Quantum Commitments from Complexity Assumptions

André Chailloux
LRI, Université Paris-Sud
andre.chailloux@lri.fr

Iordanis Kerenidis
LIAFA, CNRS, Université Paris 7
jkeren@liafa.jussieu.fr

Bill Rosgen
CQT, National University of Singapore
bill.rosgen@nus.edu.sg

July 25, 2011

Abstract

Bit commitment schemes are at the basis of modern cryptography. Since information-theoretic security is impossible both in the classical and the quantum regime, we examine computationally secure commitment schemes. In this paper we study worst-case complexity assumptions that imply quantum bit-commitment schemes. First we show that $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ implies a computationally hiding and statistically binding auxiliary-input quantum commitment scheme. We then extend our result to show that the much weaker assumption $\mathsf{QIP} \not\subseteq \mathsf{QMA}$ (which is weaker than $\mathsf{PSPACE} \not\subseteq \mathsf{PP}$) implies the existence of auxiliary-input commitment schemes with quantum advice. Finally, to strengthen the plausibility of the separation $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ we find a quantum oracle relative to which honest-verifier QSZK is not contained in QCMA, the class of languages that can be verified using a classical proof in quantum polynomial time.

1 Introduction

The goal of modern cryptography is to design protocols that remain secure under the weakest possible complexity assumptions. Such fundamental protocols include commitment schemes, authentication, one-way functions, and pseudorandom generators. All these primitives have been shown equivalent: for example commitment schemes imply one-way functions [13] and one-way functions imply commitments [10, 11, 25].

In this paper we study complexity assumptions that imply commitment schemes, which are the basis for many cryptographic constructions, such as zero knowledge protocols for NP [3, 9]. A commitment scheme is a two-phase protocol between a sender and a receiver. In the commit phase, the sender interacts with the receiver so that by the end of the phase, the sender is bound to a specific bit, which remains hidden from the receiver until the reveal phase of the protocol, where the receiver learns the bit.

There are two security conditions for such schemes: binding (the sender cannot reveal more than one value) and hiding (the receiver has no information about the bit before the reveal phase). These conditions can hold statistically, i.e. against an unbounded adversary, or computationally, i.e. against a polynomial-time adversary. Without further assumptions these conditions cannot both hold statistically [21, 23].

The main complexity assumptions that have been used for the construction of one-way functions, and hence commitments, involve the classes of Computational and Statistical Zero Knowledge. Ostrovsky and Wigderson [27] proved that if Computational Zero Knowledge (ZK) is not trivial then there exists a family of functions that are not 'easy to invert'. The result was extended by Vadhan [33] to show that if ZK does not equal Statistical Zero Knowledge (SZK), then there exists an auxiliary-input one-way function, i.e. one can construct a one-way function given an auxiliary input (or else advice). Auxiliary-input cryptographic primitives are natural when considering worst-case complexity classes: the auxiliary input can encode a 'hard' instance of a problem known only to be hard in the worst case. Last, Ostrovsky and Wigderson also showed that if ZK contains a 'hard-on-average' problem, then 'regular' one-way functions exist.

With the advent of quantum computation and cryptography, one needs to revisit computational security, since many widely-used computational assumptions, such as the hardness of factoring or the discrete logarithm problem, become false when the adversary is a polynomial-time quantum machine [30].

In this paper, we study worst-case complexity assumptions under which quantum commitment schemes exist. As in the classical case, we obtain auxiliary-input commitments: commitments that can be constructed with classical and/or quantum advice. As our commitments are quantum, we define the computational security properties against quantum poly-time adversaries (who also receive an arbitrary quantum auxiliary input).

Our first result, involves the class of Quantum Statistical Zero Knowledge, QSZK.

Theorem 1.1. *If* $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ *there exists a non-interactive auxiliary-input quantum commitment scheme that is statistically-binding and computationally-hiding.*

Before explaining this result, let us try to see what an equivalent classical result would mean. At a high level, the classical statement would be of the following form: if SZK is not in MA, then auxiliary-input commitments exist. However, under some derandomization assumptions, we have that $\mathsf{NP} = \mathsf{MA} = \mathsf{AM}$ ([20, 24]) and since $\mathsf{SZK} \subseteq \mathsf{AM}$, we conclude that $\mathsf{SZK} \subseteq \mathsf{MA}$. Hence, the equivalent classical assumption is quite strong and, if one believes in derandomization, possibly false.

However, in the quantum setting, it would be surprising if QSZK is actually contained in QMA. We know that $\mathsf{QSZK} \subseteq \mathsf{QIP}[2]$ [37], where QIP[2] is the class of languages that have quantum interactive proofs with two messages (note that one only needs three messages to get the whole power of quantum interactive proofs). So far, any attempt to reduce QIP[2] or QSZK to QMA or find any plausible assumptions that would imply it, have not been fruitful. This seems harder than in the classical case. The main reason is that the verifier's message cannot be reduced to a public coin message nor to a pure quantum state. His message is entangled with his quantum workspace and this seems inherent for the class QIP[2] as well as for QSZK. It would be striking if one can get rid of this entanglement and reduce these classes to a single message from the prover.

If we weaken the security condition to hold against quantum adversaries with only classical auxiliary input, then the above assumption also becomes weaker, i.e. $\mathsf{QSZK} \not\subseteq \mathsf{QCMA}$, where QCMA is the class where the quantum verifier receives a single classical message from the prover. We give (quantum) oracle evidence for this by showing that

Theorem 1.2. *There exists a quantum oracle* A *such that* $\mathsf{QSZK}_{\mathsf{HV}}^{A} \not\subseteq \mathsf{QCMA}^{A}$.

Note that honest-verifier $\mathsf{QSZK}_{\mathsf{HV}} = \mathsf{QSZK}$ [37] in the unrelativized case. Our proof of this result extends Aaronson and Kuperberg's result that there is a quantum oracle A such that $\mathsf{QMA}^{A} \not\subseteq$

QCMA^A [2]. Subsequent to the completion of this work, Aaronson has shown the stronger result that there is an oracle A such that $\mathsf{SZK}^A \not\subseteq \mathsf{QMA}^A$ [1]. This result implies that our assumption that $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ is true relative to an oracle.

We then show the existence of commitment schemes based on a much weaker complexity assumption about quantum interactive proofs. More precisely, we look at the class QIP, which was first studied in [36]. This class is believed to be much larger than QSZK. We consider this class and its relation to QMA to show the following

Theorem 1.3. *If* $\mathsf{QIP} \not\subseteq \mathsf{QMA}$ *there exist non-interactive auxiliary-input quantum commitment schemes (both statistically hiding and computationally binding as well as statistically binding and computationally hiding) with quantum advice.*

Note, that $\mathsf{QIP} = \mathsf{PSPACE}$ [14] and $\mathsf{QMA} \subseteq \mathsf{PP}$ [22], so our assumption is extremely weak, in fact weaker than $\mathsf{PSPACE} \not\subseteq \mathsf{PP}$. Of course, with such a weak assumption we get a weaker form of commitment: the advice is now quantum. Thus, in order for the prover and the verifier to efficiently perform the commitment for a security parameter n, they need to receive a classical auxiliary input as well as quantum advice of size polynomial in n. This quantum advice is a quantum state on $\mathrm{poly}(n)$ qubits that is not efficiently constructible (otherwise, we could have reduced the quantum advice to classical advice by describing the efficient circuit that produces it). Moreover, the quantum advice we consider does not create entanglement between the players.

The key point behind this result is the structure of QIP. More precisely, we use the fact that there exists a QIP-complete problem where the protocol has only three rounds and the verifier's message is a single coin. The equivalent classical result would say that if three-message protocols with a single coin as a second message are more powerful than MA then commitments exist. Again, classically, if we believe that $\mathsf{AM} = \mathsf{MA}$, then this assumption is false. Taking this assumption to the quantum realm, it becomes 'almost' true, unless $\mathsf{PSPACE} = \mathsf{PP}$.

All of our commitment schemes are non-interactive, a feature that is useful in many applications. From $\mathsf{QIP} \not\subseteq \mathsf{QMA}$ we construct both statistically hiding and computationally binding commitments as well as statistically binding and computationally hiding ones, whose constructions are conceptually different. In order to prove the security of the first construction, we prove a parallel repetition theorem for protocols based on the swap test that may be of independent interest. From the $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ assumption we show here only statistically binding and computationally hiding commitments, but computationally binding and statistically hiding commitments can be similarly shown.

2 Definitions

In order to define the statistical distance between quantum states, we use the *trace norm*, given by $\|X\|_{\mathrm{tr}} = \mathrm{tr}\sqrt{X^\dagger X} = \max_U |\mathrm{tr}\, XU|$, where the maximization is taken over all unitaries of the appropriate size. Given one of two quantum states ρ, σ with equal probability, the optimal measurement to distinguish them succeeds with probability $1/2 + \|\rho - \sigma\|_{\mathrm{tr}}/4$ [12]. Note that this measurement is not generally efficient.

The *diamond norm* is a generalization of the trace norm to quantum channels that preserves the distinguishability characterization. Given one of two channels Q_0, Q_1 with equal probability, then the optimal distinguishing procedure that uses the channel only once succeeds with probability $1/2 + \|Q_0 - Q_1\|_\diamond/4$. The diamond norm is more complicated to define than the trace norm,

however, as the optimal distinguishing procedure may need to use an auxiliary space of size equal to the input space [18, 31]. For a linear map $Q\colon \mathbf{L}(\mathcal{H}) \to \mathbf{L}(\mathcal{K})$ with an auxiliary space $\mathcal{F}$ with $\dim \mathcal{F} = \dim \mathcal{H}$, the diamond norm can be defined as $\|Q\|_\diamond = \max_{X \in \mathbf{L}(\mathcal{H}\otimes\mathcal{F})} \|Q(X)\|_{\mathrm{tr}} / \|X\|_{\mathrm{tr}}$. One inconvenient property of the diamond norm is that for some maps the maximum in the definition may not be achieved on a quantum state. Fortunately, in the case of the difference of two completely positive maps this maximum is achieved by a pure state.

Lemma 2.1 ([29])**.** *Let $\Phi_0, \Phi_1\colon \mathbf{L}(\mathcal{H}) \to \mathbf{L}(\mathcal{K})$ be completely positive linear maps and let $\Phi = \Phi_0 - \Phi_1$. Then, there exists a space $\mathcal{F}$ and a state $|\phi^*\rangle \in \mathcal{F} \otimes \mathcal{H}$ such that*

$$\|\Phi\|_\diamond = \left\| (\mathbb{1}_{\mathbf{L}(\mathcal{F})} \otimes \Phi)(|\phi^*\rangle\langle\phi^*|) \right\|_{\mathrm{tr}}.$$

Closely related to the diamond norm is a norm studied in operator theory known as the completely bounded norm. An upper bound on this norm can be found in [28]. Since the diamond norm is dual to this norm, this bound may also be applied to the diamond norm. See [15] for a discussion of this bound and the relationship between the diamond and completely bounded norms.

Lemma 2.2. *Let $\Phi\colon \mathbf{L}(\mathcal{H}) \to \mathbf{L}(\mathcal{K})$ be a linear map, then*

$$\|\Phi\|_\diamond \leq (\dim \mathcal{H}) \|\Phi\|_{\mathrm{tr}} = (\dim \mathcal{H}) \sup_{X \in \mathbf{L}(\mathcal{H})} \frac{\|\Phi(X)\|_{\mathrm{tr}}}{\|X\|_{\mathrm{tr}}}.$$

In addition to these norms, we will also make use of the *fidelity* between two quantum states [16], which is given by $\mathrm{F}(\rho, \sigma) = \mathrm{tr}\sqrt{\sqrt{\sigma}\rho\sqrt{\sigma}}$. One property that is important for the results in this paper is that the fidelity only increases under the application of a quantum channel. Specifically, tracing out a portion of two states can only increase their fidelity, i.e. for ρ, σ density matrices on $\mathcal{H} \otimes \mathcal{K}$, it holds that $\mathrm{F}(\rho, \sigma) \leq \mathrm{F}(\mathrm{tr}_{\mathcal{K}} \rho, \mathrm{tr}_{\mathcal{K}} \sigma)$.

We also make significant use of the following two properties of the fidelity.

Lemma 2.3 ([8])**.** *For any density matrices ρ and σ, $1 - \mathrm{F}(\rho, \sigma) \leq \frac{1}{2}\|\rho - \sigma\|_{\mathrm{tr}} \leq \sqrt{1 - \mathrm{F}(\rho, \sigma)^2}$.*

Lemma 2.4 ([26, 32])**.** *For any density matrices ρ and σ, $\max_\xi \left(\mathrm{F}(\rho, \xi)^2 + \mathrm{F}(\xi, \sigma)^2\right) = 1 + \mathrm{F}(\rho, \sigma)$.*

2.1 Quantum Interactive Complexity Classes

The class QMA, first studied in [34], is informally the class of all problems that can be verified by a quantum polynomial-time algorithm with access to a quantum proof.

Definition 2.5. *A language L is in* QMA *if there is poly-time quantum algorithm V (called the* verifier*) such that*

1. *if $x \in L$, then there exists a state ρ such that $\Pr[V(x, \rho)$ accepts$] \geq a$,*
2. *if $x \notin L$, then for any state ρ, $\Pr[V(x, \rho)$ accepts$] \leq b$,*

where a, b are any efficiently computable functions of $|x|$ with $a > b$ with at least an inverse polynomial gap [19, 22]. If ρ is restricted to be a classical string, the class is called QCMA*.*

The class QIP, first studied in [36], consists of those problems that can be interactively verified in quantum polynomial time. A recent result is that QIP = PSPACE [14].

Definition 2.6. *A language $L \in \mathsf{QIP}$ if there is a poly-time quantum algorithm V exchanging quantum messages with an unbounded prover P such that for any input x*

1. *if $x \in L$ there exists a P such that, (V, P) accepts with probability at least a.*
2. *if $x \notin L$, then for any prover P, (V, P) accepts with probability at most b.*

As in QMA, we require only that $a > b$ with at least an inverse polynomial gap [17].

One key property of QIP is that any quantum interactive proof system can be simulated by one using only three messages [17]. This is not expected to hold in the classical case, as it would imply that $\mathsf{PSPACE} = \mathsf{AM}$. This property allows us to define simple problems involving quantum circuits that are complete for QIP.

In what follows we consider quantum unitary circuits C that output a state in the space $\mathcal{O} \otimes \mathcal{G}$. These spaces can be different for each circuit. $\mathcal{O}$ corresponds to the output space and $\mathcal{G}$ to the garbage space. For any circuit C, we define $|\phi_C\rangle = C|0\rangle$ in the space $\mathcal{O} \otimes \mathcal{G}$ to be the output of the circuit before the garbage space is traced out, and $\rho^C = \mathrm{Tr}_{\mathcal{G}}(|\phi_C\rangle\langle\phi_C|)$ to be the mixed state output by the circuit after the garbage space is traced out. We will also consider more general mixed-state quantum circuits C, that on an input state σ and output a quantum state, denoted by $C(\sigma)$. Unlike unitary circuits, mixed-state circuits are allowed to introduce ancillary qubits and trace out qubits during the computation. Note that circuits of this form can (approximately) represent any quantum channel. The size of a circuit C is equal to the number of gates in the circuit plus the number of qubits used by the circuit, denoted $|C|$. We will also use $|\mathcal{H}|$ to refer to the size of a Hilbert space $\mathcal{H}$ i.e. $|\mathcal{H}| = \lceil \log_2 \dim \mathcal{H} \rceil$. We use $\mathbf{L}(\mathcal{H})$ to refer to the set of all linear operators on $\mathcal{H}$, and $\mathbf{D}(\mathcal{H})$ to denote the subset of these operators that are density matrices. We consider two complete problems for QIP.

Definition 2.7 (QCD Problem)**.** *Let μ be a negligible function. We define the promise problem $\mathrm{QCD} = \{\mathrm{QCD}_Y, \mathrm{QCD}_N\}$ with input two mixed-state quantum circuits C_0, C_1 of size n as*

- $(C_0, C_1) \in \mathrm{QCD}_Y \Leftrightarrow \|C_0 - C_1\|_\diamond \geq 2 - \mu(n)$
- $(C_0, C_1) \in \mathrm{QCD}_N \Leftrightarrow \|C_0 - C_1\|_\diamond \leq \mu(n)$

Definition 2.8 (Π Problem)**.** *Let μ be a negligible function. We define the promise problem $\Pi = \{\Pi_Y, \Pi_N\}$ with input two mixed-state quantum circuits C_0, C_1 of size n, where for each i $C_i : \mathbf{D}(\mathcal{X} \otimes \mathcal{Y}) \to \{0, 1\}$, as*

- $(C_0, C_1) \in \Pi_Y \Leftrightarrow \exists \rho^0, \rho^1 \in \mathbf{D}(\mathcal{X} \otimes \mathcal{Y})$ *with* $tr_{\mathcal{X}}(\rho^0) = tr_{\mathcal{X}}(\rho^1)$ *such that*

$$\frac{1}{2}\left(\Pr[C_0(\rho^0) = 1] + \Pr[C_1(\rho^1) = 1]\right) = 1$$

- $(C_0, C_1) \in \Pi_N \Leftrightarrow \forall \rho^0, \rho^1 \in \mathbf{D}(\mathcal{X} \otimes \mathcal{Y})$ *with* $tr_{\mathcal{X}}(\rho^0) = tr_{\mathcal{X}}(\rho^1)$ *we have*

$$\frac{1}{2}\left(\Pr[C_0(\rho^0) = 1] + \Pr[C_1(\rho^1) = 1]\right) \leq \frac{1}{2} + \mu(n)$$

QCD is QIP-complete [29]. The QIP-completeness of Π follows from a characterization of QIP due to Mariott and Watrous [22] that states that any problem in QIP has a three message protocol where the challenge from the Verifier consists of a single coin flip. We may also assume that this protocol has perfect completeness and soundness error negligibly larger than 1/2. Taking the circuits C_0 and C_1 as the final circuit of the Verifier in such a proof system when the challenge is either 0 or 1 results in an instance of the problem Π. The QIP-completeness of Π then follows directly from the completeness and soundness conditions on the proof system.

The complexity class QSZK, introduced in [35], is the class of all problems that can be interactively verified by a quantum verifier who learns nothing beyond the truth of the assertion being verified. In the case that the verifier is *honest*, i.e. does not deviate from the protocol in an attempt to gain information, this class can be defined as

Definition 2.9. *A language $L \in \mathsf{QSZK}_{\mathsf{HV}}$ if*

1. *There is a quantum interactive proof system for L.*
2. *If $x \in L$, the state of the verifier in this proof system after the sending of each message can be approximated, within negligible trace distance, by a polynomial-time preparable quantum state.*

If we insist that item 2 holds when the Verifier departs from the protocol, the result is the class QSZK. Watrous has shown that $\mathsf{QSZK}_{\mathsf{HV}} = \mathsf{QSZK}$ [37]. This class has complete problems. We use the following QSZK-complete problem [35].

Definition 2.10 (QSD Problem)**.** *Let μ be a negligible function.* $\mathrm{QSD} = \{\mathrm{QSD}_Y, \mathrm{QSD}_N\}$ *is the promise problem on input (C_0, C_1), unitary circuits of size n with m output qubits, such that*

- $(C_0, C_1) \in \mathrm{QSD}_Y \Leftrightarrow \left\| \rho^{C_0} - \rho^{C_1} \right\|_{\mathrm{tr}} \geq 2 - \mu(n)$
- $(C_0, C_1) \in \mathrm{QSD}_N \Leftrightarrow \left\| \rho^{C_0} - \rho^{C_1} \right\|_{\mathrm{tr}} \leq \mu(n)$

2.2 Quantum Computational Distinguishability

The following definitions may be found in [37].

Definition 2.11. *Two mixed states ρ^0 and ρ^1 on m qubits are (s, k, ε)-distinguishable if there exists a mixed state σ on k qubits and a quantum circuit D of size s that performs a two-outcome measurement on $(m + k)$ qubits, such that $|\Pr[D(\rho^0 \otimes \sigma) = 1] - \Pr[D(\rho^1 \otimes \sigma) = 1]| \geq \varepsilon$. If ρ^0 and ρ^1 are not (s, k, ε)-distinguishable, then they are (s, k, ε)-indistinguishable.*

Let $I \subseteq \{0,1\}^*$ and let an *auxiliary-input state ensemble* be a collection of mixed states $\{\rho_x\}_{x \in I}$ on $r(|x|)$ qubits for polynomial r with the property that ρ_x can be efficiently generated given x.

Definition 2.12. *Two auxiliary-input state ensembles $\{\rho_x^0\}$ and $\{\rho_x^1\}$ on I are* quantum computationally indistinguishable *if for all polynomials p, s, k and for all but finitely many $x \in I$, ρ_x^0 and ρ_x^1 are $(s(|x|), k(|x|), 1/p(|x|))$-indistinguishable. Ensembles $\{\rho_x^0\}$ and $\{\rho_x^1\}$ on I are* quantum computationally distinguishable *if there exist polynomials p, s, k such that for all $x \in I$, ρ_x^0 and ρ_x^1 are $(s(|x|), k(|x|), 1/p(|x|))$-distinguishable.*

At first glance these definitions of distinguishability and indistinguishability are not complementary. We require distinguishability for all $x \in I$, but require indistinguishability in only all but finitely many $x \in I$. This is because $|x|$ will be our security parameter, and so while a polynomially-bounded adversary may be able to distinguish the two ensembles for a finite number of (small) values of $|x|$, as the parameter grows no efficient algorithm can distinguish the two ensembles.

Key to this definition is that if two ensembles are computationally distinguishable, then for all x there exists an efficient procedure in $|x|$ that distinguishes ρ_x^0 and ρ_x^1 with probability at least $1/2 + 1/p(|x|)$. Note that this is not a uniform procedure: the circuit that distinguishes the two states may depend on x.

Definition 2.13. *Two auxiliary-input state ensembles $\{\rho_x^0\}$ and $\{\rho_x^1\}$ on I are* quantum statistically indistinguishable *if for any polynomial p and for all but finitely many $x \in I$, $\left\|\rho_x^0 - \rho_x^1\right\|_{\mathrm{tr}} \leq 1/p(|x|)$.*

Definition 2.14. *Two admissible superoperators Φ^0 and Φ^1 from t qubits to m qubits are (s, k, ε)-distinguishable if there exists a mixed state σ on $t + k$ qubits and a quantum circuit D of size s that performs a two-outcome measurement on $(m + k)$ qubits, such that $|\Pr[D((\Phi^0 \otimes \mathbb{1}_k)(\sigma)) = 1] - \Pr[D((\Phi^1 \otimes \mathbb{1}_k)(\sigma)) = 1]| \geq \varepsilon$, where $\mathbb{1}_k$ denotes the identity superoperator on k qubits. If the superoperators Φ^0 and Φ^1 are not (s, k, ε)-distinguishable, then they are (s, k, ε)-indistinguishable.*

Let $I \subseteq \{0,1\}^*$ and let an *auxiliary-input superoperator ensemble* be a collection of superoperators $\{\Phi_x\}_{x \in I}$ from $q(|x|)$ to $r(|x|)$ qubits for some polynomials q, r, where as in the case of states, given x the superoperators can be performed efficiently in $|x|$.

Definition 2.15. *Two auxiliary-input superoperator ensembles $\{\Phi_x^0\}$ and $\{\Phi_x^1\}$ on I are* quantum computationally indistinguishable *if for all polynomials p, s, k and for all but finitely many $x \in I$, Φ_x^0 and Φ_x^1 are $(s(|x|), k(|x|), 1/p(|x|))$-indistinguishable. Auxiliary-input ensembles $\{\Phi_x^0\}$ and $\{\Phi_x^1\}$ on I are* quantum computationally distinguishable *if there exist polynomials p, s, k such that for all $x \in I$, Φ_x^0 and Φ_x^1 are $(s(|x|), k(|x|), 1/p(|x|))$-distinguishable.*

If two superoperator ensembles are computationally distinguishable then there is an efficient (nonuniform) procedure (in $|x|$) to distinguish them with probability at least $1/2 + 1/p(|x|)$ for some polynomial p. If the property of being (s, k, ε)-indistinguishable holds for all (unbounded) s and all polynomial $k, 1/\varepsilon$, then we call an ensemble statistically indistinguishable. Note that these definitions provide a strong quantum analogue of the classical non-uniform notion of computational indistinguishability, since the non-uniformity includes an arbitrary quantum state as advice to the distinguisher.

We define a new notion that we will use later on. Intuitively, two circuits that take input in the space $\mathcal{X} \otimes \mathcal{Y}$ and output a single bit are witnessable if there exist two input states that are identical on $\mathcal{Y}$ and are accepted by the two circuits with high probability.

Definition 2.16. *Two superoperators Φ^0 and Φ^1 from $\mathbf{L}(\mathcal{X} \otimes \mathcal{Y})$ to a single bit are (s, k, p)-witnessable if there exist two input states $\rho^0, \rho^1 \in \mathbf{L}(\mathcal{X} \otimes \mathcal{Y})$ such that*

1. $\frac{1}{2}\left(\Pr[\Phi^0(\rho^0) = 1] + \Pr[\Phi^1(\rho^1) = 1]\right) \geq 1/2 + 1/p(n)$
2. *there exists a state $\sigma \in \mathbf{L}(\mathcal{W} \otimes \mathcal{X} \otimes \mathcal{Y})$ with $|\mathcal{W}| = k$ and $\mathrm{tr}_{\mathcal{W}}\, \sigma = \rho_0$, and an admissible superoperator $\Psi : \mathbf{L}(\mathcal{W} \otimes \mathcal{X}) \to \mathbf{L}(\mathcal{X})$ of size s, such that $\rho^1 = (\Psi \otimes \mathbb{1}_{\mathbf{L}(\mathcal{Y})})(\sigma)$ where $\mathbb{1}_{\mathbf{L}(\mathcal{Y})}$ denotes the identity on $\mathbf{L}(\mathcal{Y})$.*

If Φ^0 and Φ^1 are not (s,k,p)-witnessable, then they are (s,k,p)-unwitnessable.

Let $I \subseteq \{0,1\}^*$ and let an *auxiliary-input superoperator ensemble* be a collection of superoperators $\{\Phi_x\}_{x\in I}$ from $q(|x|)$ to 1 bit for a polynomial q, where given x the superoperators can be performed efficiently in $|x|$.

Definition 2.17. *Auxiliary-input superoperator ensembles $\{\Phi_x^0\}$ and $\{\Phi_x^1\}$ on I are quantum computationally witnessable if there are polynomials s,k,p such that for all $x \in I$, Φ_x^0 and Φ_x^1 are $(s(|x|),k(|x|),p(|x|))$-witnessable. Ensembles $\{\Phi_x^0\}$ and $\{\Phi_x^1\}$ on I are quantum computationally unwitnessable if for all polynomials s,k,p and all but finitely many $x \in I$, Φ_x^0 and Φ_x^1 are $(s(|x|),k(|x|),p(|x|))$-unwitnessable.*

2.3 Quantum Commitments

Definition 2.18. *A quantum commitment scheme (resp. with quantum advice) is an interactive protocol $Com = (S,R)$ with the following properties*

- *The sender S and the receiver R have common input a security parameter 1^n (resp. both S and R have a copy of a quantum state $|\phi\rangle$ of $\mathrm{poly}(n)$ qubits). The sender has private input the bit $b \in \{0,1\}$ to be committed. Both S and R are quantum algorithms that run in time $\mathrm{poly}(n)$ that may exchange quantum messages.*
- *In the* commit *phase, S interacts with R in order to commit to b.*
- *In the* reveal *phase, S interacts with R in order to reveal b. R decides to accept or reject depending on the revealed value of b and his final state. We say that S reveals b, if R accepts the revealed value. In the honest case, R always accepts.*

A commitment scheme is non-interactive *if the commit and the reveal phase each consist of a single message from S to R. When the commit phase is non-interactive, we call ρ_S^b the state sent by the honest sender during the commit phase when his bit is b.*

Definition 2.19. *A* non-interactive auxiliary-input quantum commitment scheme (with quantum advice) on *I is a collection of non-interactive quantum commitment schemes (with advice) $\mathcal{C} = \{Com_x = (S_x, R_x)\}_{x\in I}$ such that*

- *there exists a quantum circuit Q of size polynomial in $|x|$, that given as input x for any $x \in I$, can apply the same maps that S_x and R_x apply during the commitment scheme in time polynomial in $|x|$.*
- *(statistically/computationally hiding) the two auxiliary-input state ensembles sent by the honest sender when committing to 0 or 1, which are given by $\{\rho_{S_x}^0\}_{x\in I}$ and $\{\rho_{S_x}^1\}_{x\in I}$, are quantum statistically/computationally indistinguishable.*
- *(statistically/computationally binding) for all but finitely many $x \in I$, for all polynomial p and for any unbounded/polynomial dishonest senders $S_{x,0}^*$, $S_{x,1}^*$ that send the same state in the commit phase*

$$P_{S_x^*} = \frac{1}{2}\left(\Pr[S_{x,0}^* \text{ reveals } b=0] + \Pr[S_{x,1}^* \text{ reveals } b=1]\right) \leq \frac{1}{2} + \frac{1}{p(|x|)}$$

When referring to a commitment scheme, we will use the (b_s, h_c) and (b_c, h_s) to denote schemes that are statistically binding and computationally hiding and schemes that are computationally binding and statistically hiding, respectively.

At a high level, the distinction between the two notions, without or with quantum advice, is the following. We can assume that the two players decide to perform a commitment scheme and agree on a security parameter n. Then, in the first case, a trusted party can give them the description of the circuits (C_0, C_1) so that the players can perform the commitment scheme themselves. One can think of the string (C_0, C_1) as classical advice to the players. In the second case, the trusted party gives them the description of the circuits, as well as one copy of a quantum state each. This quantum state is of polynomial size, however it is not efficiently constructible, otherwise the trusted party could have given the players the classical description of the circuit that constructs it. Hence, in the second notion the players receive both classical and quantum advice.

3 Quantum Commitments Unless $\mathsf{QSZK} \subseteq \mathsf{QMA}$

The idea of the proof is to start from pairs of circuits (C_0, C_1) which are in QSD_Y which means that their mixed state outputs ρ^{C_0} and ρ^{C_1} are statistically far from each other. We want to use ρ^{C_b} as a commitment state for the bit b. Since the states are statistically far away, such a commitment will be statistically binding. For the hiding property, we distinguish two cases. If the Receiver can distinguish in polynomial time (with some quantum auxiliary input) the two states for all but finitely many such pairs of circuits then we show that $\mathsf{QSZK} \subseteq \mathsf{QMA}$. If the Receiver cannot distinguish the two states for an infinite set I of pairs of circuits, we show how to construct a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme on I. More formally:

Theorem 1.1. *If* $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$*, then there exists a non-interactive auxiliary-input quantum* (b_s, h_c)*-commitment scheme on an infinite set* I*.*

Proof. First, we show the following

Lemma 3.1. *If* $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$ *then there exist two auxiliary-input state ensembles that are quantum computationally indistinguishable on an infinite set* I*.*

Proof. Let us consider the complete problem $\mathrm{QSD} = \{\mathrm{QSD}_Y, \mathrm{QSD}_N\}$ for $\mathsf{QSZK}_{\mathsf{HV}}$. We may restrict attention to the honest verifier case, since it is known that $\mathsf{QSZK} = \mathsf{QSZK}_{\mathsf{HV}}$ [37]. Let $n = |(C_0, C_1)|$ and define $|\phi_{C_b}\rangle = C_b(|0\rangle)$ in the space $\mathcal{O} \otimes \mathcal{G}$ to be the entire output state of the circuit on input $|0\rangle$ and $\rho^{C_b}_{(C_0,C_1)} = \mathrm{Tr}_{\mathcal{G}}(|\phi_{C_b}\rangle\langle\phi_{C_b}|)$ be the output of circuit C_b on $m(n)$ qubits for a polynomial m.

Recall that the set QSD_Y consists of pairs of circuits (C_0, C_1), such that the trace norm satisfies $\left\|\rho^{C_0}_{(C_0,C_1)} - \rho^{C_1}_{(C_0,C_1)}\right\|_{\mathrm{tr}} \geq 2 - \mu(n)$. We now consider the two auxiliary-input state ensembles $\{\rho^{C_0}_{(C_0,C_1)}\}$ and $\{\rho^{C_1}_{(C_0,C_1)}\}$ for $(C_0, C_1) \in \mathrm{QSD}_Y$. Assume for contradiction that they are quantum computationally distinguishable on QSD_Y, i.e. for some polynomials p, s, k and for all $(C_0, C_1) \in \mathrm{QSD}_Y$, the states $\rho^{C_0}_{(C_0,C_1)}$ and $\rho^{C_1}_{(C_0,C_1)}$ are $(s(n), k(n), 1/p(n))$-distinguishable. In other words, for polynomials p, s, k and for all $(C_0, C_1) \in \mathrm{QSD}_Y$ there exists a state σ on $k(n)$ qubits and a quantum circuit Q of size $s(n)$ that performs a two-outcome measurement on $m(n) + k(n)$ qubits, such that

$$|\Pr[Q(\rho^{C_0}_{(C_0,C_1)} \otimes \sigma) = 1] - \Pr[Q(\rho^{C_1}_{(C_0,C_1)} \otimes \sigma) = 1]| \geq \frac{1}{p(n)}.$$

We now claim that this implies that $\mathsf{QSZK} \subseteq \mathsf{QMA}$, which is a contradiction. For any input (C_0, C_1) the prover can send the classical polynomial size description of Q to the verifier as well as the mixed state σ with polynomial number of qubits. Then, for all $(C_0, C_1) \in \mathrm{QSD}_Y$, the verifier with the help of Q and σ can distinguish between the two circuits with probability at least $1/2 + 1/(2p(n))$. On the other hand, for all $(C_0, C_1) \in \mathrm{QSD}_N$, no matter what Q and σ the prover sends, since $\left\| \rho^{C_0}_{(C_0,C_1)} - \rho^{C_1}_{(C_0,C_1)} \right\|_{\mathrm{tr}} \leq \mu(n)$ the verifier can only distinguish the two circuits with probability at most $1/2 + \mu(n)/2$. This implies that there is an inverse polynomial gap between the acceptance probabilities in the two cases. By applying standard error reduction tools for QMA [19, 22], we obtain a QMA protocol to solve QSD.

This implies that if $\mathsf{QSZK} \not\subseteq \mathsf{QCMA}$ then there exists a non empty set $I \subseteq \mathrm{QSD}_Y$ such that the two auxiliary-input state ensembles $\{\rho^{C_0}_{(C_0,C_1)}\}$ and $\{\rho^{C_1}_{(C_0,C_1)}\}$ are quantum computationally indistinguishable on I. Notice that we may take the set I to be infinite, since if I is finite, then by hard-wiring this finite number of instances into the QMA verifier (who always accepts these instances), we have again that $\mathsf{QSZK} \subseteq \mathsf{QMA}$. □

We now show how to construct a commitment scheme from these ensembles.

Lemma 3.2. *The two auxiliary-input state ensembles given by* $\{\rho^{C_0}_{(C_0,C_1)}\}_{(C_0,C_1)\in I}$ *and* $\{\rho^{C_1}_{(C_0,C_1)}\}_{(C_0,C_1)\in I}$ *that are computationally indistinguishable on the infinite set* I *imply a non-interactive auxiliary-input quantum* (b_s, h_c)*-commitment scheme on* I*.*

Proof. For each $(C_0, C_1) \in I$ we define a scheme with security parameter $n = |(C_0, C_1)|$.

- Commit phase: To commit to bit b, the sender S runs the quantum circuit C_b with input $|0\rangle$ to create $|\phi_{C_b}\rangle = C_b(|0\rangle)$ and sends $\rho^{C_b}_{(C_0,C_1)}$ to the receiver R, which is the portion of $|\phi_{C_b}\rangle$ in the space $\mathcal{O}$.

- Reveal phase: To reveal bit b, the sender S sends the remaining qubits of the state $|\phi_{C_b}\rangle$ to the receiver R, which lie in the space $\mathcal{G}$ (the honest sender sends $|\phi'\rangle = C_b|0\rangle$). The receiver applies the circuit $C_b^\dagger$ on his entire state and then measures all his qubits in the computational basis. He accepts if and only if the outcome is $|0\rangle$.

Note that all operations of the sender and the receiver in the above protocol can be computed in time polynomial in n given the input (C_0, C_1), including the receiver's test during the reveal phase. The protocol is computationally hiding since $\{\rho^{C_0}_{(C_0,C_1)}\}$ and $\{\rho^{C_1}_{(C_0,C_1)}\}$ are quantum computationally indistinguishable.

The fact that the protocol is statistically binding follows from the fact that for the states $\{\rho^{C_0}_{(C_0,C_1)}\}$ and $\{\rho^{C_1}_{(C_0,C_1)}\}$ (for $(C_0, C_1) \in I \subseteq \mathrm{QSD}_Y$) we know that $\|\rho^{C_0}_{(C_0,C_1)} - \rho^{C_1}_{(C_0,C_1)}\|_{\mathrm{tr}} \geq 2 - \mu(n)$, for a negligible function μ. More precisely, if ξ is the total quantum state sent by a dishonest sender S^* in the commit and reveal phases of the protocol, then the probability that ξ can be revealed as the bit b is

$$\Pr[S^* \text{ reveals } b \text{ from } \xi] = \mathrm{tr}(|0\rangle\langle 0| C_b^\dagger \xi C_b) = \mathrm{F}(C_b|0\rangle, \xi)^2 \leq \mathrm{F}(\rho^{C_b}_{(C_0,C_1)}, \mathrm{tr}_{\mathcal{G}}\, \xi)^2$$

using the monotonicity of the fidelity with respect to the partial trace. This calculation follows the proof of Watrous that QSZK is closed under complementation [35]. In what follows we consider a dishonest sender that, after the commit phase, sends one of two different states in the reveal phase, so the state held by the Receiver is either ξ_0 or ξ_1. Notice that in either case the Sender sends the

same state in the commit phase, so that we have $\mathrm{tr}_{\mathcal{G}}\,\xi_0 = \mathrm{tr}_{\mathcal{G}}\,\xi_1 = \gamma$ for some $\gamma \in \mathbf{D}(\mathcal{O})$. Using this, as well as the previous equation and properties of the fidelity

$$\begin{aligned} P_{S^*} &= \frac{1}{2}\left(\Pr[S^* \text{ reveals } b=0 \text{ from } \xi_0] + \Pr[S^* \text{ reveals } b=1 \text{ from } \xi_1]\right) \\ &\leq \max_{\gamma\in\mathbf{D}(\mathcal{O})} \frac{1}{2}\left(\mathrm{F}(\rho^{C_0}_{(C_0,C_1)},\gamma)^2 + \mathrm{F}(\rho^{C_1}_{(C_0,C_1)},\gamma)^2\right) \\ &= \frac{1}{2}\left(1 + \mathrm{F}(\rho^{C_0}_{(C_0,C_1)},\rho^{C_1}_{(C_0,C_1)})\right) \leq \frac{1}{2} + \frac{\sqrt{\mu(n)}}{2}. \end{aligned}$$

The final inequality follows from the relationship between the fidelity and the trace norm as well as the fact that $\|\rho^{C_0}_{(C_0,C_1)} - \rho^{C_1}_{(C_0,C_1)}\|_{\mathrm{tr}} \geq 2 - \mu(n)$. This implies that the protocol is statistically binding. □

By combining the above Lemmas: if $\mathsf{QSZK} \not\subseteq \mathsf{QMA}$, then there exists a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme on an infinite set I. □

If we are willing to relax the indistinguishability condition, i.e. enforce the indistinguishability against a quantum algorithm that has only classical auxiliary input (i.e. get rid of σ in Definition 2.11), then the condition becomes $\mathsf{QSZK} \not\subseteq \mathsf{QCMA}$. In Section 6 we give oracle evidence that this this condition is true. Notice also that the result of Crépeau, Légaré, and Salvail [7] allows this commitment scheme to be used as a subroutine to construct a scheme that is statistically hiding and computationally binding.

4 Quantum (b_s, h_c)-commitments unless $\mathsf{QIP} \subseteq \mathsf{QMA}$

First, let us note that $\mathsf{QIP} \subseteq \mathsf{QMA}$ implies that $\mathsf{PSPACE} \subseteq \mathsf{PP}$ which is widely believed not to be true. Hence, the commitments we exhibit are based on a very weak assumption. Using this weaker assumption, we obtain a weaker commitment scheme, in the sense that it requires quantum advice. Note that our definitions of security are against quantum adversaries that also receive arbitrary quantum advice, hence our honest players are never more powerful than the dishonest ones. Moreover, the quantum advice does not create entanglement between the two players.

In our first construction, we start from pairs of circuits (Q_0, Q_1) in QCD_Y which means that there is a common input $|\phi^*\rangle$ such that their outputs ρ^{Q_0} and ρ^{Q_1} are statistically far from each other. We use ρ^{Q_b} as a commitment state for b. The quantum advice needed for the commitment is the following: the Sender receives a copy of $|\phi^*\rangle$ to create the states ρ^{Q_0} and ρ^{Q_1} and the Receiver also gets a copy of $|\phi^*\rangle$ to check via a SWAP test that the Sender did not cheat. Using the fact that the states are statistically far apart and a parallel repetition theorem for our swap-test based protocol we obtain negligible binding error. Similarly to the QSZK construction, we show that if QCD cannot be solved in QMA then our scheme is also computationally hiding.

The remainder of this section provides the proof of this result. As a first step, we give a scheme with constant binding error based on the swap test (see [6] for an exposition of the swap test). Following this result, we prove a parallel repetition theorem for non-interactive swap-test based protocols, which we then use to obtain a scheme with negligible error.

Proposition 4.1. *If $\mathsf{QIP} \not\subseteq \mathsf{QMA}$, then there exists a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme with quantum advice on an infinite set I. This scheme has constant binding error.*

Proof. We first show the following

Lemma 4.2. *If* $\mathsf{QIP} \not\subseteq \mathsf{QMA}$*, there exist two auxiliary-input superoperator ensembles* $\{Q^0\}_{(Q^0,Q^1)\in I}$ *and* $\{Q^1\}_{(Q^0,Q^1)\in I}$ *that are quantum computationally indistinguishable on an infinite set* I*.*

Proof. Suppose $\mathsf{QIP} \not\subseteq \mathsf{QMA}$. Let us consider the complete problem QCD for QIP with input the mixed-state circuits (Q^0, Q^1). Let $n = |(Q^0, Q^1)|$. Let $\mathcal{I}$ denote the input space, $\mathcal{O}$ the output space and $\mathcal{G}$ the output garbage space of the circuits Q^0, Q^1.

Consider the set QCD_Y, whose elements are pairs of circuits (Q^0, Q^1), such that the diamond norm satisfies $\left\|Q^0 - Q^1\right\|_\diamond \geq 2 - \mu(n)$, and the two auxiliary-input superoperator ensembles $\{Q^0\}_{(Q^0,Q^1)\in\mathrm{QCD}_Y}$ and $\{Q^1\}_{(Q^0,Q^1)\in\mathrm{QCD}_Y}$. Assume for contradiction that they are quantum computationally distinguishable on QCD_Y, i.e. for some polynomials p, s, k and all $(Q^0, Q^1) \in \mathrm{QSD}_Y$, the superoperators Q^0 and Q^1 are $(s(n), k(n), 1/p(n))$-distinguishable. In other words, for polynomials p, s, k and for all $(Q^0, Q^1) \in \mathrm{QSD}_Y$ there exists a mixed state σ on $t(n) + k(n)$ qubits and a quantum circuit D of size $s(n)$ that performs a two-outcome measurement on $(m(n)+k(n))$ qubits, such that

$$|\Pr[D((Q^0 \otimes \mathbb{1}_k)(\sigma)) = 1] - \Pr[D((Q^1 \otimes \mathbb{1}_k)(\sigma)) = 1]| \geq \frac{1}{p(n)}$$

We now claim that this implies that $\mathsf{QIP} \subseteq \mathsf{QMA}$, which is a contradiction. For any input (Q^0, Q^1) the QMA-prover can send to the verifier the classical polynomial size description of D as well as the mixed state σ with $\mathrm{poly}(n)$ qubits. Then, for all $(Q^0, Q^1) \in \mathrm{QCD}_Y$, the verifier with the help of D and σ can distinguish between the two circuits with probability higher than $1/2 + 1/(2p(n))$. On the other hand, for all $(Q^0, Q^1) \in \mathrm{QCD}_N$, no matter what D and σ the prover sends, since $\left\|Q^0 - Q^1\right\|_\diamond \leq \mu(n)$ the verifier can only distinguish the two circuits with probability at most $1/2 + \mu(n)/2$. Hence, there is at least an inverse polynomial gap between the two probabilities, so we can use error reduction [19, 22] to obtain a QMA protocol that solves QCD with high probability.

Thus $\mathsf{QIP} \not\subseteq \mathsf{QMA}$ implies that there exists a non-empty set $I \subseteq \mathrm{QCD}_Y$ and two auxiliary-input superoperator ensembles $\{Q^0\}_{(Q^0,Q^1)\in\mathrm{QCD}_Y}$ and $\{Q^1\}_{(Q^0,Q^1)\in\mathrm{QCD}_Y}$ which are quantum computationally indistinguishable on I. Once again, the set I must be infinite, as if I is finite then by hard-wiring this finite number of instances into the QMA verifier (who always accepts these instances), we have again that $\mathsf{QIP} \subseteq \mathsf{QMA}$. □

We now need to show how to construct a commitment scheme on I based on these indistinguishable superoperator ensembles. The protocol we obtain has only constant binding error: the average of the probability of successfully revealing 0 and the probability of successfully revealing 1 is negligibly larger than 3/4. Following this Lemma we prove a parallel repetition result for this protocol that reduces this error to a negligible function.

Lemma 4.3. *The two auxiliary-input superoperator ensembles* $\{Q^0\}_{(Q^0,Q^1)\in I}$ *and* $\{Q^1\}_{(Q^0,Q^1)\in I}$*, which are quantum computationally indistinguishable on the infinite set* $I \subseteq \mathrm{QCD}_Y$*, imply a non-interactive auxiliary-input quantum* (b_s, h_c)*-commitment scheme with quantum advice on* I*. This protocol has constant binding error.*

Proof. For every $(Q^0, Q^1) \in I$ we define a quantum commitment scheme with quantum advice. For convenience we let U^b be the unitary operation that simulates the admissible map Q^b, in other

words we have that $Q^b(\rho) = \mathrm{tr}_G\, U^b(\rho \otimes |0\rangle\langle 0|)(U^b)^\dagger$. Note that any Q^b can be efficiently converted to a unitary circuit U^b. Let also $|\phi^*\rangle$ be the pure state from Lemma 2.1, such that

$$\left\|Q^0 - Q^1\right\|_\diamond = \left\|(\mathbb{1}_{\mathbf{L}(\mathcal{F})} \otimes (Q^0 - Q^1))(|\phi^*\rangle\langle\phi^*|)\right\|_{\mathrm{tr}}.$$

- Define $n = |(Q^0, Q^1)|$ to be the security parameter. S and R also receive as advice a copy of the state $|\phi^*\rangle$ on poly(n) qubits.
- Commit phase: To commit to bit b, the sender S runs the quantum circuit $\mathbb{1}_{\mathcal{F}} \otimes U^b$ with input $|\phi^*\rangle|0\rangle$. The entire output of the circuit is a state in the space $\mathcal{F} \otimes \mathcal{O} \otimes \mathcal{G}$. The sender then sends the qubits in the space $\mathcal{O} \otimes \mathcal{F}$ to the receiver R.
- Reveal phase: To reveal bit b, the sender S sends the remaining qubits of the state $(\mathbb{1}_{\mathcal{F}} \otimes U^b)(|\phi^*\rangle|0\rangle)$ in the space $\mathcal{G}$ to the receiver R. The receiver first applies the operation $\mathbb{1}_{\mathcal{F}} \otimes (U^b)^\dagger$ to the entire state he received from the sender and then performs a swap test between this state and his copy of $|\phi^*\rangle|0\rangle$.

Let us analyze the above scheme. First, note that all operations of the sender and the receiver in the above protocol can be computed in time polynomial in n given the input (Q^0, Q^1). This includes the receiver's test during the reveal phase, since given a description of a unitary circuit it can be inverted by simply taking the inverse of each gate and running the circuit in reverse and the swap test is also efficient.

The protocol is computationally hiding since the superoperators Q^0 and Q^1 are quantum computationally indistinguishable.

The fact that the protocol is statistically binding (with constant error) follows from the fact that we have $\left\|Q^0 - Q^1\right\|_\diamond \geq 2 - \mu(n)$ for a negligible function μ. More precisely, let σ^b be the state sent by the sender with $\mathrm{tr}_{\mathcal{G}}\, \sigma^0 = \mathrm{tr}_{\mathcal{G}}\, \sigma^1 = \sigma_{\mathcal{OF}}$ (the honest sender sends the pure state $(\mathbb{1}_{\mathcal{F}} \otimes U^b)(|\phi^*\rangle|0\rangle)$). Then the receiver accepts if and only if the output of $(\mathbb{1}_{\mathcal{F}} \otimes (U^b)^\dagger)\sigma^b(\mathbb{1}_{\mathcal{F}} \otimes U_b)$ and his copy of $|\phi^*\rangle|0\rangle$ pass the swap test. This probability is equal to

$$\begin{aligned}
\Pr[S^* \text{ reveals } b \text{ from } \sigma^b] &= \frac{1}{2} + \frac{1}{2}\,\mathrm{tr}\left[(|\phi^*\rangle\langle\phi^*| \otimes |0\rangle\langle 0|)(\mathbb{1}_{\mathcal{F}} \otimes (U^b)^\dagger)\sigma^b(\mathbb{1}_{\mathcal{F}} \otimes U_b)\right] \\
&= \frac{1}{2} + \frac{1}{2}\,\mathrm{F}((\mathbb{1}_{\mathcal{F}} \otimes U_b)(|\phi^*\rangle\langle\phi^*| \otimes |0\rangle\langle 0|)(\mathbb{1}_{\mathcal{F}} \otimes (U^b)^\dagger), \sigma^b)^2 \\
&\leq \frac{1}{2} + \frac{1}{2}\,\mathrm{F}(\mathbb{1}_{\mathcal{F}} \otimes Q^b(|\phi^*\rangle\langle\phi^*|), \mathrm{tr}_{\mathcal{G}}\, \sigma^b)^2 \\
&\leq \frac{1}{2} + \frac{1}{2}\,\mathrm{F}(\mathbb{1}_{\mathcal{F}} \otimes Q^b(|\phi^*\rangle\langle\phi^*|), \sigma_{\mathcal{OF}})^2
\end{aligned}$$

where we have used the fact that the swap test on a state $\rho \otimes \sigma$ returns the symmetric outcome with probability $\frac{1}{2} + \frac{1}{2}\,\mathrm{tr}\,\rho\sigma$, as well as the monotonicity of the fidelity with respect to the partial trace.

Using this calculation, the binding property of the protocol is given by

$$\begin{aligned} P_{S^*} &= \frac{1}{2}\left(\Pr[S^* \text{ reveals } b=0] + \Pr[S^* \text{ reveals } b=1]\right) \\ &\leq \frac{1}{2} + \frac{1}{4}\left(\mathrm{F}(\mathbb{1}_{\mathcal{F}} \otimes Q^0(|\phi^*\rangle\langle\phi^*|), \mathrm{tr}_{\mathcal{G}}\,\sigma)^2 + \mathrm{F}(\mathbb{1}_{\mathcal{F}} \otimes Q^1(|\phi^*\rangle\langle\phi^*|), \mathrm{tr}_{\mathcal{G}}\,\sigma)^2\right) \\ &\leq \frac{1}{2} + \frac{1}{4}\left(1 + \mathrm{F}(\mathbb{1}_{\mathcal{F}} \otimes Q^0(|\phi^*\rangle\langle\phi^*|), \mathbb{1}_{\mathcal{F}} \otimes Q^1(|\phi^*\rangle\langle\phi^*|))\right) \\ &\leq \frac{3}{4} + \frac{\sqrt{\mu(n)}}{4}, \end{aligned}$$

where we have used Lemma 2.1 and Lemma 2.4. □

From the above two Lemmata, we have that if $\mathsf{QIP} \not\subseteq \mathsf{QMA}$, then there exists a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme with quantum advice on an infinite set I, with constant binding error. □

In the remainder of this section we show how to reduce the cheating probability of the sender to $1/2 + \mathrm{neg}(n)$. To do this, we will use parallel repetition of the above protocol.

Proposition 4.4. *Consider a k-fold repetition of the above bit commitment protocol. This is a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme with quantum advice on I.*

Proof. The two things we have to make sure of is that the computationally hiding property remains under parallel repetition and that the cheating probability of the sender decreases as a negligible function in k. To show that the protocol is computationally hiding, we use the following Lemma.

Lemma 4.5 ([37])**.** *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.*

From this Lemma, we easily have that if the superoperators Q_0 and Q_1 are quantum computationally indistinguishable then the output states of the superoperators $Q_0^{\otimes k}$ and $Q_1^{\otimes k}$ applied to any product state are quantum computationally indistinguishable for any k of polynomial size. This proves that the repeated protocol remains computationally hiding, since the honest Sender prepares a product state.

We now need to prove that the statistical binding property decreases to $1/2 + \mathrm{neg}(n)$. We first prove the following Lemma that applies to the ideal case, i.e. the Receiver applies the swap test to one of two states with orthogonal reduced states. The calculation that this strategy (approximately) generalizes to the case of states that are *almost* orthogonal states follows the proof of the Lemma.

Lemma 4.6. *Let $|\phi_0\rangle, |\phi_1\rangle \in \mathcal{A} \otimes \mathcal{B}$ be states such that $\mathrm{tr}_{\mathcal{B}} |\phi_0\rangle\langle\phi_0|$ and $\mathrm{tr}_{\mathcal{B}} |\phi_1\rangle\langle\phi_1|$ are orthogonal, and let ρ_0, ρ_1 be two states on $(\mathcal{A} \otimes \mathcal{B})^{\otimes k} = \mathcal{A}_1 \otimes \mathcal{B}_1 \otimes \cdots \otimes \mathcal{A}_k \otimes \mathcal{B}_k$ such that*

$$\mathrm{tr}_{\mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_k} \rho_0 = \mathrm{tr}_{\mathcal{B}_1 \otimes \cdots \otimes \mathcal{B}_k} \rho_1.$$

Consider the following test:

> *Test b: Take k copies of $|\phi_b\rangle$ and apply for each $i \in \{1, \ldots, k\}$ the swap test between each copy and the state in $\mathcal{A}_i \otimes \mathcal{B}_i$. Accept if all the swap tests accept.*

For any ρ_0 and ρ_1 with equal reduced states on $\mathcal{A}_1 \otimes \cdots \otimes \mathcal{A}_k$, we have

$$\frac{1}{2}\left(\Pr[\rho_0 \textit{ passes Test } 0] + \Pr[\rho_1 \textit{ passes Test } 1]\right) \leq \frac{1}{2} + \frac{1}{2^{k+1}}$$

Proof. We prove the result by induction on k. For $k = 1$. We have

$$\begin{aligned}\Pr[\rho_b \text{ passes Test } b] &= 1/2 + \langle\phi_b|\rho_b|\phi_b\rangle/2 \\ &= 1/2 + \mathrm{F}(|\phi_b\rangle\langle\phi_b|, \rho_b)^2/2 \\ &\leq 1/2 + \mathrm{F}(\mathrm{tr}_{\mathcal{B}}\,|\phi_b\rangle\langle\phi_b|, \mathrm{tr}_{\mathcal{B}}\,\rho_b)^2/2.\end{aligned}$$

Since $\mathrm{tr}_{\mathcal{B}}\,\rho_0 = \mathrm{tr}_{\mathcal{B}}\,\rho_1$, this implies that

$$\begin{aligned}&\frac{1}{2}\left(\Pr[\rho_0 \text{ passes Test } 0] + \Pr[\rho_1 \text{ passes Test } 1]\right) \\ &\quad\leq \frac{1}{2} + \frac{1}{4}(\mathrm{F}(\mathrm{tr}_{\mathcal{B}}\,|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{B}}\,\rho_0)^2 + \mathrm{F}(\mathrm{tr}_{\mathcal{B}}\,|\phi_1\rangle\langle\phi_1|, \mathrm{tr}_{\mathcal{B}}\,\rho_1)^2) \\ &\quad\leq \frac{1}{2} + \frac{1}{4}(1 + \mathrm{F}(\mathrm{tr}_{\mathcal{B}}\,|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{B}}\,|\phi_1\rangle\langle\phi_1|)) = \frac{3}{4}\end{aligned}$$

since the reduced states of $|\phi_0\rangle, |\phi_1\rangle$ are orthogonal.

Now we suppose the Lemma is true for k and show it for $k+1$. For convenience we set $\mathcal{S}_i = \mathcal{A}_i \otimes \mathcal{B}_i$. We take a reference space $\mathcal{R}$ of sufficient size to consider purifications of ρ_0 and ρ_1. Let $\rho_b = \mathrm{tr}_{\mathcal{R}}\,|\psi_b\rangle\langle\psi_b|$ be these (arbitrary) purifications. Using this notation, we write

$$|\psi_0\rangle = \alpha_0|\phi_0\rangle_{\mathcal{S}_1}|\Omega_0\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\otimes\mathcal{R}} + \alpha_1|\phi_1\rangle_{\mathcal{S}_1}|\Omega_1\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\otimes\mathcal{R}} + \alpha_2\sum_{i=2}^{n}|\phi_i\rangle|\Omega_i\rangle \tag{1}$$

and

$$|\psi_1\rangle = \beta_0|\phi_0\rangle_{\mathcal{S}_1}|\Gamma_0\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\otimes\mathcal{R}} + \beta_1|\phi_1\rangle_{\mathcal{S}_1}|\Gamma_1\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\otimes\mathcal{R}} + \beta_2\sum_{i=2}^{n}|\phi_i\rangle|\Gamma_i\rangle \tag{2}$$

where each $|\phi_i\rangle, |\phi_j\rangle$ are orthogonal for $i \neq j$ (for $|\phi_0\rangle$ and $|\phi_1\rangle$ this follows from the fact that the reduced states on $\mathcal{A}_1$ are orthogonal). Since the goal is to pass swap tests with $|\phi_0\rangle$ and $|\phi_1\rangle$, we can easily see that we can take $\alpha_2 = \beta_2 = 0$ without loss of generality, since this state will only have larger probability of passing the tests. As one final notational convenience, let $p_i = |\alpha_i|^2$ and $q_i = |\beta_i|^2$.

Before we analyze the probability that the swap tests pass, we show that the probabilities p_0 and q_1 satisfy $p_0 + q_1 \leq 1$. By Equation (1) we have

$$\begin{aligned}p_0 = |\alpha_0|^2 &= \mathrm{tr}((|\phi_0\rangle\langle\phi_0| \otimes \mathbb{1})|\psi_0\rangle\langle\psi_0|) \\ &\leq \mathrm{F}(|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}\,|\psi_0\rangle\langle\psi_0|)^2 \\ &\leq \mathrm{F}(\mathrm{tr}_{\mathcal{B}_1}\,|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}\,|\psi_0\rangle\langle\psi_0|)^2.\end{aligned}$$

By a similar calculation, we have

$$q_1 = |\beta_1|^2 \leq \mathrm{F}(\mathrm{tr}_{\mathcal{B}_1}\,|\phi_1\rangle\langle\phi_1|, \mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}\,|\psi_1\rangle\langle\psi_1|)^2.$$

Then, using the fact that $\mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}|\psi_0\rangle\langle\psi_0| = \mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}|\psi_1\rangle\langle\psi_1|$, as well as the fact that $\mathrm{tr}_{\mathcal{B}_1}|\phi_0\rangle\langle\phi_0|$ and $\mathrm{tr}_{\mathcal{B}_1}|\phi_1\rangle\langle\phi_1|$ are orthogonal, we have

$$\begin{aligned} p_0 + q_1 &\leq \mathrm{F}(\mathrm{tr}_{\mathcal{B}_1}|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}|\psi_0\rangle\langle\psi_0|)^2 + \mathrm{F}(\mathrm{tr}_{\mathcal{B}_1}|\phi_1\rangle\langle\phi_1|, \mathrm{tr}_{\mathcal{B}_1\mathcal{S}_2\ldots\mathcal{S}_{k+1}\mathcal{R}}|\psi_1\rangle\langle\psi_1|)^2 \\ &\leq 1 + \mathrm{F}(\mathrm{tr}_{\mathcal{B}_1}|\phi_0\rangle\langle\phi_0|, \mathrm{tr}_{\mathcal{B}_1}|\phi_1\rangle\langle\phi_1|) \\ &= 1. \end{aligned} \tag{3}$$

We now analyze the probability that the swap tests pass. Consider applying test 0 on $|\psi_0\rangle$. When applying the swap test between $|\phi_0\rangle$ and $|\phi_0\rangle$, the result is the state $|0\rangle|\phi_0\rangle|\phi_0\rangle$ where the first register corresponds to the acceptance of the swap test (0 corresponds to accept). When applying the swap test between the two states $|\phi_0\rangle$ and $|\phi_1\rangle$, the result before measuring the first qubit is

$$\frac{1}{\sqrt{2}}\left(|0\rangle(|\phi_0\rangle|\phi_1\rangle + |\phi_1\rangle|\phi_0\rangle) + |1\rangle(|\phi_0\rangle|\phi_1\rangle - |\phi_1\rangle|\phi_0\rangle)\right).$$

So the swap test on the space $\mathcal{S}_1$ accepts with probability $p_0 + p_1/2$. Conditioned on this test passing, we have the state:

$$\frac{1}{\sqrt{p_0 + p_1/2}}\left[\alpha_0|\phi_0\rangle|\phi_0\rangle|\Omega_0\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\mathcal{R}} + \frac{\alpha_1}{\sqrt{2}}(|\phi_0\rangle|\phi_1\rangle + |\phi_1\rangle|\phi_0\rangle)|\Omega_1\rangle_{\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\mathcal{R}}\right]$$

Discarding the first system results in the state in $\mathcal{S}_2 \otimes \cdots \otimes \mathcal{S}_{k+1} \otimes \mathcal{R}$ (using orthogonality of $|\phi_0\rangle$ and $|\phi_1\rangle$) given by

$$\sigma = \frac{p_0}{p_0 + \frac{p_1}{2}}|\Omega_0\rangle\langle\Omega_0| + \frac{\frac{p_1}{2}}{p_0 + \frac{p_1}{2}}|\Omega_1\rangle\langle\Omega_1|$$

Let $T_0(\xi)$ be the probability that a state $\xi \in \mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}\otimes\mathcal{R}$ passes all swap tests in $\mathcal{S}_2\otimes\cdots\otimes\mathcal{S}_{k+1}$ with $|\phi_0\rangle$. We include the space $\mathcal{R}$ for convenience only: notice that the choice of purification in the space $\mathcal{R}$ has no effect on this probability. Using this notation, we have

$$\begin{aligned} \Pr[\rho_0 \text{ passes Test 0}] &= (p_0 + \frac{p_1}{2}) \cdot \left(\frac{p_0}{p_0 + \frac{p_1}{2}}T_0(|\Omega_0\rangle\langle\Omega_0|) + \frac{\frac{p_1}{2}}{p_0 + \frac{p_1}{2}}T_0(|\Omega_1\rangle\langle\Omega_1|)\right) \\ &= p_0 T_0(|\Omega_0\rangle\langle\Omega_0|) + \frac{p_1}{2}T_0(|\Omega_1\rangle\langle\Omega_1|) \end{aligned}$$

Similarly, we define $T_1(\xi)$ for any ξ and we have

$$\Pr[\rho_1 \text{ passes Test 1}] = \frac{q_0}{2}T_1(|\Gamma_0\rangle\langle\Gamma_0|) + q_1 T_1(|\Gamma_1\rangle\langle\Gamma_1|)$$

which gives us

$$\begin{aligned} P &= \frac{1}{2}\left(\Pr[\rho_0 \text{ passes Test 0}] + \Pr[\rho_1 \text{ passes Test 1}]\right) \\ &= \frac{1}{2}\left(p_0 T_0(|\Omega_0\rangle\langle\Omega_0|) + \frac{p_1}{2}T_0(|\Omega_1\rangle\langle\Omega_1|) + \frac{q_0}{2}T_1(|\Gamma_0\rangle\langle\Gamma_0|) + q_1 T_1(|\Omega_1\rangle\langle\Omega_1|)\right) \end{aligned} \tag{4}$$

Consider the states $\xi_0 = p_0|\Omega_0\rangle\langle\Omega_0| + p_1|\Omega_1\rangle\langle\Omega_1|$ and $\xi_1 = q_0|\Gamma_0\rangle\langle\Gamma_0| + q_1|\Gamma_1\rangle\langle\Gamma_1|$. These states are obtained from ρ_0 and ρ_1 by discarding the system in $\mathcal{S}_1$. This implies that they have the properties

in the statement of the Lemma, i.e. the reduced states of ξ_0 and x_1 on $\mathcal{A}_2 \otimes \cdots \otimes \mathcal{A}_{k+1}$ are equal. Thus, by induction, we know that $\frac{1}{2}\left(T_0(\xi_0) + T_1(\xi_1)\right) \leq \frac{1}{2} + \frac{1}{2^{k+1}}$. This means that:

$$\frac{1}{2}\left(p_0 T_0(|\Omega_0\rangle\langle\Omega_0|) + p_1 T_0(|\Omega_1\rangle\langle\Omega_1|) + q_0 T_1(|\Gamma_0\rangle\langle\Gamma_0|) + q_1 T_1(|\Gamma_1\rangle\langle\Gamma_1|)\right) \leq \frac{1}{2} + \frac{1}{2^{k+1}}$$

Using this, as well as Equation (4), we have

$$\begin{aligned} P &= \frac{1}{2}\left(p_0 T_0(|\Omega_0\rangle\langle\Omega_0|) + \frac{p_1}{2} T_0(|\Omega_1\rangle\langle\Omega_1|) + \frac{q_0}{2} T_1(|\Gamma_0\rangle\langle\Gamma_0|) + q_1 T_1(|\Gamma_1\rangle\langle\Gamma_1|)\right) \\ &= \frac{1}{4} + \frac{1}{2^{k+2}} + \frac{p_0}{4} T_0(|\Omega_0\rangle\langle\Omega_0|) + \frac{q_1}{4} T_1(|\Gamma_1\rangle\langle\Gamma_1|) \\ &\leq \frac{1}{2} + \frac{1}{2^{k+2}}, \end{aligned}$$

where the final inequality is by Equation (3). □

Notice that in the original bit commitment protocol the Receiver applies the swap test to $|\phi^*\rangle|0\rangle$ and the output of $(U_b^\dagger \otimes \mathbb{1})(\sigma_b)(U_b \otimes \mathbb{1})$ where σ_b is the state sent during the protocol. Since $U_b^\dagger$ is unitary, this is equivalent to applying the swap test between σ_b and the state $|\phi_b\rangle = (U_b \otimes \mathbb{1})|\phi^*\rangle|0\rangle$, for whatever value of b the Sender has revealed. Viewed in this way, the receiver applies the swap test between σ_b and one of two *almost* orthogonal states. Furthermore, these two states have the property that the reduced states on the space $\mathcal{O}$ have negligible fidelity. Notice also that the Sender may send one of two states σ_0 and σ_1 depending on the value that he wishes to reveal. Since we are interested in the sum of the probabilities that the Sender can successfully reveal both 0 and 1 in a given instance of the protocol, we may assume that the first message stays the same, i.e. that $\mathrm{tr}_{\mathcal{G}}\,\sigma_0 = \mathrm{tr}_{\mathcal{G}}\,\sigma_1$. This is exactly the condition in Lemma 4.6 with the exception that instead of the orthogonality of the states $|\phi_i\rangle$ we have only approximate orthogonality. We are able to overcome this obstacle with the following Lemma, the proof of which makes significant use of the fact that the trace norm can be written in terms of the projectors onto the positive and negative eigenspaces of a matrix. In particular, when applied to a Hermitian operator X the trace norm is given by $\mathrm{tr}(\Pi_+ X) - \mathrm{tr}(\Pi_- X)$, where Π_+ and Π_- are the projectors onto the positive and negative eigenspaces of X, respectively. This fact follows from the definition of the trace norm.

Lemma 4.7. *Let $|\phi_0\rangle, |\phi_1\rangle \in \mathcal{A} \otimes \mathcal{B}$ such that $\|\mathrm{tr}_{\mathcal{B}}\,|\phi_0\rangle\langle\phi_0| - \mathrm{tr}_{\mathcal{B}}\,|\phi_1\rangle\langle\phi_1|\|_{\mathrm{tr}} \geq 2 - \varepsilon$. Then there exist states $|\phi_0'\rangle, |\phi_1'\rangle \in \mathcal{A} \otimes \mathcal{B}$ such that*

1. *$\langle\phi_i'|\phi_i\rangle \geq 1 - \varepsilon$ for $i \in \{0, 1\}$,*
2. *$\mathrm{tr}_{\mathcal{B}}\,|\phi_0'\rangle\langle\phi_0'|$ and $\mathrm{tr}_{\mathcal{B}}\,|\phi_1'\rangle\langle\phi_1'|$ are orthogonal.*

Proof. For simplicity, let $\rho_i = \mathrm{tr}_{\mathcal{B}}\,|\phi_i\rangle\langle\phi_i|$. We have

$$2 - \varepsilon \leq \|\rho_0 - \rho_1\|_{\mathrm{tr}} = \mathrm{tr}\,|\rho_0 - \rho_1| = \mathrm{tr}\,\Pi_+(\rho_0 - \rho_1) - \mathrm{tr}\,\Pi_-(\rho_0 - \rho_1), \tag{5}$$

where Π_+ and Π_- are the projectors onto the positive and negative eigenspaces of $\rho_0 - \rho_1$ respectively. Notice that

$$\mathrm{tr}(\Pi_+\rho_0) = \mathrm{tr}(\Pi_+(\rho_0 - \rho_1)) + \mathrm{tr}(\Pi_+\rho_1) \geq \mathrm{tr}(\Pi_+(\rho_0 - \rho_1)),$$

and similarly $\operatorname{tr}(\Pi_- \rho_1) \geq -\operatorname{tr}(\Pi_-(\rho_0 - \rho_1))$, which implies that

$$\operatorname{tr}(\Pi_+ \rho_0) + \operatorname{tr}(\Pi_- \rho_1) \geq \operatorname{tr}(\Pi_+(\rho_0 - \rho_1)) - \operatorname{tr}(\Pi_-(\rho_0 - \rho_1)) \geq 2 - \varepsilon,$$

by Equation (5). This implies that $\operatorname{tr}(\Pi_+ \rho_0) \geq 1 - \varepsilon$ and $\operatorname{tr}(\Pi_- \rho_1) \geq 1 - \varepsilon$.

We introduce the states ρ_i' given by the (renormalized) projection of ρ_0 and ρ_1 into the spaces spanned by Π_+ and Π_-, respectively. Since these are orthogonal projectors the states ρ_0' and ρ_1' are orthogonal. Notice also that

$$\left\| \rho_0 - \rho_0' \right\|_{\mathrm{tr}} = \operatorname{tr} \left| \rho_0 - \rho_0' \right| = \operatorname{tr}(\Gamma_+(\rho_0 - \rho_0')) - \operatorname{tr}(\Gamma_-(\rho_0 - \rho_0')) = 2 \operatorname{tr}(\Gamma_+(\rho_0 - \rho_0')),$$

where Γ_+, Γ_- are the projectors onto the positive and negative eigenspaces of $\rho_0 - \rho_0'$, and we have also used the fact that $\operatorname{tr}(\rho_0 - \rho_0') = 0$, which implies that the positive portion of $\rho_0 - \rho_0'$ has the same trace as the negative portion. Consider the positive eigenspace of $\rho_0 - \rho_0'$. This is precisely the subspace spanned by the support of ρ_0 that lies outside the support of ρ_0', i.e. this is exactly the space spanned by the projector $\Pi_- = \Gamma_+$. Using this observation

$$\left\| \rho_0 - \rho_0' \right\|_{\mathrm{tr}} = 2 \operatorname{tr}(\Gamma_+(\rho_0 - \rho_0')) = 2 \operatorname{tr}(\Pi_- \rho_0) \leq 2\varepsilon, \tag{6}$$

where we have used the fact that $\operatorname{tr}(\Pi_- \rho_0) = 1 - \operatorname{tr}(\Pi_+ \rho_0) \leq \varepsilon$. A similar argument establishes the fact that

$$\left\| \rho_1 - \rho_1' \right\|_{\mathrm{tr}} = 2 \operatorname{tr}(\Pi_+ \rho_1) \leq 2\varepsilon. \tag{7}$$

Finally, we note that Equations (6) and (7) and Uhlmann's theorem imply that there exist purifications $|\phi_0'\rangle, |\phi_1'\rangle \in \mathcal{A} \otimes \mathcal{B}$ of ρ_0' and ρ_1' such that

$$\langle \phi_i' | \phi_i \rangle = \mathrm{F}(\rho_i', \rho_i) \geq 1 - \varepsilon.$$

This, combined with the orthogonality of ρ_0' and ρ_1', completes the proof. □

This Lemma shows that we may replace the two states that are almost orthogonal with nearby states that have exactly the orthogonality property required by Lemma 4.6, which we can in turn use to show that the protocol repeated k times is statistically binding. To do so, notice that the two states $|\phi_0\rangle$ and $|\phi_1\rangle$, which are given by applying the circuits Q_0 and Q_1 to the state $|\phi^*\rangle|0\rangle$, satisfy

$$\begin{aligned} \| |\phi_0\rangle\langle\phi_0| - |\phi_1\rangle\langle\phi_1| \|_{\mathrm{tr}} &\geq \| \operatorname{tr}_{\mathcal{G}}(|\phi_0\rangle\langle\phi_0| - |\phi_1\rangle\langle\phi_1|) \|_{\mathrm{tr}} \\ &= \| ((Q_0 - Q_1) \otimes I)(|\psi^*\rangle\langle\psi^*|) \|_{\mathrm{tr}} \\ &= \| Q_0 - Q_1 \|_{\diamond} \\ &\geq 2 - \mu(n), \end{aligned}$$

These states are not orthogonal, but are nearly so. We may, however, use Lemma 4.7 to obtain $|\phi_0'\rangle$ and $|\phi_1'\rangle$ that have the orthogonality property required by Lemma 4.6 that have inner product at least $1 - \mu(n)$ with the original states $|\phi_0\rangle$ and $|\phi_1\rangle$, respectively.

We now relate the probability that the state ρ passes our Test 0, i.e. the k swap tests with the state $|\phi_0\rangle^{\otimes k}$ to the probability that the same state ρ passes the k swap tests with the state $|\phi_0'\rangle^{\otimes k}$ (denoted by Test′ 0). The difference of these probabilities is upper bounded by the trace

distance of the difference of the states $|\phi_0\rangle^{\otimes k}$ and $|\phi_0'\rangle^{\otimes k}$, since we can view the swap test with ρ as a measurement to distinguish these two states. This gives

$$\begin{aligned} |\Pr[\rho \text{ passes Test } 0] - \Pr[\rho \text{ passes Test}' \, 0]| &\leq \left\| (|\phi_0\rangle\langle\phi_0|)^{\otimes k} - (|\phi_0'\rangle\langle\phi_0'|)^{\otimes k} \right\|_{\mathrm{tr}} \\ &= 2\sqrt{1 - |\langle\phi_0'|\phi_0\rangle|^{2k}} \\ &\leq 2\sqrt{1 - (1-\mu(n))^{2k}} \\ &\leq 2\sqrt{2k\mu(n)}, \end{aligned}$$

where the final inequality is Bernoulli's inequality. Similarly we have

$$|\Pr[\rho \text{ passes Test } 1] - \Pr[\rho \text{ passes Test}' \, 1]| \leq 2\sqrt{2k\mu(n)}$$

Hence, for the binding property of our scheme we have

$$\begin{aligned} &\frac{1}{2}\left(\Pr[\rho \text{ passes Test } 0] + \Pr[\rho \text{ passes Test } 1]\right) \\ &\quad\leq \frac{1}{2}\left(\Pr[\rho \text{ passes Test}' \, 0] + \Pr[\rho \text{ passes Test}' \, 1]\right) + 2\sqrt{2k\mu(n)} \\ &\quad\leq \frac{1}{2} + \frac{1}{2^{k+1}} + 2\sqrt{2k\mu(n)}. \end{aligned}$$

since, for the Test$'$ 0 and Test$'$ 1 we can use Lemma 4.6 for the perfect case. This quantity is negligibly larger than 1/2, as we may take k any polynomial and μ is a negligible function. □

This proposition, when combined with Proposition 4.1, gives the main result of this section.

Theorem 4.8. *If* QIP $\not\subseteq$ QMA*, then there exists a non-interactive auxiliary-input quantum (b_s, h_c)-commitment scheme with quantum advice on an infinite set I.*

5 Quantum (b_c, h_s)-commitments unless QIP $\subseteq$ QMA

To obtain protocols that are computationally binding and statistically hiding, we use instances of the QIP-complete problem Π to construct a (b_c, h_s)-commitment scheme with quantum advice under the assumption that QIP $\not\subseteq$ QMA. We start from pairs of circuits $Q_0, Q_1 \in \Pi_Y$ and the corresponding input states ρ^0, ρ^1 (see Definition 2.8) that will be given to the Sender as quantum advice. An honest Sender commits to b by sending half of ρ^b to the Receiver. By definition of ρ^0, ρ^1, the protocol is statistically hiding (in fact it is perfectly hiding). During the reveal phase, the Sender sends the second half of ρ^b. If $\Pi \notin$ QMA, we show that this protocol is also computationally binding, using our notion of computationally unwitnessable superoperators.

Theorem 5.1. *If* QIP $\not\subseteq$ QMA*, then there exists a non-interactive auxiliary-input quantum (b_c, h_s)-commitment scheme with quantum advice on an infinite set I.*

Proof. Recall the Complete problem $\Pi = \{\Pi_Y, \Pi_N\}$ from Definition 2.8 with inputs the mixed-state circuits (Q^0, Q^1) from $\mathbf{D}(\mathcal{X} \otimes \mathcal{Y})$ to a single bit and $n = |(Q^0, Q^1)|$. To show this Theorem, we use the following Lemma, the proof of which is very similar to the proof of Lemma 3.1.

Lemma 5.2. *If* QIP $\not\subseteq$ QMA*, there exist two auxiliary-input superoperator ensembles* $\{Q^0\}_{(Q^0,Q^1)\in I}$ *and* $\{Q^1\}_{(Q^0,Q^1)\in I}$ *that are quantum computationally unwitnessable on an infinite set* I*.*

Proof. Let us consider the set Π_Y and suppose for contradiction that the two auxiliary-input superoperator ensembles $\{Q^0\}_{(Q^0,Q^1)\in\Pi_Y}$ and $\{Q^1\}_{(Q^0,Q^1)\in\Pi_Y}$ are quantum computationally witnessable, i.e. there exist polynomials (s,k,p) such that for all $(Q^0,Q^1)\in\Pi_Y$ the superoperators Q^0 and Q^1 are $(s(n),k(n),p(n))$-witnessable. In other words, there exist polynomials (s,k,p) such that for all $(Q^0,Q^1)\in\Pi_Y$ there exist two input states $\rho^0,\rho^1\in\mathbf{L}(\mathcal{X}\otimes\mathcal{Y})$ such that first, there exists a state $\sigma\in\mathbf{L}(\mathcal{W}\otimes\mathcal{X}\otimes\mathcal{Y})$ with $|\mathcal{W}|=k$ and $\mathrm{tr}_{\mathcal{W}}\,\sigma=\rho_0$, and there exists an admissible superoperator $\Psi:\mathbf{L}(\mathcal{W}\otimes\mathcal{X})\to\mathbf{L}(\mathcal{X})$ of size s, such that $\rho^1=(\Psi\otimes\mathbb{1}_{\mathcal{Y}})(\sigma)$; and second

$$\frac{1}{2}\left(\Pr[Q^0(\rho^0)=1]+\Pr[Q^1(\rho^1)=1]\right)\geq\frac{1}{2}+\frac{1}{p(n)}.$$

Then, we provide a QMA protocol for the problem Π. Merlin sends σ (which is of size polynomial in the input, since $k(n)=|\mathcal{W}|$) and the classical description of Ψ (of size $s(n)$). Arthur with probability $1/2$ applies Q^0 on ρ^0 (which he obtains from σ by discarding the space $\mathcal{W}$) and accepts if he gets 1; and with probability $1/2$ he first creates ρ^1 from Ψ and σ, then applies Q^1 on it and also accepts if he gets 1.
(*Completeness*) If $(Q^0,Q^1)\in\Pi_Y$, we have

$$\Pr[\text{Arthur accepts}]=\frac{1}{2}\left(\Pr[Q^0(\rho^0)=1]+\Pr[Q^1(\rho^1)=1]\right)\geq\frac{1}{2}+\frac{1}{p(n)}$$

(*Soundness*) If $(Q^0,Q^1)\in\Pi_N$, then for any cheating Merlin, Arthur receives a state ρ^0_*, from which he constructs (with half probability) a state ρ^1_* each in space $\mathcal{X}\otimes\mathcal{Y}$ such that $\mathrm{tr}_{\mathcal{X}}\,\rho^0_*=\mathrm{tr}_{\mathcal{X}}\,\rho^1_*$. By the definition of Π_N, we have

$$\Pr[\text{Arthur accepts}]=\frac{1}{2}\left(\Pr[Q^0(\rho^0_*)=1]+\Pr[Q^1(\rho^1_*)=1]\right)\leq\frac{1}{2}+\mu(n)$$

We have an inverse polynomial gap between completeness and soundness and hence we conclude that $\Pi\in$ QMA. This proves that there is a nonempty I that satisfies the property of our Lemma. Note that if I is finite, then by hard-wiring this finite number of instances into the QMA verifier (who always accepts these instances), we have again that QIP $\subseteq$ QMA. So if QIP $\not\subseteq$ QMA then the set I can be taken to be infinite. □

To finish the proof of the Theorem, we now need to show the following.

Lemma 5.3. *Auxiliary-input superoperator ensembles* $\{Q^0\}_{(Q^0,Q^1)\in I}$ *and* $\{Q^1\}_{(Q^0,Q^1)\in I}$ *that are quantum computationally unwitnessable on an infinite set* $I\subseteq\Pi_Y$ *imply a non-interactive quantum* (b_c,h_s)*-commitment scheme with quantum advice on* I*.*

Proof. Commitment scheme Each $(Q^0,Q^1)\in I\subseteq\Pi_Y$ gives the following scheme

- Let $n=|(Q^0,Q^1)|$ be the security parameter. The sender receives as advice $\rho^0,\rho^1\in\mathcal{X}^i\otimes\mathcal{Y}^i$ such that $\mathrm{tr}_{\mathcal{X}}\,\rho^0=\mathrm{tr}_{\mathcal{X}}\,\rho^1$ and $\frac{1}{2}\left(\Pr[Q^0(\rho^0)=1]+\Pr[Q^1(\rho^1)=1]\right)\geq 1-\mu(n)$. For consistency with our definitions, we also suppose that the Receiver gets a copy of ρ^0,ρ^1. These states will not be used in the honest case and they will not harm the security for a cheating Receiver.

- (Commit phase) To commit to b, the Sender sends the state in $\mathcal{Y}^b$ to the Receiver.
- (Reveal phase) To reveal b, the Sender sends the state in $\mathcal{X}^b$. The Receiver applies Q^b on the space $\mathcal{X}^b \otimes \mathcal{Y}^b$ and accepts if he gets 1.

Statistical hiding property: The states that the receiver gets in the commit phase satisfy $\mathrm{tr}_{\mathcal{X}}\, \rho^0 = \mathrm{tr}_{\mathcal{X}}\, \rho^1$ and hence our scheme is perfectly hiding.
Computationally binding property: The property follows from the fact that the two auxiliary-input superoperator ensembles $\{Q^0\}_{(Q^0,Q^1)\in I}$ and $\{Q^1\}_{(Q^0,Q^1)\in I}$ are quantum computationally unwitnessable. Fix $(Q^0, Q^1) \in I$ with $|(Q^0, Q^1)| = n$. After the reveal phase, the Receiver has ρ_*^b in space $\mathcal{X} \otimes \mathcal{Y}$, where b is the revealed bit. Since we consider dishonest senders $S^*_{(Q^0,Q^1)}$ that are quantum polynomial time machines with quantum advice, the states ρ_*^0 and ρ_*^1 satisfy property 2 of Definition 2.16. Thus, for all but finitely many $(Q^0, Q^1) \in I$ they do not have property 1 of Definition 2.16. Then, for such $(Q^0, Q^1) \in I$ we have

$$\begin{aligned} P_{S^*_{(Q^0,Q^1)}} &= \frac{1}{2}\left(\Pr[S^*_{(Q^0,Q^1)} \text{ reveals } b = 0] + \Pr[S^*_{(Q^0,Q^1)} \text{ reveals } b = 1]\right) \\ &= \frac{1}{2}\left(\Pr[Q_0(\rho_*^0) = 1] + \Pr[Q_1(\rho_*^1) = 1]\right) \leq \frac{1}{2} + \frac{1}{p(n)} \end{aligned}$$

for all polynomials p □

From the above two Lemmas, unless $\mathsf{QIP} \subseteq \mathsf{QMA}$ there exists a non-interactive auxiliary-input quantum (b_c, h_s)-commitment scheme with quantum advice on infinite set I. □

This result, combined with Theorem 4.8 completes the proof of Theorem 1.3.

6 Quantum Oracle Relative to Which $\mathsf{QSZK_{HV}} \not\subseteq \mathsf{QCMA}$

In order to prove the desired result we find a problem in $\mathsf{QSZK_{HV}}$ and prove a black-box lower bound in the QCMA model. We end up with a quantum oracle, as the constructed problem makes essential use of quantum information. This approach is due to Aaronson and Kuperberg [2], who prove a similar result for QMA versus QCMA. The argument given here is related to the argument of Aaronson and Kuperberg, both in structure and in the fact that we make use of a bound on the expected overlap of a state drawn from a p-uniform distribution with a fixed state. The main difference is that in the problem we consider we need to extend the proof to the case where it is a unitary operator that is hidden inside the oracle, not a pure state. Note that subsequent to the completion of this work, Aaronson has shown the stronger result that there is an oracle relative to which $\mathsf{SZK} \not\subseteq \mathsf{QMA}$ [1].

For our result we consider a black-box that takes as input a control qubit, chooses a random pure state $|\psi\rangle$ and applies a fixed but hidden d by d unitary U to half of $|\psi\rangle$, controlled by the input qubit. The hidden unitary U can be inverted by a QSZK prover, but in the QCMA model, the Verifier cannot invert U and recover the input with making an exponential number of queries to the black-box. We prove a lower bound on the number of queries needed by a QCMA Verifier to distinguish this black-box from one that simply generates random pure states.

Theorem 1.2. *There exists a quantum oracle A such that $\mathsf{QSZK}^A_{\mathsf{HV}} \not\subseteq \mathsf{QCMA}^A$.*

6.1 Background

Before proving the oracle result we review some background on measures on quantum states and channels that will be used in the proof.

Let $\mathbf{U}(\mathcal{H})$ be the group of unitary matrices acting on a Hilbert space $\mathcal{H}$. When no confusion is likely to arise, we will also use the notation $\mathbf{U}(d)$, where $\dim \mathcal{H} = d$. The set of pure states on $\mathcal{H}$, i.e. the unit sphere in $\mathcal{H}$, is given by $\mathbf{S}(\mathcal{H})$ or $\mathbf{S}^{d-1}$. We refer to d-dimensional spaces for convenience: in general $d = 2^n$ for some space of n qubits.

Throughout this section, the *uniform* measure on states and unitaries is given by the Haar measure. In the case of unitaries, we use $\mu_{\mathbf{U}(\mathcal{H})}$ to denote the Haar measure on the unitaries on $\mathcal{H}$, that is, the unique left and right invariant measure normalized so that $\mu_{\mathbf{U}(\mathcal{H})}(\mathbf{U}(\mathcal{H})) = 1$. When the space in question is clear we will drop the subscript and use only μ to refer to this measure. The Haar measure on $\mathbf{S}(\mathcal{H})$ can be obtained by applying a random $U \in \mathbf{U}(\mathcal{H})$ to a fixed pure state (the invariance of the Haar measure implies that the choice of the fixed state does not matter). We will use $\mu_{\mathbf{S}(\mathcal{H})}$ to refer to this measure.

Essential to our argument is the notion of a probability measure that is *nearly* uniform. Following Aaronson and Kuperberg [2], given a measure σ we say that it is *p-uniform* if $p\sigma \leq \mu$, where μ is the uniform measure over the space in question. This notion is directly related to the class QCMA by the fact that if the verifier starts with a uniform measure and conditions on a m-bit classical message, the result is a (2^{-m})-uniform measure. The main technical result of this section will be to show that such a measure over $\mathbf{U}(d)$ does not help the verifier identify a particular unitary, unless $m \in \Omega(d)$. This result follows by a reduction to the pure state case, which is the key to the quantum oracle that separates QMA and QCMA [2].

Before doing this, we highlight two straightforward properties of p-uniform measures on $\mathbf{U}(d)$ and $\mathbf{S}^{d-1}$.

Proposition 6.1. *Let σ be a p-uniform measure on $\mathbf{U}(d)$.*

1. *For any $U \in \mathbf{U}(d)$ the measure $U\sigma$ remains p-uniform.*
2. *For any $|\psi\rangle \in \mathbf{S}^{d-1}$, the measure τ on $\mathbf{S}^{d-1}$ given by*

$$\tau(A) = \sigma(\{U : U|\psi\rangle \in A\})$$

is p-uniform.

Proof. The left-invariance of $\mu_{\mathbf{U}(d)}$ gives the first property, since for any $A \subseteq \mathbf{U}(d)$,

$$p(U\sigma)(A) = p\sigma(U^\dagger A) \leq \mu(U^\dagger A) = \mu(A).$$

The second property follows from the definition of $\mu_{\mathbf{S}^{d-1}}$,

$$p\tau(A) = p\sigma(\{U : U|\psi\rangle \in A\}) \leq \mu_{\mathbf{U}(d)}(\{U : U|\psi\rangle \in A\}) = \mu_{\mathbf{S}^{d-1}}(A).$$

where right-invariance of $\mu_{\mathbf{U}(d)}$ implies that the choice of $|\psi\rangle$ does not matter. □

6.2 Oracle Separation

We now define our problem.

Problem 6.2. *Given a quantum oracle $O\colon \mathcal{A} \to \mathcal{A} \otimes \mathcal{H} \otimes \mathcal{K}$, where $\dim \mathcal{H} = \dim \mathcal{K} = d$ and $\dim \mathcal{A} = 2$. The problem is to decide between the two cases*

1. *there exists a unitary $U \in \mathbf{U}(\mathcal{H})$ such that the oracle O performs the map*

$$\alpha|0\rangle + \beta|1\rangle \mapsto \frac{1}{d^2}\Big(|\alpha|^2 |0\rangle\langle 0| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + \alpha\bar{\beta}|0\rangle\langle 1| \otimes U^\dagger \otimes \mathbb{1}_{\mathcal{K}} + \bar{\alpha}\beta|1\rangle\langle 0| \otimes U \otimes \mathbb{1}_{\mathcal{K}} + |\beta|^2 |1\rangle\langle 1| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}\Big).$$

   This map can be implemented in the following way: the oracle chooses a pure state $|\psi\rangle \in \mathcal{H} \otimes \mathcal{K}$ from the Haar measure and then performs the map

$$\alpha|0\rangle + \beta|1\rangle \mapsto \alpha|0\rangle|\psi\rangle + \beta|1\rangle(U \otimes \mathbb{1}_{\mathcal{K}})|\psi\rangle.$$

2. *the oracle O preforms the map*

$$\alpha|0\rangle + \beta|1\rangle \mapsto \frac{1}{d^2}\left(|\alpha|^2 |0\rangle\langle 0| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + |\beta|^2 |1\rangle\langle 1| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}\right).$$

   for example by measuring the input qubit and appending the maximally mixed state.

We defined the oracles as superoperators, but one can think of them as unitaries in larger spaces. The key idea is that in the first case the coherence of the input qubit can be recovered, provided the hidden unitary U can be inverted, whereas in the second case this coherence is irretrievably lost. The prover in a QSZK protocol, given only the portion of the state in the space $\mathcal{H}$ and a copy of the input qubit, is able to apply $U^\dagger$ in order to disentangle the input space from $\mathcal{H} \otimes \mathcal{K}$. To prove a lower bound on this problem, we argue that with at most a small amount of knowledge about the hidden operator U, an oracle of the first type appears much the same as an oracle of the second type.

Before proving this lower bound, we give an interactive protocol for the problem. The idea behind the protocol is that when the input to the oracle is one half of a maximally entangled state then in the first case a prover is able to assist the verifier in recovering the original input state, but in the second case no action of the prover can recover the state.

Protocol 6.3. *Let O be the oracle in Problem 6.2.*

1. *V, prepares the state $|\phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2} \in \mathcal{B} \otimes \mathcal{A}$, and uses as input to the oracle O the portion of the state in $\mathcal{A}$. V then sends the state in $\mathcal{A} \otimes \mathcal{H}$ to P.*
2. *P applies the unitary $U^\dagger$ on $\mathcal{H}$ controlled on the qubit in $\mathcal{A}$.*
3. *V receives a state from P in the space $\mathcal{A} \otimes \mathcal{H}$ and measures the operator $|\phi^+\rangle\langle\phi^+|$ on the space $\mathcal{B} \otimes \mathcal{A}$, accepting if and only if the outcome is one.*

In the following theorem we prove the completeness and soundness of Protocol 6.3. The fact that it is also zero-knowledge is argued as part of the proof of Theorem 1.2.

Theorem 6.4. *Let V be the verifier in Protocol 6.3.*

1. *If the oracle is of type 1, there is a prover P that causes V to accept with certainty.*
2. *If the oracle is of type 2, then for any P, V accepts with probability at most 1/2.*

Proof. To prove completeness (item 1), notice that when the oracle is of type 1, the state of the verifier before sending the message to the prover is

$$\frac{1}{2d^2}\left[|00\rangle\langle 00| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + |00\rangle\langle 11| \otimes U^\dagger \otimes \mathbb{1}_{\mathcal{K}} + |11\rangle\langle 00| \otimes U \otimes \mathbb{1}_{\mathcal{K}} + |11\rangle\langle 11| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}.\right]$$

If the honest prover applies $U^\dagger$ on the space $\mathcal{H}$, controlled on the qubit in $\mathcal{A}$, the state of the verifier at the start of Step 3 is

$$\frac{1}{2d^2}\big(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|\big) \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} = |\phi^+\rangle\langle\phi^+| \otimes \frac{\mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}}{d^2}$$

and so the projective measurement on $\mathcal{A}\otimes\mathcal{B}$ given by $\{|\phi^+\rangle\langle\phi^+|, \mathbb{1} - |\phi^+\rangle\langle\phi^+|\}$ always results in the first outcome. This implies that the verifier can always be made to accept an oracle of type 1.

To prove soundness (item 2) we show that the verifier rejects an oracle of type 2 with probability at least 1/2, regardless of the strategy of the prover. In this case the state of the verifier before sending the message is given by the mixture

$$\frac{1}{2d^2}\left(|00\rangle\langle 00| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + |11\rangle\langle 11| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}\right).$$

After the prover applies an arbitrary transformation to $\mathcal{A}\otimes\mathcal{H}$, the result is

$$\frac{1}{2d}\left(|0\rangle\langle 0| \otimes \rho_0 \otimes \mathbb{1}_{\mathcal{K}} + |1\rangle\langle 1| \otimes \rho_1 \otimes \mathbb{1}_{\mathcal{K}}\right)$$

for some mixed states ρ_0, ρ_1 on $\mathcal{A}\otimes\mathcal{H}$. The probability that the verifier's measurement results in the outcome $|\phi^+\rangle\langle\phi^+|$ on this state is given by

$$\frac{1}{2d}\operatorname{tr}\left[|\phi^+\rangle\langle\phi^+|\left(|0\rangle\langle 0| \otimes \rho_0 \otimes \mathbb{1}_{\mathcal{K}} + |1\rangle\langle 1| \otimes \rho_1 \otimes \mathbb{1}_{\mathcal{K}}\right)\right] = \frac{1}{4}\left(\langle 0|\rho_0|0\rangle + \langle 1|\rho_1|1\rangle\right) \leq \frac{1}{2},$$

which implies that the verifier accepts with probability at most 1/2 when O is of type 2. In fact, the best strategy for a cheating prover is not to change the control bit in $\mathcal{A}$ at all. □

A central component of the argument that a QCMA verifier cannot identify a pure state hidden in an oracle is a geometric bound on the expected overlap between any fixed state and a state drawn from a p-uniform distribution.

Lemma 6.5 (Aaronson and Kuperberg [2])**.** *For any p-uniform measure σ on $\mathbf{S}^{d-1}$ and any state ρ*

$$\mathop{\mathbb{E}}_{|\psi\rangle\in\sigma}\left[\langle\psi|\rho|\psi\rangle\right] \in O\left(\frac{1+\log 1/p}{d}\right)$$

Our argument requires a similar geometric bound, except that we have a p-uniform measure over unitaries and not the pure states. We obtain a reduction from $\mathbf{U}(d)$ to $\mathbf{S}^{d-1}$, which allows us to extend the bound in Lemma 6.5.

Lemma 6.6. *If σ is a p-uniform measure on $\mathbf{U}(d)$, then*

$$\left\| \mathop{\mathbb{E}}_{U\in\sigma} U \right\|_{\mathrm{tr}} \in O\left(\sqrt{d(1+\log 1/p)}\right)$$

Proof. Let σ be an arbitrary p-uniform measure, then

$$\left\| \mathop{\mathbb{E}}_{U\in\sigma} [U] \right\|_{\mathrm{tr}} = \max_{V\in\mathbf{U}(d)} \left| \mathrm{tr} \mathop{\mathbb{E}}_{U\in\sigma} [U]\, V \right| = \max_{V\in\mathbf{U}(d)} \left| \mathop{\mathbb{E}}_{U\in\sigma} [\mathrm{tr}\, UV] \right| = \max_{V\in\mathbf{U}(d)} \left| \mathop{\mathbb{E}}_{U\in\sigma V} [\mathrm{tr}\, U] \right|.$$

Notice however that the measure σV is p-uniform whenever σ is, and so by Proposition 6.1 we may, since σ is arbitrary, discard the maximization over V. Doing so, the desired quantity is

$$\left| \mathop{\mathbb{E}}_{U\in\sigma} \mathrm{tr}\, U \right| \leq \mathop{\mathbb{E}}_{U\in\sigma} |\mathrm{tr}\, U| = \mathop{\mathbb{E}}_{U\in\sigma} \sum_{i=1}^{d} |\langle i|U|i\rangle| = \sum_{i=1}^{d} \mathop{\mathbb{E}}_{|\psi_i\rangle\in\tau_i} |\langle i|\psi_i\rangle|, \tag{8}$$

where for each i, τ_i is the p-uniform measure on $\mathbf{S}^{d-1}$ obtained by applying a σ-distributed unitary U to the state $|i\rangle$. Having reduced the problem to an expectation over a p-uniform measure on pure states, we apply the bound in Lemma 6.5 to Equation (8) to get

$$\left\| \mathop{\mathbb{E}}_{U\in\sigma} [U] \right\|_{\mathrm{tr}} \leq \sum_{i=1}^{d} O\left(\sqrt{\frac{1+\log 1/p}{d}}\right) = O\left(\sqrt{d(1+\log 1/p)}\right),$$

as in the statement of the Lemma. □

Theorem 6.7. *Any* QCMA *protocol for problem 6.2 with an m-bit witness uses $\Omega(\sqrt{d/(m+1)})$ calls to the oracle.*

Proof. Consider any QCMA protocol with any m-bit witness. We will show that this protocol requires at least $\Omega(\sqrt{d/(m+1)})$ calls to the oracle to determine whether it is an oracle of the first or second type.

We use the hybrid approach of Bennet et al. [4]. Let ρ_0 be the initial state of the algorithm. Let ρ_i be the state of the algorithm immediately after the ith call to an oracle of type 2. After T calls to such an oracle, we denote the final state of the algorithm (before the measurement of whether or not to accept) as ρ_T. In the case that the algorithm is run on an oracle of type 1, we denote the final state by ξ_T. Our goal is to show that the distance between ρ_T and ξ_T is small, unless T, the number of oracle calls, is sufficiently large. We will do this by considering running the algorithm for $(i-1)$ queries on an oracle of type 2 and then switching the oracle to type 1. We denote the state obtained in this way by ρ'_i. We prove that this state is very close to the state ρ_i, which will give the desired result, since $\|\xi_T - \rho_T\|_{\mathrm{tr}} \leq \sum_{i=1}^{T} \|\rho_i - \rho'_i\|_{\mathrm{tr}}$ by the triangle inequality.

Let $|\nu\rangle = \alpha|0\rangle + \beta|1\rangle$ and let $\nu = |\nu\rangle\langle\nu|$ be the input to the $(k+1)$st call to the oracle, after the algorithm has been run for k queries to an oracle of type 2. Strictly speaking, ν may be mixed state, but a convexity argument implies that a pure input state will maximize the distance between the output states of the two oracles. The output of the O_2 on the pure state ν is the mixed state

$$O_2(\nu) = \frac{1}{d^2}\left(|\alpha|^2\,|0\rangle\langle 0| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + |\beta|^2\,|1\rangle\langle 1| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}\right). \tag{9}$$

The output of the oracle O_1, for a fixed hidden unitary U, is

$$O_1^U(\nu) = \frac{1}{d^2}\Big(|\alpha|^2 |0\rangle\langle 0| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}} + \alpha\bar{\beta}|0\rangle\langle 1| \otimes U^\dagger \otimes \mathbb{1}_{\mathcal{K}} + \bar{\alpha}\beta|1\rangle\langle 0| \otimes U \otimes \mathbb{1}_{\mathcal{K}} + |\beta|^2 |1\rangle\langle 1| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}\Big).$$

However, since this is the first query the algorithm has made to the oracle O_1, it has no information about the hidden unitary U, except the m-bit classical message from the QCMA prover. This information constrains the unitary U to a 2^{-m}-uniform distribution σ, so that the output of oracle O_1 can be represented by the mixture of the previous equation over all $U \in \sigma$, which is

$$O_1(\nu) = \mathop{\mathbb{E}}_{U\in\sigma}\left[O_1^U(\nu)\right] \tag{10}$$

One way to think about this, is that the oracle O_1 has another space which is initialized to be a uniform superposition of descriptions of all possible unitaries. Then the oracle uses this register as a control in order to apply the mapping O_1^U. The classical QCMA message could be thought of as an outcome to a partial measurement on this register, which resulted in the collapse of the uniform superposition to a p-uniform superposition of the unitaries consistent with the measurement outcome. The verifier's view can be calculated by tracing out this register.

The remaining task is to compute the diamond norm of the difference of Equations (9) and (10), which will measure the maximum probability that any measurement can distinguish whether or not a single call to the oracle O_1 has been replaced by a call to O_2.

$$\|O_1(\nu) - O_2(\nu)\|_{\mathrm{tr}} = \frac{1}{d^2}\left\| \alpha\bar{\beta}|0\rangle\langle 1| \otimes \mathop{\mathbb{E}}_{U\in\sigma}[U^\dagger \otimes \mathbb{1}_{\mathcal{K}}] + \bar{\alpha}\beta|1\rangle\langle 0| \otimes \mathop{\mathbb{E}}_{U\in\sigma}[U \otimes \mathbb{1}_{\mathcal{K}}] \right\|_{\mathrm{tr}}$$

We then use the fact that $\left\| |0\rangle\langle 1| \otimes A^\dagger + |1\rangle\langle 0| \otimes A \right\|_{\mathrm{tr}} = 2\|A\|_{\mathrm{tr}}$ (see [5, Section II.1] for the relationship between the eigenvalues of an operator of this form and the singular values of A). This implies that

$$\|O_1(\nu) - O_2(\nu)\|_{\mathrm{tr}} = \frac{2|\alpha||\beta|}{d^2}\left\| \mathop{\mathbb{E}}_{U\in\sigma}[U \otimes \mathbb{1}] \right\|_{\mathrm{tr}} = \frac{2|\alpha||\beta|}{d}\left\| \mathop{\mathbb{E}}_{U\in\sigma}[U] \right\|_{\mathrm{tr}}.$$

Finally, since σ is a 2^{-m} uniform measure on $\mathbf{U}(d)$ we apply Lemma 6.6 to obtain

$$\|O_1(\nu) - O_2(\nu)\|_{\mathrm{tr}} \in O\left(\sqrt{\frac{1+m}{d}}\right). \tag{11}$$

This equation bounds the trace distance of the output states of the two oracles. The maximum distance between the states ρ_i and ρ_i' is upper bounded by the diamond norm, which takes into account the fact that the algorithm may use an ancillary space to better distinguish the two oracles. Using the fact that the diamond norm of the difference of two channels is achieved by a pure quantum state [29], we have shown that there exists some pure state ν such that for all $i \in \{1, \ldots, T\}$

$$\left\|\rho_i - \rho_i'\right\|_{\mathrm{tr}} \leq \|O_1 - O_2\|_\diamond \leq 2\|O_1(\nu) - O_2(\nu)\|_{\mathrm{tr}} \in O\left(\sqrt{(1+m)/d}\right),$$

where we have used Lemma 2.2 to upper bound the diamond norm by the trace norm. The triangle inequality implies that replacing all T calls to O_1 with calls to O_2 results in states ρ_T and ξ_T with trace distance

$$\|\rho_T - \xi_T\|_{\mathrm{tr}} \leq \sum_{i=1}^{T} \left\|\rho_i - \rho_i'\right\|_{\mathrm{tr}} \in O\left(T\sqrt{(1+m)/d}\right).$$

This implies that in order for a black-box algorithm to distinguish O_1 and O_2 with constant probability it is required to make $T = \Omega(\sqrt{d/(1+m)})$ calls to the oracle. □

We now use Protocol 6.3 and the lower bound in Theorem 6.7 to obtain an oracle relative to which QSZK is not contained in QCMA. The proof of this follows very closely the argument of Aaronson and Kuperberg [2], who establish an oracle relative to which QMA is not in QCMA.

Strictly speaking, we find a quantum oracle A such that $\mathsf{QSZK}_{\mathsf{HV}}^A \not\subseteq \mathsf{QCMA}^A$, i.e. we deal only with the honest verifier case. While it is known that $\mathsf{QSZK}_{\mathsf{HV}} = \mathsf{QSZK}$ [37], we do not know if this is still the case given access to the oracle A.

Theorem 1.2. *There exists a quantum oracle A such that $\mathsf{QSZK}_{\mathsf{HV}}^A \not\subseteq \mathsf{QCMA}^A$*

Proof. Let L be a random unary language that we will use to define the oracle $A = \{A_n\}$. For each n, A_n takes $2n$ qubits as input (so that $d = 2^n$ in Problem 6.2). For each n there are two cases. If $1^n \in L$ then A_n is an oracle of type 1 in Problem 6.2, i.e. A_n implements some hidden unitary U on half of the input qubits. On the other hand, if $1^n \notin L$, then A_n is of type 2.

We use Theorem 6.4 to give an honest-verifier QSZK protocol for L, given access to the oracle A. For a given input 1^n, the Verifier first runs protocol 6.3 to determine the type of the oracle. The verifier accepts that $1^n \in L$ if and only if this protocol accepts. The completeness and soundness of the protocol have already been shown. Last, it is easy to show that the protocol is zero knowledge for the honest verifier. The state of the verifier after Step 1 can be simulated by the simulator, since it has at its disposal both the honest verifier and the oracle. After the prover's message, in the 'yes' case, the state is equal to

$$|\phi^+\rangle\langle\phi^+| \otimes \mathbb{1}_{\mathcal{H}\otimes\mathcal{K}}/d^2$$

which can also be easily simulated, and so the protocol is (honest-verifier) zero-knowledge. This implies that $L \in \mathsf{QSZK}_{\mathsf{HV}}^A$.

We then use the lower bound in Theorem 6.7 to show that $L \notin \mathsf{QCMA}^A$, with probability one (over the choice of L and the hidden unitary U in the oracle). This portion of the proof is identical to the proof in [2], but for clarity we repeat it here. Fix M an arbitrary QCMA verifier and let $S_M(n)$ represent the event that the verifier M succeeds on the input 1^n, i.e. either $1^n \in L$ and there exists a witness string w such that M^A accepts with probability at least $2/3$, or $1^n \notin L$ and no witness w causes M to accept with probability larger than $1/3$. Theorem 6.7 implies that M fails for large enough n, i.e. that for some N it holds that for all $n \geq N$

$$\Pr_{L,V}[S_M(n)|S_M(1), \ldots, S_M(n-1)] \leq \frac{2}{3}.$$

This implies that the probability that M works on all n is 0, i.e.

$$\Pr_{L,V}[S_M(1) \wedge S_M(2) \cdots] = 0.$$

Finally, since there are only a countably infinite number of QCMA verifiers (by the Solovay-Kitaev Theorem [18]), the union bound implies that with probability one we have $L \notin \mathsf{QCMA}$. □

Acknowledgements

BR is supported by the Centre for Quantum Technologies, which is funded by the Singapore Ministry of Education and the Singapore National Research Foundation. AC and IK are supported by projects ANR-09-JCJC-0067-01, ANR-08-EMER-012 and QCS (grant 255961) of the E.U.

References

[1] S. Aaronson. Impossibility of succinct quantum proofs for collision-freeness. arXiv:1101.0403 [quant-ph], 2011.

[2] S. Aaronson and G. Kuperberg. Quantum versus classical proofs and advice. *Theory of Computing*, 3(7):129–157, 2007. DOI: 10.4086/toc.2007.v003a007. EPRINT: arXiv:quant-ph/0604056.

[3] M. Ben-Or, O. Goldreich, S. Goldwasser, J. Håstad, J. Kilian, S. Micali, and P. Rogaway. Everything provable is provable in zero-knowledge. In *CRYPTO 1988*, volume 403 of *LNCS*, pp. 37–56. 1990. DOI: 10.1007/0-387-34799-2_4.

[4] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997. DOI: 10.1137/S0097539796300933. EPRINT: arXiv:quant-ph/quant-ph/9701001.

[5] R. Bhatia. *Matrix Analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer, 1997.

[6] H. Buhrman, R. Cleve, J. Watrous, and R. de Wolf. Quantum fingerprinting. *Phys. Rev. Lett.*, 87(16):167902, 2001. DOI: 10.1103/PhysRevLett.87.167902. EPRINT: arXiv:quant-ph/0102001.

[7] C. Crépeau, F. Légaré, and L. Salvail. How to convert the flavor of a quantum bit commitment. In *EUROCRYPT2001*, volume 2045 of *LNCS*, pp. 60–77. 2001. DOI: 10.1007/3-540-44987-6_5.

[8] C. A. Fuchs and J. van de Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Trans. Inf. Theory*, 45(4):1216–1227, 1999. DOI: 10.1109/18.761271. EPRINT: arXiv:quant-ph/9712042.

[9] O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *J. ACM*, 38(3), 1991. DOI: 10.1145/116825.116852.

[10] I. Haitner, M.-H. Nguyen, S. J. Ong, O. Reingold, and S. Vadhan. Statistically hiding commitments and statistical zero-knowledge arguments from any one-way function. *SIAM J. Comput.*, 39(3):1153–1218, 2009. DOI: 10.1137/080725404.

[11] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999. DOI: 10.1137/S0097539793244708.

[12] C. W. Helstrom. Detection theory and quantum mechanics. *Inform. Control*, 10(3):254–291, 1967. DOI: 10.1016/S0019-9958(67)90302-6.

[13] R. Impagliazzo and M. Luby. One-way functions are essential for complexity based cryptography. In *IEEE Symp. Found. Comput. Sci. (FOCS)*, pp. 230 – 235. 1989. DOI: 10.1109/SFCS.1989.63483.

[14] R. Jain, Z. Ji, S. Upadhyay, and J. Watrous. QIP = PSPACE. In *ACM Symp. Theory Comput. (STOC)*. 2010. EPRINT: arXiv:0907.4737 [quant-ph].

[15] N. Johnston, D. W. Kribs, and V. I. Paulsen. Computing stabilized norms for quantum operations via the theory of completely bounded maps. *Quantum Inf. Comput.*, 9(1&2):16–35, 2009. EPRINT: arXiv:0711.3636 [quant-ph].

[16] R. Jozsa. Fidelity for mixed quantum states. *J. Mod. Opt.*, 41(12):2315–2323, 1994. DOI: 10.1080/09500349414552171.

[17] A. Kitaev and J. Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *ACM Symp. Theory Comput. (STOC)*, pp. 608–617. 2000. DOI: 10.1145/335305.335387.

[18] A. Y. Kitaev. Quantum computations: algorithms and error correction. *Russ. Math. Surveys*, 52(6):1191–1249, 1997. DOI: 10.1070/RM1997v052n06ABEH002155.

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

[20] A. R. Klivans and D. van Melkebeek. Graph Nonisomorphism has subexponential size proofs unless the polynomial-time hierarchy collapses. *SIAM J. Comput.*, 31(5):1501–1526, 2002. DOI: 10.1137/S0097539700389652.

[21] H.-K. Lo and H. F. Chau. Is quantum bit commitment really possible? *Phys. Rev. Lett.*, 78:3410, 1997. DOI: 10.1103/PhysRevLett.78.3410. EPRINT: arXiv:quant-ph/9603004.

[22] C. Marriott and J. Watrous. Quantum Arthur-Merlin games. *Comput. Complex.*, 14(2):122–152, 2005. DOI: 10.1007/s00037-005-0194-x. EPRINT: arXiv:cs/0506068.

[23] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Phys. Rev. Lett.*, 78:3414, 1997. DOI: 10.1103/PhysRevLett.78.3414. EPRINT: arXiv:quant-ph/9605044.

[24] P. B. Miltersen and N. V. Vinodchandran. Derandomizing Arthur-Merlin games using hitting sets. *Comput. Complex.*, 14(3):256–279, 2006. DOI: 0.1007/s00037-005-0197-7.

[25] M. Naor. Bit commitment using pseudorandomness. *Journal of Cryptology*, 4(2):151–158, 1991. DOI: 10.1007/BF00196774.

[26] A. Nayak and P. Shor. Bit-commitment-based quantum coin flipping. *Phys. Rev. A*, 67(1):012304, 2003. DOI: 10.1103/PhysRevA.67.012304. EPRINT: arXiv:quant-ph/0206123.

[27] R. Ostrovsky and A. Wigderson. One-way functions are essential for non-trivial zero-knowledge. In *2nd Israel Symposium on Theory and Computing Systems*, pp. 3–17. 1993. DOI: 10.1109/ISTCS.1993.253489.

[28] V. Paulsen. *Completely Bounded Maps and Operator Algebras*, volume 78 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, 2002.

[29] B. Rosgen and J. Watrous. On the hardness of distinguishing mixed-state quantum computations. In *Conf. Comput. Compl. (CCC)*, pp. 344–354. 2005. DOI: 10.1109/CCC.2005.21. EPRINT: arXiv:cs/0407056.

[30] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26(5):1484–1509, 1997. DOI: 10.1137/S0097539795293172. EPRINT: arXiv:quant-ph/9508027.

[31] R. R. Smith. Completely bounded maps between C*-algebras. *J. London Math. Soc.*, s2-27(1):157, 1983. DOI: 10.1112/jlms/s2-27.1.157.

[32] R. W. Spekkens and T. Rudolph. Degrees of concealment and bindingness in quantum bit commitment protocols. *Phys. Rev. A*, 65(1):012310, 2001. DOI: 10.1103/PhysRevA.65.012310. EPRINT: arXiv:quant-ph/0106019.

[33] S. Vadhan. An unconditional study of computational zero knowledge. *SIAM J. Comput.*, 36(4):1160–1214, 2006. DOI: 10.1137/S0097539705447207.

[34] J. Watrous. Succinct quantum proofs for properties of finite groups. *IEEE Symp. Found. Comput. Sci. (FOCS)*, pp. 537 – 546, 2000. DOI: 10.1109/SFCS.2000.892141. EPRINT: arXiv:cs/0009002.

[35] J. Watrous. Limits on the power of quantum statistical zero-knowledge. In *IEEE Symp. Found. Comput. Sci. (FOCS)*, pp. 459 – 468. 2002. DOI: 10.1109/SFCS.2002.1181970. EPRINT: arXiv:quant-ph/0202111.

[36] J. Watrous. PSPACE has constant-round quantum interactive proof systems. *Theoretical Computer Science*, 292(3):575–588, 2003. DOI: 10.1016/S0304-3975(01)00375-9. EPRINT: arXiv:cs/9901015.

[37] J. Watrous. Zero-knowledge against quantum attacks. *SIAM J. Comput.*, 39(1):25–58, 2009. DOI: 10.1137/060670997. EPRINT: arXiv:quant-ph/0511020.