2005.

[Gra97] J. van de Graaf. *Towards a formal definition of security for quantum protocols*. PhD thesis, Université de Montréal, 1997.

[Gro96] L. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, pages 212–219, 1996.

[Gro97] L. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79(2):325–328, 1997.

[GSV98] O. Goldreich, A. Sahai, and S. Vadhan. Honest verifier statistical zero knowledge equals general statistical zero knowledge. In *Proceedings of the 30th Annual ACM Symposium on Theory of Computing*, pages 23–26, 1998.

[GV99] O. Goldreich and S. Vadhan. Comparing entropies in statistical zero-knowledge with applications to the structure of SZK. In *Proceedings of the 14th Annual IEEE Conference on Computational Complexity*, pages 54–73, 1999.

[HILL99] J. Håstad, R. Impagliazzo, L. Levin, and M. Luby. A pseudorandom function from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[Jor75] C. Jordan. Essai sur la géométrie à n dimensions. *Bulletin de la Société Mathématique de France*, 3:103–174, 1875.

[Kit97] A. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[Kob03] H. Kobayashi. Non-interactive quantum perfect and statistical zero-knowledge. In *ISAAC 2003 – Proceedings of the 14th International Symposium on Algorithms and Computation*, volume 2906 of *Lecture Notes in Computer Science*, pages 178–188. Springer-Verlag, 2003.

[KSV02] A. Kitaev, A. Shen, and M. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

[KW00] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof system. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, pages 608–617, 2000.

[LC97] H.-K. Lo and H. F. Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78(17):3410–3413, 1997.

[Mat06] K. Matsumoto. A simpler proof of zero-knowledge against quantum attacks using Grover's amplitude amplification. Available as arXiv.org e-print quant-ph/0602186, 2006.

[May97] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78(17):3414–3417, 1997.

[May01] D. Mayers. Unconditional security in quantum cryptography. *Journal of the ACM*, 48(3):351–406, 2001.

[MRV07] C. Moore, A. Russell, and U. Vazirani. A classical one-way function to confound quantum adversaries. Available as arXiv.org e-print quant-ph/0701115, 2007.

[MW05] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005.

[Nao91] M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991.

[NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[Pap94] C. Papadimitriou. *Computational Complexity*. Addison-Wesley, 1994.

[Reg06] O. Regev. Personal communication, 2006.

[RSA78] R. Rivest, A. Shamir, and L. Adleman. A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21(2):120–126, 1978.

[Sho97] P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997.

[SP00] P. Shor and J. Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Physical Review Letters*, 85(2):441–444, 2000.

[SV03] A. Sahai and S. Vadhan. A complete promise problem for statistical zero-knowledge. *Journal of the ACM*, 50(2):196–249, 2003.

[Wat02] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *Proceedings of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 459–468, 2002.

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