Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 140 (2022)

Commitments to Quantum States

Sam Gunn* Nathan Ju† Fermi Ma‡ Mark Zhandry§

November 4, 2022

Abstract

What does it mean to commit to a quantum state? In this work, we propose a simple answer: a commitment to quantum messages is *binding* if, after the commit phase, the committed state is *hidden* from the sender's view. We accompany this new definition with several instantiations. We build the first non-interactive *succinct* quantum state commitments, which can be seen as an analogue of collision-resistant hashing for quantum messages. We also show that hiding quantum state commitments (QSCs) are implied by any commitment scheme for classical messages. All of our constructions can be based on quantum-cryptographic assumptions that are implied by *but are potentially weaker than* one-way functions.

Commitments to quantum states open the door to many new cryptographic possibilities. Our flagship application of a succinct QSC is a quantum-communication version of Kilian's succinct arguments for any language that has quantum PCPs with constant error and polylogarithmic locality. Plugging in the PCP theorem, this yields succinct arguments for NP under significantly weaker assumptions than required classically; moreover, if the quantum PCP conjecture holds, this extends to QMA. At the heart of our security proof is a new rewinding technique for extracting quantum information.

*UC Berkeley. Email: gunn@berkeley.edu. Supported by a Google PhD Fellowship.
†UC Berkeley. Email: nju@berkeley.edu.
‡Simons Institute & UC Berkeley. Email: fermima@alum.mit.edu.
§NTT Research & Princeton University. Email: mzhandry@gmail.com.

ISSN 1433-8092

Contents

1 Introduction **1**
- 1.1 Our results . . . 2

2 Technical overview **6**
- 2.1 Definitions . . . 6
- 2.2 Constructions . . . 10
- 2.3 Applications . . . 12
- 2.4 Related works . . . 17

3 Preliminaries **20**

4 Defining quantum state commitments (QSCs) **22**
- 4.1 Syntax . . . 22
- 4.2 The swap binding definition . . . 22
- 4.3 Additional definitions: hiding and succinctness . . . 23
- 4.4 Hiding-binding duality . . . 24
- 4.5 Parallel composition . . . 26

5 Constructions of QSCs **28**
- 5.1 Succinct QSCs from one-time quantum encryption . . . 28
- 5.2 Domain extension . . . 30
- 5.3 Formalizing the folklore: QSCs from QBCs . . . 35

6 The oracle security lemma **41**
- 6.1 The admissible oracle lemma . . . 42
- 6.2 Swap binding implies oracle swap binding . . . 47

7 A succinct quantum argument protocol **51**
- 7.1 Preliminaries . . . 51
- 7.2 Quantum tree commitments . . . 53
- 7.3 The protocol description . . . 54

8 Security of the quantum succinct argument protocol **56**
- 8.1 CMSZ preliminaries . . . 56
- 8.2 Description of the knowledge extractor . . . 59
- 8.3 Correctness of our extractor . . . 62

9 Computationally sound quantum sigma protocols **69**
- 9.1 Argument of knowledge . . . 71
- 9.2 Zero knowledge . . . 71

A Removing interaction from any QSC **81**

B Equivalences between binding definitions **84**
- B.1 Pauli binding . . . 84
- B.2 A statistical extraction-based definition . . . 88

C Quantum encryption with short keys from PRUs **91**

1 Introduction

Is it possible to commit to a quantum state? There is a simple "folklore" proposal: apply a quantum one-time pad $X^r Z^s$ to the state $\boldsymbol{\rho}$, and send the resulting state along with a commitment to the classical string (r, s). While this scheme has appeared implicitly in several quantum cryptographic protocols [BJSW16, CVZ20, BG20], our understanding of commitments to quantum states is extremely limited. For example, consider the following basic questions:

- How should binding even be *defined* for commitments to quantum messages? A proper definition of binding for quantum messages should capture the semantic meaning of *binding*, while also being useful for cryptographic applications.
- The commitment in the folklore construction is always larger than the original message. Is it possible to *succinctly* commit to a quantum state? (For example, is it possible to commit to a $2n$-qubit state with an n-qubit commitment?)
- What kinds of cryptographic protocols are enabled by commitments to quantum states?

More broadly, commitments are central to cryptography and appear in many different forms and contexts. Does an analogous picture exist in the quantum setting?

This work. We initiate a formal study of quantum-state commitments (QSCs) and their applications. We develop general techniques to answer the above questions and position QSCs to play a vital role in quantum cryptography. Our contributions are the following:

- **Definitions.** We provide a definition of binding for QSCs: informally, committing to a quantum state $\boldsymbol{\rho}$ should *erase* it from the sender's view. Our definition captures computational and statistical security, composes across multiple commitments, and we show that it naturally generalizes Unruh's collapse-binding definition for (quantum-secure) *classical* commitments [Unr16]. Perhaps most surprisingly, we show that the notions of binding and hiding for QSCs satisfy a *duality* that has no known analogue in classical cryptography.
- **Constructions.** Our new definition directly enables new constructions. We build the first non-interactive *succinct* QSCs, which can be seen as an analogue of collision-resistant hashing for quantum messages. We also formalize the "folklore" proposal to show that hiding and binding QSCs exist if and only if hiding and binding quantum bit commitments (QBCs) exist (i.e., commitments to *classical* messages). All of our constructions can be based on assumptions that are implied by *but are potentially weaker than one-way functions*.
- **Applications.** Finally, we use QSCs to build protocols. Our flagship application of a succinct QSC is a quantum version of Kilian's succinct arguments [Kil92, CM21] for any language that has quantum PCPs with constant error and polylogarithmic locality. Plugging in the PCP theorem, this yields succinct arguments for NP from significantly weaker assumptions than in the classical setting; moreover, if the quantum PCP conjecture holds, this result extends to all of QMA. Proving security is the most technical component of this work. Our proof

develops a new rewinding technique for extracting quantum information, combining ideas from [CMSZ21] (which handles quantum attacks on *classical* protocols) with new tools and abstractions designed for quantum protocols.

We now give a more detailed description of our results.

1.1 Our results

1.1.1 Definitions

Syntax for quantum state commitments. Before we present our binding definition, we establish some basic syntax for *non-interactive* QSCs:

- The sender commits to a state $|\psi\rangle$ by applying a public unitary Com to $|\psi\rangle\,|0^\lambda\rangle$, obtaining a state on registers (C,D). The commitment is the state on register C.
- The sender can later open the commitment by sending the register D. The receiver verifies by applying $\mathsf{Com}^\dagger$ and checking that the last λ qubits are 0. If verification succeeds, the receiver recovers the committed state $|\psi\rangle$ from the unmeasured registers.[1]

There is a simple transformation that turns any interactive QSC into a non-interactive one: define Com to run the honest interactive commit phase coherently (see Appendix A for more details).[2] We will therefore focus our attention on non-interactive QSCs.

Swap binding: a definition for quantum messages. Our new binding definition — which we call *swap binding* — requires that once the adversary has sent the commitment C, it can no longer distinguish its original decommitment from one where the committed message has been swapped with junk. In other words, sending C erases the committed message from the sender's view. We make this concrete with the following security game:

1. The adversary sends (C,D) to the challenger.
2. The challenger applies $\mathsf{Com}^\dagger$ and verifies that the last λ qubits are 0, and if not, aborts.[3] Next, it samples a random bit $b \leftarrow \{0,1\}$ and does one of the following:
   - If $b = 0$, it applies Com and sends D back to the adversary.

[1] A similar *syntax* appeared in prior work of Chen and Movassagh [CM21], who proposed applying a Haar random unitary to $|\psi\rangle\,|0^\lambda\rangle$. However, since their goal was to build quantum tree commitments (see Section 2.4 for further discussion), they did not formalize this syntax for individual commitments.

[2] In the setting of quantum bit commitments, i.e., quantum commitments to *one-bit classical messages*, a similar transformation and syntax for non-interactive commitments previously appeared in [Yan22].

[3] Since a successful verification measurement completely collapses the last λ qubits, it would be equivalent to consider a security game where, instead of sending (C,D), the adversary sends the challenger the committed message in the clear; indeed, this fact is crucial for our results on the "hiding-binding duality." However, allowing the adversary to specify (C,D) yields a more robust binding definition, since it would also capture schemes where verification does not completely collapse the last λ qubits.

- If $b = 1$, it initializes an ancilla register to $|0\rangle$, and applies SWAP to replace the committed message with $|0\rangle$.[4] Then it applies Com and sends D back to the adversary.

3. Finally, the adversary wins if it can guess b.

Definition 1.1 (Swap binding; see Definition 4.2)**.** *A non-interactive QSC is computationally (resp. statistically) binding if no efficient (resp. inefficient) adversary can guess b with advantage greater than $1/2 + \mathsf{negl}(\lambda)$.*[5]

Swap binding satisfies a number of desirable properties:

- It handles both computational and statistical binding.
- It works for message spaces of any dimension.
- It composes in parallel and in sequence across multiple commitments (see Sections 4.5, 5.2 and 7.2).
- It encompasses binding for classical messages. In fact, we show that our definition can be viewed as a natural generalization of Unruh's collapse-binding definition [Unr16], which was originally defined for post-quantum-secure *classical commitments* (see the discussion in Section 2.1.2).

A particularly interesting property of swap binding is that it guarantees that QSCs *respect entanglement across multiple commitments*. For instance, if a sender prepares an EPR pair $(|00\rangle + |11\rangle)/\sqrt{2}$ and commits to the two qubits with two independent QSCs, our definition ensures that if the sender opens both commitments, it must be to the same entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$. This "entanglement-respecting" property will be crucial for all of our applications, which involve committing to a large global state using many commitments.

Hiding-binding duality. A surprising property of swap binding is that it is *dual* to existing notions of *hiding* for quantum messages (hiding was previously defined for quantum messages in prior works on quantum encryption, e.g., [BJ15, Definition 3.3]) in the following sense: hiding requires that C alone reveals no information about the committed message $|\psi\rangle$, while binding requires that D alone reveals no information about $|\psi\rangle$.

This has several interesting consequences that have no analogue in classical cryptography. For instance, given a computationally binding, statistically hiding QSC with commitment C and decommitment D, there is a simple "dual" scheme satisfying statistical binding and computational hiding: just send D as the commitment and use C as the opening! We give further details in Section 4.4. We also point out a duality between hiding and binding for quantum commitments to *classical messages* (see Section 4.4.1).

[4] Recall that SWAP maps $|\psi\rangle|\phi\rangle$ to $|\phi\rangle|\psi\rangle$ for all pure states $|\psi\rangle$, $|\phi\rangle$.

[5] Our swap-binding definition is stated for *non-interactive* QSCs. Despite the fact that any interactive QSC can be made non-interactive, we believe it could still be useful to have a definition that directly handles interactive QSCs. In Appendix B.1, we give an alternative definition of binding for QSCs that we call "Pauli binding" and show that (1) Pauli binding is equivalent to swap binding for non-interactive QSCs and (2) Pauli binding captures both interactive and non-interactive QSCs. However, it is unclear how to use interactive QSCs that satisfy Pauli binding to build secure interactive protocols. We leave this as an open question for future work.

1.1.2 Constructions

Our new definition directly enables new constructions of QSCs. We build the first *succinct* QSC, which allows a sender to commit to a $\mathsf{poly}(\lambda)$-qubit quantum state with a λ-qubit commitment C. These succinct QSCs behave like a quantum analogue of collision-resistant hash functions (CRHFs) for quantum states; for instance, we show that classical techniques for composing CRHFs such as Merkle-Damgård and Merkle tree commitments naturally generalize to succinct QSCs.

However, unlike CRHFs, our succinct QSCs are un-keyed and thus are *completely non-interactive*.[6] Moreover, the assumptions required for succinct QSCs appear to be much weaker than those required for CRHFs.

Theorem 1.2 (Succinct QSCs from encryption; see Section 5.1)**.** *Assume the existence of one-time secure private-key encryption for $(\lambda+1)$-qubit quantum messages with λ-bit classical keys. Then for any $k = \mathsf{poly}(\lambda)$, there exists a non-interactive succinct* QSC *for committing to k-qubit quantum messages with commitments of only λ qubits.*

Such one-time secure encryption schemes are implied by one-way functions, and even potentially weaker assumptions such as (one-time secure) pseudorandom unitaries.

We also use our definition to formalize security of the "folklore" scheme described above. As a consequence, hiding and binding QSCs exist as long as hiding and binding quantum *bit* commitments (QBCs) — i.e., quantum commitments to classical messages — exist. Since QSCs encompass commitments to classical messages, this implies that the existence of QBCs and QSCs are equivalent assumptions.

Theorem 1.3 (Section 5.3)**.** *Quantum state commitments exist (satisfying hiding and binding for quantum messages) if and only if quantum bit commitments exist (satisfying hiding and binding for classical messages).*

Quantum bit commitments are implied by one-way functions, and even potentially weaker assumptions [Yan22, Kre21, AQY22, MY22, BCQ22].

1.1.3 Applications

Our flagship application is a three-message quantum succinct argument based on succinct QSCs. The protocol is a direct quantum analogue of Kilian's PCP-based succinct arguments [Kil92]. Recall that Kilian's protocol assumes the existence of a succinct classical commitment (i.e., a collision-resistant/collapsing hash function) and can be instantiated for any language that has classical PCPs with constant error and polylogarithmic locality, i.e., any NP language [BFLS91, FGL+91, AS98, ALM+98]. Correspondingly, our quantum Kilian protocol assumes the existence of a succinct QSC and can be instantiated for any language that has quantum PCPs[7] with constant error and polylogarithmic locality.[8] The set of such languages includes NP, and is famously conjectured to include all of QMA [AALV09, AAV13].

[6] While un-keyed CRHFs are impossible classically due to non-uniform attacks, such attacks can be ruled out in the quantum setting by monogamy of entanglement!

[7] A quantum PCP is a quantum proof that can be probabilistically checked by measuring a few qubits.

[8] Our protocol is inspired by a proposal of Chen and Movassagh [CM21], who showed that Kilian's protocol has a *syntactic* quantum analogue and conjecture (but do not prove) its security in the "Haar Random Oracle Model."

Theorem 1.4 (Quantum succinct arguments; see Theorem 8.1)**.** *Assuming the existence of succinct QSCs, there is a three-message quantum-communication succinct argument for any language that has quantum PCPs with constant error and polylogarithmic locality.*

Corollary 1.5. *Assuming the existence of succinct QSCs, there is a three-message quantum succinct argument for* NP.[9]

Corollary 1.6. *If the quantum PCP conjecture holds [AALV09, AAV13], then assuming the existence of succinct QSCs, there is a three-message quantum succinct argument for* QMA.

Even in the context of succinct arguments for NP, Corollary 1.5 lowers the round complexity[10] and relies on weaker assumptions than Kilian's result.

Swap-based rewinding for quantum protocols. Proving Theorem 8.1 is the most technical component of this work. At a high level, proving soundness requires extracting *quantum information* from the malicious prover. This is qualitatively different from prior works on extracting from quantum adversaries [Unr12, Unr16, CMSZ21] which only needed to extract classical information and relied heavily on the ability to record (i.e., copy) the adversary's responses.

Our rewinding procedure works by swapping out the messages underlying the adversary's responses. This explains how QSCs make cryptographic applications possible: *swap binding enables undetectable extraction of quantum information in a rewinding-based security analysis.*

While the swapping idea is simple, it introduces a fundamental problem that was not present in the classical setting. In our setting, by swapping out the adversary's message, we have implicitly forced the adversary to *forget* part of its committed PCP. As a result, [CMSZ21]-style state repair — the only known technique for rewinding succinct protocols — is now *information-theoretically impossible*. We explain our new rewinding approach in Section 2.3.

Quantum sigma protocols. Finally, as an additional application of our techniques, we prove in Section 9 that any [GMW86, BG20]-style quantum sigma protocol is sound when instantiated with our hiding and binding QSCs. We formalize this using a quantum version of the zero-knowledge PCP framework of [IKOS07]. This also provides a proof of computational soundness for a quantum sigma protocol due to [BG20], which was missing a security analysis (see Section 2.4 for additional details).

[9] This follows by combining Theorem 1.4 with the classical PCP theorem [BFLS91, FGL+91, AS98, ALM+98], since quantum PCPs encompass classical PCPs.

[10] [BKP18] constructs three message succinct arguments for NP under a non-standard assumption called "keyless multi-collision-resistant hash functions" and the hardness of learning with errors.

2 Technical overview

2.1 Definitions

Syntax. Recall our syntax for non-interactive quantum state commitments from Section 1.1.1:

- The sender commits to a state $|\psi\rangle$ by applying a public unitary Com to $|\psi\rangle\ |0^\lambda\rangle$, obtaining a state on registers (C, D). The commitment is the register C.
- The sender can later open the commitment by sending the register D. The receiver verifies by applying $\mathsf{Com}^\dagger$ and checking that the last λ qubits are 0. If verification succeeds, the receiver recovers the committed message $|\psi\rangle$ from the unmeasured registers.

We emphasize one important difference from the classical setting: classically, the decommitment d can always be assumed to contain m in the clear, i.e., $d = (m, r)$ where m is the committed message and r is the randomness/opening information. For QSCs, however, the D register crucially *does not* contain the committed message in the clear.

2.1.1 What does binding mean?

Our first goal in this overview is to answer the following question:

What does binding mean for a commitment to quantum messages?

To understand the subtleties that arise, we first recall the classical definition of binding.

Definition 2.1 (Classical binding, informal)**.** *A commitment scheme is binding if an adversary cannot generate a commitment c along with two valid decommitments $d_1 = (m_1, r_1)$ and $d_2 = (m_2, r_2)$ for two different messages $m_1 \neq m_2$.*

What happens if we try to use this definition for commitments to quantum states? Two closely related issues arise:

- First, this definition requires the challenger to verify that the two messages are actually different. If the messages are arbitrary quantum states, how should the challenger implement this check? A natural idea is to run a swap test[11], but performing the swap test on message registers $\mathsf{M}_1, \mathsf{M}_2$ requires the challenger to obtain two openings *simultaneously*. This brings us to our second issue.
- A quantum adversary does not have to produce two openings simultaneously. Consider a quantum adversary that prepares a valid commitment-decommitment pair (C, D) corresponding to a message $|\psi\rangle$, and suppose it has the ability to modify the state on D so that the commitment opens to a completely different message $\left|\psi^\perp\right\rangle$. This adversary clearly violates any reasonable notion of binding, but since it produces the two openings sequentially rather than in parallel, it is not captured by the classical-style definition.[12]

[11] The swap test on two quantum states $|\psi\rangle, |\phi\rangle$ outputs 1 with probability $\frac{1}{2} + \frac{1}{2}|\langle\psi|\phi\rangle|^2$.

[12] This issue was first observed by [ARU14, Unr16] in the setting of classical commitments secure against quantum attacks.

At first glance, it may seem impossible to handle both issues at once. How can a challenger possibly verify that two arbitrary quantum states are different if it only gets to see one state at a time?

We propose the following solution. In the security game, the adversary will open the commitment twice *in sequence*, i.e., it sends registers (C, D) to the challenger (corresponding to the first opening), the challenger later returns register D to the adversary, and finally the adversary sends the challenger a new state on D (corresponding to a second opening). Meanwhile, the challenger picks one of the openings at random and swaps out the committed message M into an internal register M', replacing the committed message with a junk $|0\rangle$ state. At the end of the experiment, the challenger sends M' to the adversary and asks the adversary to distinguish whether it came from its first opening or its second opening. If the adversary can distinguish the two cases, the challenger can be convinced the two messages are different!

Remark 2.2. *This idea is reminiscent of the [GMW87] zero-knowledge protocol for graph non-isomorphism (in which the prover demonstrates that two graphs are not isomorphic by showing that it can distinguish them), except that we are using it to define a security property rather than to build a protocol.*

In slightly more detail, the security game is the following:

`DoubleOpenExpt`:

1. The adversary sends (C, D).
2. The challenger checks that the decommitment is valid by applying $\mathsf{Com}^\dagger$ and measuring the last λ qubits (and aborts if it's not 0^λ). The challenger samples a bit $b \leftarrow \{0,1\}$ and does one the following:
   - (If $b = 0$) It applies Com to recompute the commitment/decommitment.
   - (If $b = 1$) It applies $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ where M is the opened message, and M' is initialized to $|0\rangle$. It then applies Com to recompute the commitment/decommitment.

   Finally, it sends D back to the adversary (but not C).
3. The adversary sends another decommitment on D back to the challenger.
4. The challenger checks that the decommitment is valid by applying $\mathsf{Com}^\dagger$ and measuring the last λ qubits (and aborts if it's not 0^λ). Then:
   - (If $b = 0$) It applies $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ where M is the opened message, and M' is initialized to $|0\rangle$. It then applies Com to recompute the commitment/decommitment.
   - (If $b = 1$) It applies Com to recompute the commitment/decommitment.

   Finally, it sends all of its registers (including M') back to the adversary.
5. The adversary outputs a guess b' and wins if $b' = b$.

Definition 2.3 (Binding for QSCs, Attempt 1)**.** *A quantum state commitment is binding if it is hard to win* `DoubleOpenExpt` *with advantage better than* $1/2 + \mathsf{negl}(\lambda)$.

Intuitively, this definition captures the *semantic meaning* of binding for quantum states. If an adversary can change its committed quantum state across two openings, it should also be able to distinguish between the two opened messages. But this definition requires several rounds of back-and-forth interaction — can we make it any simpler?

A simpler definition: swap binding. It turns out that the second opening in `DoubleOpenExpt` is unnecessary, and a simpler "single opening" game suffices:

`BindExpt`:

1. The adversary sends the registers (C, D).
2. The challenger checks that the decommitment is valid by applying $\mathsf{Com}^\dagger$ and measuring the last λ qubits (and aborts if it's not 0^λ). The challenger samples a bit $b \leftarrow \{0,1\}$ and does one of the following:
   - (If $b = 0$) It applies Com to recompute the commitment/decommitment.
   - (If $b = 1$) It applies $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ where M is the opened message, and M' is initialized to $|0\rangle$. It then applies Com to recompute the commitment/decommitment.

   Finally, it sends the D register back to the adversary.
3. The adversary outputs a guess b' and wins if $b' = b$.

This gives our final definition, which we call *swap binding*.

Definition 2.4 (Swap binding)**.** *A quantum state commitment is swap binding if it is hard to win* `BindExpt` *with advantage better than* $1/2 + \mathsf{negl}(\lambda)$.

Why is this definition *equivalent* to Definition 2.3? At a very high level, any adversary for the double-opening game can be "folded" into an adversary for the single opening game by running the first and second opening rounds in superposition. In slightly more detail, let $|\psi\rangle_{\mathsf{C},\mathsf{D},\mathsf{E}}$ (where E is the internal register of the double opening adversary) be the initial state of the adversary for the double opening game, and let U denote the unitary the adversary applies before sending the second opening. Consider the following adversary for the single opening game:

- The adversary prepares a control qubit initialized to $|+\rangle$, and applies U controlled on $|+\rangle$ to obtain the state $(|0\rangle |\psi\rangle_{\mathsf{C},\mathsf{D},\mathsf{E}} + |1\rangle U |\psi\rangle_{\mathsf{C},\mathsf{D},\mathsf{E}})/\sqrt{2}$. It then sends the registers (C, D) to the challenger.
- Upon receiving the D register back from the challenger, it applies controlled-$U^\dagger$, measures the control qubit in the Hadamard basis, and uses the resulting outcome to guess b.

It is possible to show that if the original adversary can win the double opening game, this adversary will win the single opening game.

2.1.2 Consequences

We briefly discuss some consequences of our new definition:

- **Relationship to classical binding.** Since quantum information encompasses classical information, we should expect our definition to capture binding for classical messages. Indeed, we can show this is the case:

  - By a hybrid argument, we can replace the $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ operation in `BindExpt` with any other quantum operation P that acts on M and an ancilla held by the challenger.[13]
  - Define P to be the operation that applies Z^s to M for a random s; this is equivalent to *measuring* M in the standard basis and discarding the outcome. Our definition implies that the adversary cannot distinguish its original decommitment from a decommitment in which the message has been measured in the standard basis. This is essentially Unruh's definition [Unr16] of collapse-binding for classical messages![14]

  In fact, we can say more. In `BindExpt`, we could initialize M' to the maximally mixed state (instead of $|0\rangle$), and the definition would remain the same. From the adversary's point of view, $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ is equivalent to the challenger applying $X^r Z^s$ to M for uniformly random strings r, s, since this also maximally mixes the message. This gives an alternative view of our definition: a QSC is binding for quantum states if it is (collapse-)binding for classical messages in both the standard and Hadamard bases!

- **Hiding-binding duality.** We have not yet defined hiding for QSCs, but it is not hard to write down a definition: the adversarial receiver sends a register M containing its message $|\psi\rangle$, the challenger either sends a commitment to $|\psi\rangle$ or a commitment to an unrelated message, and the adversary must guess which one it received.[15] Perhaps the most surprising consequence of our definition is that for *non-interactive* QSCs, this hiding experiment is the same as our binding experiment — except that the challenger sends C instead of D to the adversary. Indeed, sending registers (C, D) containing a valid commitment/decommitment state is equivalent to sending a register M containing the committed message state in the clear, since a valid commitment/decommitment is $\mathsf{Com}\,|\psi\rangle_{\mathsf{M}}\,|0^\lambda\rangle$.

  One consequence of this duality is that if we have, for example, a statistically binding and computationally hiding QSC, there is a simple "dual" QSC satisfying computationally binding and statistical hiding: just send D as the commitment and use C as the opening!

[13] We sketch the argument here. In Hybrid 0, the challenger simply sends D back to the adversary after verifying the commitment. In Hybrid 1, the challenger performs $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$ before sending D back; this is indistinguishable from Hybrid 0 by our binding definition. In Hybrid 2, the challenger applies P to M and then applies $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$; this is perfectly indistinguishable from Hybrid 1 since the register P acts on is independent of the adversary's view. Finally, in Hybrid 3, the challenger only applies P; this is indistinguishable from Hybrid 2 by our definition, which allows us to undetectably remove $\mathsf{SWAP}_{\mathsf{M},\mathsf{M}'}$.

[14] Technically, Unruh's definition is for classical commitments to classical messages. However, it is not hard to adapt his definition to the syntax of quantum-communication commitments.

[15] This hiding definition is virtually identical to existing definitions of semantic security for quantum encryption schemes [BJ15, Definition 3.3].

2.2 Constructions

2.2.1 Warm-up: hiding and binding QSCs

Having defined the security guarantees of QSCs, we will now prove that the "folklore" construction is secure. Recall that the sender commits to $|\psi\rangle$ by sampling random r, s and sending $X^r Z^s |\psi\rangle$ together with a commitment $|c_{r,s}\rangle$ to the classical string (r, s). Writing this as a non-interactive QSC (which requires replacing the randomness of r, s with a corresponding superposition), the commitment/decommitment state is (up to normalization)

$$\sum_{r,s}(|r,s\rangle\,|d_{r,s}\rangle)_{\mathsf{D}}(X^r Z^s\,|\psi\rangle\,|c_{r,s}\rangle)_{\mathsf{C}},$$

where $|d_{r,s}\rangle$ is the decommitment for $|c_{r,s}\rangle$.

- For hiding, we must show that the C register alone does not reveal $|\psi\rangle$. This can be easily formalized with a hybrid argument. First, we can replace $|c_{r,s}\rangle$ with $|c_{0,0}\rangle$ by the hiding of the commitment to (r, s). Then, because the superposition over r, s is held on the challenger's registers, the adversary's view in this hybrid is just a maximally mixed state and a commitment to $(0, 0)$.
- For binding, we must show that the D register alone does not reveal $|\psi\rangle$. By collapse-binding security of the $|c_{r,s}\rangle$ commitment, the adversary cannot distinguish a superposition of valid decommitments from one in which the message (r, s) is measured. In particular, this means the adversary's decommitment D is indistinguishable from $\sum_{r,s}|r,s\rangle\langle r,s|\otimes|d_{r,s}\rangle\langle d_{r,s}|$. But this mixed state is just a uniform mixture of valid decommitments to uniformly random r, s and is therefore independent of $|\psi\rangle$.

We therefore conclude the following.

Theorem 2.5. *Hiding and binding QSCs exist if and only if hiding and binding QBCs exist (i.e., quantum-communication commitments to classical bits).*

A pair of recent works [MY22, AQY22] showed that hiding and binding QBCs — which were previously known assuming one-way functions — can be built from quantum assumptions that are potentially weaker than one-way functions [Kre21]. Consequently, hiding and binding QSCs exist under the same weak assumptions.

2.2.2 Beyond the folklore construction: succinct QSCs.

While we have successfully formalized the security of the folklore construction using swap binding, a definition really begins to shine when it enables new constructions. As mentioned in Section 1, the folklore construction requires the commitment C to be *at least as large* as the committed message. We will now see how to construct a *succinct* QSC, meaning a binding QSC for n-qubit messages where the commitment C is smaller than n qubits.[16]

[16] Notice that we have dropped the hiding property; as in the classical setting, many applications of succinct commitments do not explicitly require hiding. We remark that in the classical setting, succinct commitments are known to imply standard hiding-binding commitments (and thus one-way functions).

A succinct QSC from any collapsing hash function. A natural starting point is to assume the existence of collision-resistant hash functions, or rather, their post-quantum analogue: collapsing hash functions [Unr16]. Collapsing hash functions are succinct, collapse-binding commitments to *classical messages*. That is, given a hash function description $h \leftarrow H_\lambda$, a sender can succinctly commit to a message m by sending the classical string $h(m) = y$, and can open by revealing m.

How can we turn a succinct commitment h for classical messages into one for quantum messages? Our idea is simple: use h to commit to the quantum state in both the standard and Hadamard bases. Concretely, given a collapsing hash function $h : \{0,1\}^n \to \{0,1\}^{n/4}$, the sender commits to an n-qubit message M as follows:

1. Interpret M in the standard basis and evaluate h coherently, writing the output onto a fresh register C_0.
2. Apply the n-bit Hadamard transform $H^{\otimes n}$ to M.
3. Apply h again, this time writing the output onto another fresh register C_1. Finally, send the $(n/2)$-qubit state $\mathsf{C} = (\mathsf{C}_0, \mathsf{C}_1)$ as the commitment.
4. To decommit, the sender provides the $\mathsf{D} = \mathsf{M}$ register. This allows the receiver to uncompute the commitment, check validity, and recover the original message.

Our swap-binding definition makes it easy to prove this scheme secure assuming that h is collapse-binding. Recall that collapse-binding of h guarantees that if an adversary sends a superposition of inputs for h to a challenger, then it cannot distinguish between the following:

- the challenger evaluates h in superposition, measures the output, and then returns the input registers, or
- the challenger simply measures the input registers and returns them.

In our scheme, since the challenger keeps the register $\mathsf{C} = (\mathsf{C}_0, \mathsf{C}_1)$, from the adversary's point of view, both C_0 and C_1 are measured (in the standard basis). By carefully invoking collapse-binding security twice, once for each coherent evaluation of h in our scheme, we can show that the decommitment held by the adversary is indistinguishable from one that is generated as follows: measure M in the standard basis, apply $H^{\otimes n}$, and measure M again. But this gives a totally mixed state, and so the decommitment completely hides the original quantum message.

Succinct QSCs from even weaker assumptions. The above construction demonstrates that succinct QSCs can be built from collapsing hash functions (succinct classical commitments). Perhaps surprisingly, we show something stronger: succinct QSCs can even exist in a world where succinct classical commitments do not!

For instance, suppose we are only given a (post-quantum) one-way function, or equivalently, a (post-quantum) pseudorandom generator (PRG) $G : \{0,1\}^{n/2} \to \{0,1\}^{2n}$ [HILL99, Zha12]. We show that it is still possible to commit to an n-qubit state $|\psi\rangle$, as follows:

1. Initialize an $n/2$-qubit register to $|0\rangle$ and apply $H^{\otimes \frac{n}{2}}$ to obtain $2^{-n/4} \sum_{k \in \{0,1\}^{n/2}} |k\rangle$.

2. Controlled on k, apply $X^{G_0(k)}Z^{G_1(k)}$ to $|\psi\rangle$, where $G_0(k), G_1(k)$ are the first and last $n/2$ bits of $G(k)$, respectively. The resulting state is

$$\frac{1}{2^{n/4}} \sum_{k\in\{0,1\}^{n/2}} |k\rangle_{\mathsf{C}} X^{G_0(k)}Z^{G_1(k)} |\psi\rangle_{\mathsf{D}} ,$$

where the registers containing k are the succinct commitment C, and the one-time padded message is the decommitment D.

Our swap-binding definition makes it easy to prove this scheme secure:

- When the challenger holds the C register (i.e., it is traced out in the adversary's view), the decommitment for $|\psi\rangle$ is the mixed state resulting from applying a $X^{G_0(k)}Z^{G_1(k)}$ to $|\psi\rangle$ for a uniformly random k.
- By PRG security, this is indistinguishable from $X^r Z^s |\psi\rangle$ for uniformly random r, s, which is the maximally mixed state so swap-binding follows.

More generally, all we need is a one-time secure encryption scheme for n-qubit quantum messages with short keys. In fact, any scheme with keys shorter than n bits will suffice, since it turns out that using a quantum analogue of Merkle-Damgård domain extension (see Section 5.2), succinct QSCs with 1 qubit of compression can be composed to achieve any desired compression!

As we just saw, one-way functions give one possible instantiation of such an encryption scheme: to encrypt a state $|\psi\rangle$ with key k, just apply $X^{G_0(k)}Z^{G_1(k)}$. However, there are other instantiations from assumptions that may be even weaker than one-way functions. In particular, we show that one-time secure quantum encryption with short keys is implied by any pseudorandom unitary [JLS18] (see Claim 5.5), and the latter is known to be black-box separated from one-way functions [Kre21]. In summary, we have the following theorem.

Theorem 2.6 (Succinct QSCs from encryption)**.** *Assuming the existence of one-time secure encryption for n-bit quantum messages with $(n-1)$-bit classical keys, succinct QSCs exist. Such one-time encryption schemes are implied by one-way functions, but their existence is a potentially weaker assumption.*

2.3 Applications

So far, we have proposed a definition and argued that it captures the semantic meaning of binding for quantum messages. We have also shown several instantiations of schemes achieving this definition. However, there is still a crucial missing piece to our theory: can commitments satisfying our definition actually be used to achieve cryptographic ends?

In this section, we describe our flagship application: a three-message quantum succinct argument system based on succinct QSCs. Recall that a succinct argument is a protocol that enables a prover to convince a verifier that a claim is true (e.g., an NP statement), where the total communication and verifier runtime is extremely short. Our protocol is a direct quantum analogue of Kilian's classical succinct argument protocol [Kil92] and is based on a recent proposal of Chen and

Movassagh [CM21] (see Section 2.4 for more details). While the protocol is simple to state, proving security will be far more difficult. Even for Kilian's classical protocol, security against quantum attacks was a long-standing open question that was only recently settled by [CMSZ21]. Moreover, the [CMSZ21] technique fundamentally relies on the ability to copy information and is therefore limited to extracting classical information from classical protocols.

In this work, we resolve several challenges unique to the quantum setting and obtain a general framework for proving security of quantum protocols via *rewinding*. This is our most technically involved contribution. For this overview, we will present these techniques in the context of quantum succinct arguments (see Section 9 for applications to quantum sigma protocols).

2.3.1 The quantum succinct argument protocol

Kilian's protocol. Kilian's protocol [Kil92] is a compiler that uses succinct commitments to turn a probabilistically checkable proof (PCP) into a four-message interactive succinct argument. At the start of the protocol, the verifier samples a classical key for the succinct commitment, i.e., a hash function description h, and sends it to the prover. The prover uses h to generate a *tree commitment* c to the PCP and sends c to the verifier.[17] Next, the verifier samples random coins corresponding to the randomness of the PCP verifier. Finally, the prover sends *local openings* for the indices S specified by the random coins, which are at most $|S| \cdot \mathsf{poly}(\lambda)$ bits long in total. The verifier accepts if the PCP verifier accepts and the openings are valid.

The quantum protocol. Following Kilian's template, our quantum protocol is a compiler that uses succinct QSCs to turn any *classical or quantum* PCP into a quantum succinct argument. As mentioned earlier, the resulting protocol is very similar to a heuristic proposal of Chen and Movassagh [CM21] (see Section 2.4 for more details).

- **First message.** The prover generates a *quantum tree commitment* to its PCP using a succinct QSC that maps n-qubit messages to $n/2$-qubit commitments. The quantum tree commitment is computed as follows:

  1. The prover partitions the PCP into blocks of size $n/2$.
  2. Starting at the leaves, it creates a binary tree of commitments up to the root where each node register is a succinct commitment to its two children, and
  3. It sends the root commitment as the commitment to the whole PCP.

  We emphasize one difference from the classical setting: since the prover is committing to quantum states, the committed PCP is unavailable to the prover once it sends the root.

[17] Assuming $h : \{0,1\}^n \to \{0,1\}^{n/2}$, the tree commitment is generated as follows: (1) partition the PCP into blocks of size $n/2$, (2) create a binary tree of hash values where the PCP blocks form the leaves and each internal node is the hash of its two children, and (3) send the root as the commitment. The local opening for a PCP index i is a set consisting of (i) the PCP block containing i and (ii) the values of every node adjacent to the path from block containing i up to the root. The opening for a set of PCP indices S is the union of the sets corresponding to the decommitments for each $i \in S$.

- **Second message.** The verifier sends random coins corresponding to the PCP verifier randomness.
- **Third message.** The prover sends local openings for the PCP indices S specified by the verifier's challenge. For an individual leaf, the corresponding opening consists of every register on or adjacent to the root-to-leaf path (excluding the root itself). The opening for a set S consists of all registers involved in the opening for each $i \in S$. We refer the reader to Section 7.2 and Fig. 1 for further details.
- **Verification.** Starting from the root, the verifier works its way down the tree, using the provided openings to reveal each subsequent layer of commitments. Eventually, this reveals the qubits of the PCP at the locations in S. The verifier accepts if every individual QSC was opened correctly and the PCP verifier would accept.

This template is compatible with both classical and quantum PCPs. We emphasize that even when the PCP is classical, our protocol still requires the ability to commit to arbitrary *quantum states*. In particular, a tree commitment requires composing quantum commitments in sequence, i.e., committing to commitments, so even if the leaves are classical, the internal nodes are still commitments to quantum messages.

The goal for the remainder of this subsection will be to prove that this protocol is sound for any classical or quantum PCP and thereby establish the following theorem.

Theorem 2.7. *Assuming the existence of succinct QSCs, every* NP *language has a three-message quantum succinct argument. If we additionally assume the quantum PCP conjecture, this extends to* QMA*.*

2.3.2 Rewinding for quantum protocols

At a very high level, we will prove soundness of our protocol — or rather, a stronger property called argument of knowledge — by showing how to extract a convincing PCP from any malicious prover $\widetilde{P}$ that makes the verifier accept with noticeable probability. In the classical setting, this is achieved by a technique called *rewinding*: the extractor queries $\widetilde{P}$ many times on random challenges, rewinding $\widetilde{P}$ back to its earlier state after every query. After sufficiently many queries, the extractor will have enough accepting responses to stitch together a PCP.

It is well known that traditional rewinding proofs completely fail to capture quantum attacks [vdG97, Wat09, ARU14]: when the extractor runs $\widetilde{P}$ and records its response, it may irreversibly disturb $\widetilde{P}$'s state, rendering it useless for further queries. Fortunately, there has been a line of work showing how to rewind quantum adversaries for some important *classical* protocols [Wat09, Unr12, Unr16, CCY21, CMSZ21, LMS22b, LMS22a].

The work that turns out to be most relevant to our setting is [CMSZ21], which showed a general technique that enables repeatedly querying $\widetilde{P}$ and recording an arbitrary number of its accepting responses. At a high level, their technique consists of two steps:

- **Step 1: Reduce the task to "one-bit" extraction.** First, they reduce the problem to a simplified setting where the extractor only measures the bit indicating whether $\widetilde{P}$ succeeds

on a query (as opposed to measuring the entire response). This step invokes collapse-binding security [Unr12, Unr16] to argue that measuring this bit is computationally indistinguishable from measuring the entire response.

- **Step 2: One-bit extraction via state repair.** The primary technical innovation of [CMSZ21] is a procedure for one-bit extraction. One of their key insights was to realize that the extractor does not need to restore the malicious prover's original state, but rather *any state that gives it the same success probability as the original state*. Then, they designed a "state repair" procedure to restore the success probability of $\widetilde{P}$ after any one-bit disturbance. Their resulting extractor simply queries $\widetilde{P}$ repeatedly on random challenges and runs the state repair procedure in between each query.

In summary, the full [CMSZ21] extractor repeats the following "measure-and-repair" step: (1) run the adversary on a random challenge, (2) check whether the response is valid; if so, measure and record the response, and finally (3) run the repair procedure to restore the success probability.

Extracting quantum information. The [CMSZ21] extractor was designed for a different setting than ours: the goal of [CMSZ21] is to extract classical information from a quantum prover $\widetilde{P}$ in a classical protocol, whereas our goal is to extract *quantum information* from a quantum prover $\widetilde{P}$ in a quantum protocol.[18] In our setting it doesn't make sense to *measure* $\widetilde{P}$'s response, but our swap-binding definition strongly suggests a quantum analogue: swap out $\widetilde{P}$'s opened message and replace it with a junk $|0\rangle$ state!

While this idea may seem promising, it introduces a fundamental problem that was not present in the classical setting. By swapping $\widetilde{P}$'s message with junk, we have implicitly forced the adversary to *forget* part of its committed PCP. As a result, repairing the adversary is now *information-theoretically impossible*: if we could repair the adversary, the extractor would eventually generate multiple copies of the quantum PCP.

The swap-augmented prover. Fortunately, there is still a way to use [CMSZ21]-style state repair. To explain our idea, suppose that the extractor has swapped a single index i of the prover's PCP onto some internal register M'_i. Our idea is to define the following *swap-augmented prover*:

Let C denote the register holding the (root of the) tree commitment.

- The state of the swap-augmented prover consists of $\widetilde{P}$'s state augmented with the registers $(\mathsf{M}'_i, \mathsf{C})$.
- To run the swap-augmented prover on a set of positions S:
  1. First, run the real prover $\widetilde{P}$ on S to generate a response on some register Z. If $i \notin S$, we are done.
  2. If $i \in S$:

[18] Even when we instantiate our quantum succinct argument with a classical PCP, we will still need to extract quantum information to invoke security of the tree commitment.

(a) Use (C, Z) to recover the message (corresponding to indices S) in the clear.

(b) Next, apply $\mathsf{SWAP}_{\mathsf{M}_i,\mathsf{M}'_i}$ where M_i is the register containing the message opened by the prover for index i.

(c) Apply the inverse of Step 2a.

Roughly speaking, the swap-augmented prover is defined to output the responses that the prover *would have given* if the swap had not occurred. More generally, at every stage in the extraction procedure, we define a swap-augmented prover that makes use of all registers that have been swapped out so far.

At the beginning of the extraction, the swap-augmented prover is simply the original malicious prover $\widetilde{P}$, which convinces the verifier to accept with non-negligible probability p. Then after each query, instead of repairing $\widetilde{P}$ itself (which is impossible), our extractor will *run state repair on the swap-augmented prover*. At first glance, this may seem somewhat strange: unlike [CMSZ21], which only runs state repair on $\widetilde{P}$'s state, we give the state repair procedure the additional freedom to modify the extracted PCP itself. Nevertheless, we show the extractor works by arguing that there is eventually a point where:

- with very high probability, the swap-augmented prover answers any PCP query using only the augmented registers $\{\mathsf{M}'_i\}$, and
- the swap-augmented prover still has success probability $\approx p$.

If this happens, then the registers $\{\mathsf{M}'_i\}_i$ must contain a convincing PCP!

The oracle security lemma. It turns out to be quite subtle to argue that swap-binding security can actually be invoked in our extraction procedure. In [CMSZ21], collapse binding is used to argue that whenever the prover produces a superposition of valid responses, we can undetectably measure it and copy it. In our setting, we would like to argue that a swap operation can be applied undetectably.

However in a quantum commitment scheme, the adversarial committer does not have access to the commitment register after sending the commitment. But our extractor explicitly needs access to the commitment (for instance, the extractor must be able to run swap-augmented prover, which makes use of the commitment to generate its responses). Thus, it is completely unclear how we can argue security while running this extractor.

Our solution is the following:

- First, we show that swap binding generically implies a stronger security property we call *oracle swap binding*. Recall that in the swap-binding security experiment, the adversary's message in the $\mathsf{Com}^\dagger$ basis is either left alone on its original message register $\mathsf{M}^{(0)}$ or swapped onto a challenger register $\mathsf{M}^{(1)}$. What the adversary actually sees is the D register in the original basis.

  We define an oracle swap-binding experiment that gives the adversary additional power. In particular, it can access an oracle $\mathcal{O}$ that performs any efficient operation on $\mathsf{M}^{(b)}$ in the $\mathsf{Com}^\dagger$ basis, where b is the challenge bit. In other words, the operation is performed on the

register that contains the adversary's originally committed message. The adversary's goal in this experiment is the same as before, which is to guess b.

It is not a priori obvious that swap binding implies oracle swap binding, but we prove this in Section 6.

- We show that the swap operations performed in our extractor are undetectable to the adversary by reduction to oracle swap binding.

  Specifically, we prove that the extractor's access to the C register can be implemented with the oracle provided in this security game. At a high level, this oracle enables the extractor to (a) check whether an opening is valid and (b) check if the opened message passes PCP checks. These functionalities turn out to be sufficient to implement the [CMSZ21] repair procedure given the oracle but not the C register, which is what allows us to appeal to oracle swap-binding.

2.4 Related works

Comparison to [CM21]. The quantum succinct argument protocol we analyze is based on a proposal from Chen and Movassagh [CM21], who were the first to realize that a quantum-communication analogue of Kilian's protocol was even possible. The syntax for our tree commitments is similar to the syntax proposed in their work (modulo some changes to handle generic succinct QSCs).

However, [CM21] did not prove any security guarantees for their protocol, nor did they define the intermediate primitive of succinct QSCs. Instead they conjectured (but did not prove) that their protocol is secure when the commitments are modeled as *Haar random oracles*. In this work, we give a full proof of security *in the plain model* under very mild cryptographic assumptions. We additionally construct and analyze other protocols that do not rely on the quantum PCP conjecture (e.g., when our succinct argument template is instantiated with classical PCPs, we obtain a protocol that requires fewer messages and weaker assumptions than Kilian's protocol).

Computational soundness of [BG20]. The quantum sigma protocol we analyze is essentially the same as the one in [BG20]. However, the proof that the protocol is computationally sound (in fact, an argument of knowledge) is new to this work. While [BG20] originally proposed this claim, a recent version of their work [BG22] retracts it since it was missing a rewinding-based analysis, which is essential in the setting of computational soundness. In Section 9, we prove computational soundness using our new rewinding techniques. As an additional contribution, we also place their protocol in a general framework using our new abstraction of hiding and binding QSCs.

Comparison to [Mah18]. [Mah18] constructs a "weak quantum state commitment" with classical communication and a classical receiver. As Mahadev explains, her protocol is only a "weak" commitment because the commit phase of does not actually bind the prover to a fixed quantum state. Instead, she shows that the commitment *together with any malicious opening attack* fixes a state. While Mahadev's construction suffices for her applications, it is not a commitment in the typical sense and her techniques do not directly yield QSCs.

Prior work on quantum *bit* commitments. The basic syntax for quantum state commitments (QSCs) is similar to the "canonical form" described in prior work of Yan [Yan22]. However, Yan only considers commitments to *single-bit classical messages*, whereas we consider commitments to general quantum states. Even if we restrict to commitments to classical messages, there is one syntactic difference between our form and [Yan22]: Yan's "canonical form" assumes that the bit commitment scheme is specified by two circuits Q_0 and Q_1 and the committed message is revealed *in the clear* in the opening phase.

It was also well known from prior work on quantum bit commitments that quantum communication enables removing interaction [KO09, KO11, YWLQ15, BB21, Yan22]. The non-interactive compiler we describe in Appendix A) follows from very similar ideas and is included for the sake of completeness.

Other succinct arguments for QMA. [CCY20, BKL+22] construct certain variants of succinct arguments for QMA with classical communication. However, both works rely on extremely heavy cryptographic hammers including post-quantum indistinguishability obfuscation (for which there are only heuristic candidates). Our succinct argument for QMA is significantly simpler and can be instantiated from far weaker cryptographic assumptions at the cost of relying on the quantum PCP conjecture.

Authentication of quantum messages. Our hiding-binding duality is reminiscent of a well-known result of [BCG+02], which showed that any secure authentication scheme for quantum messages must also encrypt the message. However, there are several major differences between our results and [BCG+02]:

- While binding and authentication both consider adversaries that attempt to change some underlying message, the two notions capture different threat models. Binding guarantees security against the sender (who prepares the initial messages), whereas authentication guarantees security against an adversary who intercepts communication between the sender and receiver.
- There is no "duality" in the setting of authentication. In particular, authentication implies encryption, but *encryption does not imply authentication.*
- [BCG+02] only consider statistical security, whereas our definitions apply to both the computational and statistical settings.

In fact, our work highlights a potential limitation of the [BCG+02] definition: their definition does not say what happens to entangled messages.[19] For instance, suppose a sender prepares an entangled state on registers (A, B) and authenticates B. Is the receiver guaranteed to obtain a state that is still entangled with A? Their definition for mixed states [BCG+02, Appendix B] guarantees

[19] It is claimed on [BCG+02, Page 1] that their definition "implies security for mixed or entangled states," but what they mean is that the mixed state of the authenticated message is preserved (see [BCG+02, Appendix B]). This does not immediately rule out the possibility we raise here, in which the mixed state is preserved but the entanglement is broken.

that if the receiver accepts, the message it obtains is negligibly close to the sender's original message (i.e., fidelity $1 - \mathsf{negl}(\lambda)$). However, it does not say that this message remains entangled with A. While it is almost certainly true that the constructions in [BCG+02] preserve entanglement in this manner, it is not clear how to deduce this from the definition.

In our setting, proving such guarantees is made possible by the fact that our definition composes in parallel. In slightly more detail, if a sender commits to B and leaves A untouched, we can view this as a parallel composition of a trivially binding (identity) commitment to A and a commitment to B. The security of this larger commitment ensures that entanglement is preserved.

3 Preliminaries

Notation. The security parameter is written as λ. A function $f : \mathbb{N} \to [0,1]$ is *negligible* (denoted $f(\lambda) = \mathsf{negl}(\lambda)$) if it decreases faster than the inverse of any polynomial. A probability is *overwhelming* if is at least $1 - \mathsf{negl}(\lambda)$ for a negligible function $\mathsf{negl}(\lambda)$. For any nonzero $n \in \mathbb{N}$, let $[n] = \{1, 2, \ldots, n\}$ and let $[0]$ be the empty set. For a set R, let $r \leftarrow R$ be a uniformly random sample from R.

Quantum information preliminaries. Let $\mathcal{H}$ be a finite-dimensional Hilbert space. Let $\mathbf{S}(\mathcal{H})$ be the space of all Hermitian operators on $\mathcal{H}$. Let $\mathbf{D}(\mathcal{H})$ be set of all $\boldsymbol{\rho} \in \mathbf{S}(\mathcal{H})$ with $\mathrm{Tr}(\boldsymbol{\rho}) = 1$. A *pure quantum state* is a unit vector $|\psi\rangle \in \mathcal{H}$. A *density matrix* (and mixed quantum state) is an operator $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H})$. For any vector $|\phi\rangle \in \mathcal{H}$, we define its norm $\||\phi\rangle\| := \sqrt{\langle\phi|\phi\rangle}$. For an operator M on $\mathcal{H}$, we define its operator norm as

$$\|M\|_{op} := \max_{|\psi\rangle : \||\psi\rangle\| = 1} \|M\,|\psi\rangle\|.$$

We use the term quantum register to refer to a collection of qubits that we wish to treat as a single unit. These will be denoted with uppercase sans serif font, e.g., $\mathsf{A}, \mathsf{B}, \mathsf{C}$. Each register is associated with a finite-dimensional Hilbert space, denoted by writing the same letter in uppercase calligraphic font, e.g., $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are the spaces corresponding to registers $\mathsf{A}, \mathsf{B}, \mathsf{C}$. For an n-qubit register A, we sometimes use the shorthand $|0\rangle$ to denote the all-zero state $|0^n\rangle$ on A.

An *observable* is represented by a Hermitian operator O on H. When $O^2 = \mathbb{I}$ (equivalently, O has eigenvalues in $\{-1, 1\}$), O is also a *binary observable*. For binary observables O we define the projector onto its $+1$ eigenspace O^+ and the projector onto its -1 eigenspace O^- so that $O = O^+ - O^-$. For any projector Π, we define a corresponding binary projective measurement $\{\Pi, \mathbb{I} - \Pi\}$ and say that the measurement *accepts* if the outcome is Π; otherwise, we say that the measurement rejects.

General (non-unitary) evolution of a quantum state can be represented via a completely-positive trace-preserving (CPTP) map $\Phi : \mathbf{S}(\mathcal{H}) \to \mathbf{S}(\mathcal{H}')$. We define the diamond norm [Kit97, KSV02] of any CPTP map $\Phi : \mathbf{S}(\mathcal{H}) \to \mathbf{S}(\mathcal{H}')$ to be

$$\|\Phi\|_\diamond := \max_{\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H} \otimes \mathcal{X})} \left\| (\Phi \otimes \mathbb{I}_{\mathbf{S}(\mathcal{X})})(\boldsymbol{\rho}) \right\|_1$$

where $\mathcal{X}$ is any Hilbert space with the same dimension as $\mathcal{H}$, and $\|\cdot\|_1$ is the Schatten 1-norm.

We occasionally denote the action of a unitary U on a mixed state $\boldsymbol{\rho}$ by $U(\boldsymbol{\rho}) := U\boldsymbol{\rho}U^\dagger$.

Computational indistinguishability. The following is taken verbatim from [BKL+22]. Two quantum state ensembles $\{\boldsymbol{\rho}_0^{(\lambda)}, \boldsymbol{\rho}_1^{(\lambda)}\}_\lambda$ are said to be *computationally indistinguishable* if for every non-uniform QPT algorithm $A = \{A^{(\lambda)}, \boldsymbol{\rho}^{(\lambda)}\}$ (that outputs a bit b), we have that

$$\left| \mathbb{E}\left[A^{(\lambda)}(\boldsymbol{\rho}^{(\lambda)}, \boldsymbol{\rho}_0^{(\lambda)})\right] - \mathbb{E}\left[A^{(\lambda)}(\boldsymbol{\rho}^{(\lambda)}, \boldsymbol{\rho}_1^{(\lambda)})\right] \right| = \mathsf{negl}(\lambda).$$

Equivalently, $\{\boldsymbol{\rho}_0^{(\lambda)}, \boldsymbol{\rho}_1^{(\lambda)}\}_\lambda$ are computationally indistinguishable if for every efficiently computable non-uniform binary observable $(R, \boldsymbol{\sigma})$, we have that

$$\left| \mathrm{Tr}(R(\boldsymbol{\rho}_0 \otimes \boldsymbol{\sigma})) - \mathrm{Tr}(R(\boldsymbol{\rho}_1 \otimes \boldsymbol{\sigma})) \right| = \mathsf{negl}(\lambda).$$

We will occasionally use the notation $\boldsymbol{\rho}_0 \approx_c \boldsymbol{\rho}_1$ to denote computational indistinguishability of $\{\boldsymbol{\rho}_0^{(\lambda)}, \boldsymbol{\rho}_1^{(\lambda)}\}_\lambda$. When $\boldsymbol{\rho}_0$ and $\boldsymbol{\rho}_1$ are statistically indistinguishable, i.e. $\|\boldsymbol{\rho}_0 - \boldsymbol{\rho}_1\|_1 = \mathsf{negl}(\lambda)$, we write $\boldsymbol{\rho}_0 \approx_s \boldsymbol{\rho}_1$.

4 Defining quantum state commitments (QSCs)

4.1 Syntax

We describe the syntax for non-interactive quantum state commitments.

Definition 4.1 (Non-interactive QSC syntax)**.** *A non-interactive quantum state commitment* QSC *is specified by a unitary* $\mathsf{Com}_{\mathsf{QSC}}$ *on* $n + \lambda$ *qubits.*

- *(Commitment phase) To commit to a* n*-qubit state in register* M*, the sender initializes register* W *to the* $|0^\lambda\rangle$ *state and applies* $\mathsf{Com}_{\mathsf{QSC}}$ *to* (M, W)*. The resulting state is divided into registers* (C, D)*. It sends* C *to the receiver.*
- *(Opening phase) The sender decommits by sending* D*. The receiver applies* $\mathsf{Com}^\dagger_{\mathsf{QSC}}$ *to the pair* (C, D)*, obtaining* (M, W)*. It verifies the decommitment by measuring* W *in the standard basis and checking that the outcome is* $|0^\lambda\rangle$*. If verification accepts, the receiver interprets the state on* M *as the committed message.*

For convenience, we will often write Com in place of $\mathsf{Com}_{\mathsf{QSC}}$.

What about interactive commitments? Our swap-binding definition (Definition 4.2) will only apply to non-interactive QSCs. This is justified in part by the fact that any interactive QSC can be made non-interactive via a simple transformation that was previously known in the setting of quantum *bit* commitments (see Appendix A for full details). Nevertheless, we believe it could be useful in some settings to have a definition that directly handles the interactive case. Towards this end, we state an alternative definition of binding that we call "Pauli binding" in Appendix B.1. Pauli binding naturally handles interactive QSCs and is equivalent to swap-binding for non-interactive QSCs. We choose to present swap binding as our main definition since we found it significantly easier to work with.

4.2 The swap binding definition

We now present our new swap binding security definition. For two quantum registers of the same size, M and M', $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$ denotes the unitary that maps $|\psi\rangle_{\mathsf{M}} |\phi\rangle_{\mathsf{M}'}$ to $|\phi\rangle_{\mathsf{M}} |\psi\rangle_{\mathsf{M}'}$.

Definition 4.2 (Swap binding)**.** *For a non-interactive quantum commitment scheme* Com*, an interactive adversary* A*, a challenge bit* $b \in \{0, 1\}$*, and a security parameter* λ*, define the swap binding security experiment* $\mathtt{SwapBindExpt}_{\mathsf{QSC},A,b}(\lambda)$ *as follows.*

$\mathtt{SwapBindExpt}_{\mathsf{QSC},A,b}(\lambda)$*:*

1. *The adversary* A *(acting as a malicious sender) sends commitment register* C *and decommitment register* D *to the challenger (acting as an honest receiver).*
2. *The challenger applies* $\mathsf{Com}^\dagger$ *and measures* $\{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$ *on* W*; if the measurement rejects, then it aborts and outputs a random* $b' \leftarrow \{0, 1\}$*.*
3. *The challenger does the following:*

- *if* $b = 0$*, the challenger simply applies* Com*, and sends the* D *register to* A*;*
- *if* $b = 1$*, the challenger replaces the contents of the* M *register with* $|0\rangle$ *(i.e., it initializes a fresh register* $\mathsf{E} = |0\rangle$ *of the same dimension as* M *and applies* $\mathsf{SWAP}[\mathsf{M}, \mathsf{E}]$*). It then applies* Com *and sends the* D *register to* A*.*

4. *The output of the experiment is* $b' \leftarrow A$*.*

Com *is computationally (resp. statistically) swap binding if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries* A*,*

$$\Pr_{b \leftarrow \{0,1\}}[\mathtt{SwapBindExpt}_{\mathsf{QSC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

4.3 Additional definitions: hiding and succinctness

We now define hiding and succinctness for QSCs. While the notion of hiding for *commitments* to quantum messages had not been defined in the literature before, it is easy to write down a definition based on existing definitions for *encryption* of quantum messages. The following definition is essentially the same as the definition of indistinguishability under chosen-plaintext-attacks (q-IND-CPA) given by [BJ15].

Definition 4.3 (Hiding for quantum state commitments)**.** *For a non-interactive quantum commitment scheme* Com*, an interactive adversary* A*, a challenge bit* $b \in \{0, 1\}$*, and a security parameter* λ*, define a security experiment* $\mathtt{HideExpt}_{\mathsf{QSC},A,b}(\lambda)$ *as follows.*

$\mathtt{HideExpt}_{\mathsf{QSC},A,b}(\lambda)$*:*

1. A *prepares a message* M *and sends it to the challenger.*
2. *Next, the challenger (acting as an honest sender) performs the commit phase with* A *(acting as the receiver) using the quantum state in* M *if* $b = 0$*, or* $|0\rangle$ *if* $b = 1$*.*
3. *The output of the experiment is* $b' \leftarrow A$*.*

Com *is computationally (resp. statistically) hiding if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries* A*,*

$$\Pr_{b \leftarrow \{0,1\}}[\mathtt{HideExpt}_{\mathsf{QSC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Definition 4.4 (Succinct non-interactive QSCs)**.** *We say that a non-interactive QSC is* succinct *if the commitment is shorter than the message, i.e., the size of register* C *is smaller than* M*.*

We remark that our definition of succinct QSCs does not require hiding. This is analogous to the definitions of succinct classical commitments (i.e., collision-resistant hash functions, or their post-quantum analogue, collapsing hash functions [Unr16]), which only require binding and succinctness but not hiding. Moreover, hiding is not necessary for our primary application to quantum succinct arguments (Section 7).

4.4 Hiding-binding duality

We observe that the hiding and binding security experiments are nearly identical for non-interactive QSCs. In particular, we can write both of the experiments as follows.

Property 4.5 (Hiding-binding duality for QSCs)**.** *For a non-interactive quantum commitment scheme* Com*, an interactive adversary A, a challenge bit $b \in \{0,1\}$, and a security parameter λ, the binding and hiding security experiments can be written as follows.*

1. *The adversary A sends (C,D) to the challenger.*
2. *The challenger applies $\mathsf{Com}^\dagger$ and measures $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort the experiment (i.e., output a random bit b'). Otherwise:*
   (a) *The challenger initializes a register* E *to $|0\rangle$ and applies* $\mathsf{SWAP}[\mathsf{M},\mathsf{E}]^b$*, then* Com*.*
   (b) *For binding, the challenger returns* D*; for hiding, the challenger returns* C*.*
3. *The output of the experiment is $b' \leftarrow A$.*

Proof. The binding game is exactly equivalent to Definition 4.2. The hiding game is equivalent to Definition 4.3, except that here the adversary sends (C,D) whereas the adversary in Definition 4.3 sends M (and there is no W measurement). However, these notions of hiding are equivalent by essentially the same argument as [Unr16, Lemma 14]. The proof follows by observing that sending (C,D) registers containing an invalid commitment-decommitment pair cannot help the adversary; if the adversary sends (C,D) registers containing a valid commitment-decommitment pair, this is equivalent to sending M (in the Com basis). □

Property 4.5 motivates the definition of a *dual scheme*, where the roles of the C and D registers are reversed.

Definition 4.6 (Dual commitments)**.** *For a non-interactive commitment scheme* Com *with commitment register* C *and decommitment register* D*, the* dual scheme *is the commitment scheme with the same commitment unitary* Com*, but* D *is used as the commitment and* C *is used as the decommitment.*

As an immediate consequence of Property 4.5, we obtain the following corollary.

Corollary 4.7. *Let* Com *be a computationally (resp. statistical) swap-binding and statistically (resp. computationally) hiding non-interactive commitment scheme. Then the dual commitment scheme* Com *is a computationally (resp. statistical) hiding and statistically (resp. computationally) swap-binding non-interactive commitment scheme.*

4.4.1 Discussion: is there a hiding-binding duality for QBCs?

As it turns out, a similar duality also exists for quantum bit(string) commitments (QBCs), i.e., quantum-communication commitments to *classical messages*. Previously, [Yan22] showed that it

was possible to transform any statistically binding, computationally hiding QBC into a computationally binding, statistically hiding QBC (or vice versa). However, using the insights developed above for QSCs, we can make this connection between hiding and binding significantly simpler.

For the following discussion, we consider QBCs satisfying our syntax for non-interactive QSCs. This is essentially Yan's "canonical form" [Yan22] with a few syntactic differences (most notably, we allow for arbitrary-length messages, not just bits).

Consider the following version of Unruh's collapse-binding definition [Unr16] for *classical messages* (restated for QBCs):

> Computational (resp. statistical) security requires that any efficient (resp. inefficient) A guesses b with probability at most $1/2 + \mathsf{negl}(\lambda)$ in the following experiment:

1. The adversary A sends (C, D) to the challenger.
2. The challenger applies $\mathsf{Com}^\dagger$ and measures $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort the experiment (i.e., output a random bit b'). Otherwise:
   (a) The challenger samples a random bitstring s (whose length equals the number of qubits of M) and applies Z^s to M.
   (b) Next, the challenger applies Com and returns D to the adversary.
3. The adversary wins the game if it guesses b.

Security requires that it is hard to guess b with advantage noticeably greater than $1/2$. How does this security experiment relate to hiding? Suppose we modify the collapse-binding experiment as follows (differences in red):

> Computational (resp. statistical) security requires that any efficient (resp. inefficient) A guesses b with probability at most $1/2 + \mathsf{negl}(\lambda)$ in the following experiment:

1. The adversary A sends (C, D) to the challenger.
2. The challenger applies $\mathsf{Com}^\dagger$ and measures $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort the experiment (i.e., output a random bit b'). Otherwise:
   (a) The challenger samples a random bitstring s (whose length equals the number of qubits of M) and applies X^s to M.
   (b) Next, the challenger applies Com and returns C to the adversary.
3. The adversary wins the game if it guesses b.

It is not hard to see that this definition, which we call "X-hiding," implies the commitment hides any *classical message*. In particular, if the adversary sends any classical message $|m\rangle$ (i.e., by initializing (C, D) to $\mathsf{Com}(|m\rangle\,|0\rangle_{\mathsf{W}})$) then the game above amounts to distinguishing a commitment to m from a commitment to a uniformly random string. However, X-hiding is stronger than ordinary hiding for classical messages, since the adversary may choose a superposition of classical messages $|m\rangle$.

This connection between X-hiding and collapse-binding (which we will call Z-binding here) leads to a simple dual scheme for any QBC satisfying our non-interactive commitment syntax. In particular, suppose Com is statistically Z-binding and computationally X-hiding. Now consider the following "dual" commitment scheme:

- To commit to an n-bit classical string $|m\rangle$, apply $H^{\otimes n}$ followed by Com. Send the decommitment register D as the commitment.
- To decommit, send the commitment C. To verify, apply $H^{\otimes n}\mathsf{Com}^\dagger$ and recover the $|m\rangle$ from the M register.

It is immediate from these definitions that this scheme scheme is *computationally* Z-binding and statistically X-hiding!

4.5 Parallel composition

In this section, we show that our binding definition for quantum messages composes in parallel.

Theorem 4.8 (Binding composes in parallel)**.** *Let* $\{\mathsf{Com}_i\}_{i\in[n]}$ *be a set of* n *non-interactive binding commitment schemes with message and ancillary registers* $(\mathsf{M}_i, \mathsf{W}_i)_{i\in[n]}$*. Then the parallel composition of the commitment schemes* $\widetilde{\mathsf{Com}} := \bigotimes \mathsf{Com}_i$ *with message register* $\mathsf{M} := (\mathsf{M}_i)_{i\in[n]}$ *and ancilla register* $\mathsf{W} := (\mathsf{W}_i)_{i\in[n]}$ *is also a binding commitment scheme.*

Proof. We will show by a sequence of $n+1$ hybrids that the parallel composition of the n commitment schemes is binding. That is, we will show that an adversary cannot distinguish between decommitments for messages of their choice in all n slots or for $|0\rangle$ in all n slots. We define hybrids H_j for $j \in \{0, 1, \ldots, n\}$ as follows.

Hybrid H_j:

1. The adversary sends n commitment and decommitment registers $(\mathsf{C}_i, \mathsf{D}_i)_{i\in[n]}$ to the challenger.
2. The challenger then does the following:
   (a) Apply $\widetilde{\mathsf{Com}}^\dagger$.
   (b) Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement accepts, then continue to the next step. Otherwise, the experiment aborts and a random bit $b' \leftarrow \{0,1\}$ is output.
   (c) Swap the messages in M_i for $i \leq j$ with $|0\rangle$.
   (d) Apply $\widetilde{\mathsf{Com}}$.
3. The challenger returns the decommitment $(\mathsf{D}_i)_{i\in[n]}$ to the adversary.

Breaking binding for the commitment scheme $\widetilde{\mathsf{Com}}$ is equivalent to distinguishing between hybrids H_0 and H_n. Suppose that an adversary A can distinguish hybrids H_0 and H_n with advantage ε. Then for some $j \in [n]$, the adversary A can distinguish hybrid H_{j-1} from hybrid H_j with advantage ε/n. We now use A to construct an adversary A' on the binding experiment for the commitment scheme Com_j. The adversary A' receives the registers $(\mathsf{C}_i, \mathsf{D}_i)_{i=1}^n$ from A, then does the following:

1. Apply $\widetilde{\mathsf{Com}}^\dagger$.
2. Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement accepts, then continue to the next step. Otherwise, output a random bit $b' \leftarrow \{0,1\}$.
3. Swap the messages in M_i for $i \leq j-1$ with $|0\rangle$.
4. Apply $\widetilde{\mathsf{Com}}$.
5. Forward $(\mathsf{C}_j, \mathsf{D}_j)$ to the challenger, receive D_j from the challenger, and send $(\mathsf{D}_i)_{i=1}^n$ to A.
6. Output $b' \leftarrow A$.

By construction of A', the view of A in which the challenger either does nothing or swaps out the jth message with $|0\rangle$ is identical to that of H_{j-1} and H_j, respectively, so A' wins the binding experiment for Com_j with advantage ε/n. □

5 Constructions of QSCs

5.1 Succinct QSCs from one-time quantum encryption

Definition 5.1 (Quantum encryption syntax)**.** *A quantum encryption scheme* QEnc *for quantum message space* M *and classical key space* $\{0,1\}^{d(\lambda)}$ *is specified by a family* $\{U_k\}_{k\in\{0,1\}^{d(\lambda)}}$ *of unitaries acting on* M*, for each security parameter* λ*. We require that for any* $k \in \{0,1\}^{d(\lambda)}$*, there is a* $\mathsf{poly}(\lambda)$*-time classical procedure to generate a description of* U_k*. The encryption of* $|\psi\rangle \in \mathsf{M}$ *under secret key* k *is* $U_k |\psi\rangle$*.*

To simplify notation, we will often drop the dependence on λ and write $\mathsf{QEnc} = \{U_k\}_k$. We say that a quantum encryption scheme has *short keys* if $d < n$. We define *one-time security* for a quantum encryption scheme QEnc with the following security game:

Definition 5.2 (One-time secure quantum encryption)**.** *For a quantum encryption scheme* QEnc*, an interactive adversary* A*, a challenge bit* $b \in \{0,1\}$*, and a security parameter* λ*, define a security experiment* $\mathtt{QEncExpt}_{\mathsf{QSC},A,b}(\lambda)$ *as follows.*

$\mathtt{QEncExpt}_{\mathsf{QEnc},A,b}(\lambda)$*:*

1. *A sends a quantum message* M *to the challenger.*
2. *The challenger samples a random* $k \leftarrow \{0,1\}^{d(\lambda)}$ *and does the following:*
   (a) *If* $b = 0$*, apply* U_k *to* M *and send* M *to* A*.*
   (b) *If* $b = 1$*, swap the contents of* M *with* $|0\rangle$*, apply* U_k *to* M*, and send* M *to* A*.*
3. *The output of the experiment is* $b' \leftarrow A$*.*

QEnc *is one-time secure if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time quantum adversaries* A*,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{QEncExpt}_{\mathsf{QEnc},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Using any one-time secure quantum encryption scheme $\mathsf{QEnc} = \{U_k\}_k$, we can define a non-interactive QSC by the following commitment unitary:

$$\mathsf{Com} = \sum_{k\in\{0,1\}^d} \left(|k\rangle\langle k|\, H^{\otimes d}\right)_{\mathsf{W}} \otimes (U_k)_{\mathsf{M}}, \tag{1}$$

using $\mathsf{C} = \mathsf{W}$ as the commitment and $\mathsf{D} = \mathsf{M}$ as the decommitment. For example, the honest commitment to the pure state $|\psi\rangle_{\mathsf{M}}$ is

$$2^{-d/2} \sum_{k\in\{0,1\}^d} |k\rangle_{\mathsf{C}} \otimes (U_k |\psi\rangle)_{\mathsf{D}}.$$

If the encryption scheme has short keys, then the scheme is succinct. It is easy to show that the scheme is also binding.

Theorem 5.3. *If $\{U_k\}_k$ is a one-time secure quantum encryption scheme, then the QSC scheme* Com *defined by Eq. (1) is binding.*

Proof. The proof is immediate from the definitions. Consider the binding security experiment for Com (Definition 4.2). We denote the register containing the adversary's internal state by R. Let $\boldsymbol{\rho}_{\mathsf{C,D,R}}$ be the state of the experiment after the adversary sends D to the challenger, let $\boldsymbol{\rho}^b_{\mathsf{D,R}}$ be the state held by the adversary after it receives D back from the challenger in the $b \in \{0,1\}$ world of the binding experiment, and let $\sigma_{\mathsf{M,R}} = \mathrm{Tr}_{\mathsf{C}}(\Pi \cdot \mathsf{Com}^\dagger \boldsymbol{\rho}_{\mathsf{C,D,R}} \mathsf{Com} \cdot \Pi)$ be the sub-normalized state on (M,R) after a successful opening. By the one-time security of the quantum encryption scheme,

$$\begin{aligned}
\boldsymbol{\rho}^0_{\mathsf{D,R}} &= 2^{-d} \sum_{k\in\{0,1\}^d} (U_k)_{\mathsf{M}} \sigma_{\mathsf{M,R}} (U_k^\dagger)_{\mathsf{M}} \\
&\approx_c 2^{-d} \sum_{k\in\{0,1\}^d} \left(U_k \left|0\right\rangle\!\left\langle 0\right|_{\mathsf{M}} U_k^\dagger\right)_{\mathsf{M}} \otimes \sigma_{\mathsf{R}} \\
&= \boldsymbol{\rho}^1_{\mathsf{D,R}}.
\end{aligned}$$

□

Claim 5.4 (One-time secure quantum encryption with short keys from OWFs)**.** *One-time secure quantum encryption with short keys exists assuming the existence of any post-quantum one-way function.*

Proof. There exists a post-quantum pseudo-random generator $G : \{0,1\}^{n/2} \to \{0,1\}^{2n}$ assuming the existence of any post-quantum one-way function [HILL99, Zha12]. Let $U_k = X^{G_0(k)} Z^{G_1(k)}$ be a quantum encryption scheme, where $G_0(k), G_1(k)$ are the first and last $n/2$ bits of $G(k)$, respectively. This scheme has key size $d = n/2$ for n-qubit messages, and security is immediate from the definitions.

Consider the security experiment for $\{U_k\}_k$ (Definition 5.2). We denote the register containing the adversary's internal state by R. Let $\boldsymbol{\rho}_{\mathsf{M,R}}$ be the state of the experiment after the adversary sends M to the challenger and let $\boldsymbol{\rho}^b_{\mathsf{M,R}}$ be the state held by the adversary after it receives M back from the challenger in the $b \in \{0,1\}$ world of the security experiment. Since G is a post-quantum pseudo-random generator,

$$\begin{aligned}
\boldsymbol{\rho}^0_{\mathsf{M,R}} &= 2^{-d} \sum_{k\in\{0,1\}^d} (X^{G_0(k)} Z^{G_1(k)})_{\mathsf{M}} \boldsymbol{\rho}_{\mathsf{M,R}} (Z^{G_1(k)} X^{G_0(k)})_{\mathsf{M}} \\
&\approx_c 2^{-d} \sum_{k\in\{0,1\}^d} (X^{G_0(k)} Z^{G_1(k)} \left|0\right\rangle\!\left\langle 0\right|_{\mathsf{M}} Z^{G_1(k)} X^{G_0(k)})_{\mathsf{M}} \otimes \boldsymbol{\rho}_{\mathsf{R}} \\
&= \boldsymbol{\rho}^1_{\mathsf{M,R}}.
\end{aligned}$$

□

Claim 5.5. *There is a quantum oracle $\mathcal{O}$ relative to which one-time secure quantum encryption with short keys exists, but* $\mathrm{BQP}^{\mathcal{O}} = \mathrm{QMA}^{\mathcal{O}}$ *(and in particular, post-quantum one-way functions do not exist).*

Proof. By [Kre21, Theorem 2], there is a quantum oracle $\mathcal{O}$ relative to which a pseudo-random unitary family $\{U_k : \mathbb{C}^{2^n} \to \mathbb{C}^{2^n}\}_{k\in\{0,1\}^n}$ exists, but $\mathrm{BQP}^{\mathcal{O}} = \mathrm{QMA}^{\mathcal{O}}$. We would like to view the

family as a quantum encryption scheme with keys $k \in \{0,1\}^n$, but this scheme does not have short keys (as $d = n$).

However, Appendix C shows that we can exchange the PRU family's many-time security for one-time security on a larger message space. We therefore use $\{\mathsf{Expand}(U_k, \ell)\}_{k\in\{0,1\}^n}$ for any constant $\ell > 1$, which is a secure one-time quantum encryption scheme acting on at least $\ell \cdot n$ qubits by Theorem C.1. □

5.2 Domain extension

In this section, we show that a quantum analogue of Merkle-Damgård domain extension for classical compressing hash functions works for succinct QSCs.

Concretely, suppose we are given a *succinct* QSC scheme sQSC for $(m+1)$-qubit messages that produces m-qubit commitments.[20] That is, $\mathsf{Com}_{\mathsf{sQSC}}$ acts on message-ancilla registers (M, W) and produces commitment-decommitment registers (C, D) where:

- M is an $(m+1)$-qubit message register,
- W is a d-qubit ancilla register,
- C is an m-qubit commitment register, and
- D is a $(d+1)$-qubit decommitment register.

Using sQSC, we construct a new succinct QSC scheme $\mathsf{sQSC}_k^{\mathsf{MD}}$ (where MD stands for Merkle-Damgård) enabling commitments to k-qubit messages with commitments of m qubits for any $k = \mathsf{poly}(\lambda)$.

In $\mathsf{sQSC}_k^{\mathsf{MD}}$, the sender commits to a k-qubit message M as follows:

> **Setup:** Initialize an $(m + kd)$-qubit register W to $|0^{m+kd}\rangle$. Separate W into subregisters $\mathsf{W} = (\mathsf{C}_0, \mathsf{Y}_0, \mathsf{Y}_1, \ldots, \mathsf{Y}_{k-1})$ where C_0 is m qubits and each Y_i is d qubits. Separate M into subregisters $\mathsf{M} = (\mathsf{M}_0, \mathsf{M}_1, \ldots, \mathsf{M}_{k-1})$ where each M_i is a single qubit register.

1. For $i = 0, \ldots, k-1$:

   > Apply $\mathsf{Com}_{\mathsf{sQSC}}$ to $(\mathsf{M}_i, \mathsf{C}_i, \mathsf{Y}_i)$, i.e., commit to the $(m+1)$-qubit message register $(\mathsf{M}_i, \mathsf{C}_i)$ using the d-qubit ancilla Y_i, to obtain $(\mathsf{C}_{i+1}, \mathsf{D}_{i+1})$.

2. The commitment is the m-qubit register $\mathsf{C} = \mathsf{C}_k$ and the decommitment is the $k\cdot(d+1)$-qubit register $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$.

Theorem 5.6 (Domain extension). *For any $k = \mathsf{poly}(\lambda)$, the scheme $\mathsf{sQSC}_k^{\mathsf{MD}}$ is swap binding assuming that* sQSC *is swap binding.*

[20] One-qubit compression is without loss of generality, since if a scheme compresses by more than one qubit, we can pad the input with 0's (and additionally require that these qubits are 0 during verification).

Proof. We prove this claim by a sequence of hybrid arguments. To simplify notation, let U_j be the unitary corresponding to the jth application of $\mathsf{Com}_{\mathsf{sQSC}}$ in $\mathsf{sQSC}_k^{\mathsf{MD}}$, so that the commitment is implemented with the unitary $U := U_k U_{k-1} \cdots U_1$. We first state all of the hybrids below and then prove indistinguishability of the hybrids in Claims 5.7 to 5.9. We highlight the differences from the previous hybrid in red.

- **Hybrid H_0**:
  1. The adversary sends the registers (C, D) to the challenger.
  2. The challenger does the following:
     (a) Apply a binary projective measurement to check that (C, D) is valid.
     (b) Return $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

Hybrid H_0 corresponds to the $b = 0$ case of the QSC binding security game Definition 4.2.

- **Hybrid H_1**:
  1. The adversary sends (C, D) to the challenger.
  2. The challenger does the following:
     (a) Apply a binary projective measurement to check that (C, D) is valid.
     (b) Apply $U^\dagger$ to recover the committed message on registers $(\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$.
     (c) For each $i = 0, \ldots, k-1$:
        i. Initialize an m-qubit register C'_i to $|0^m\rangle$.
        ii. Use Com to commit to the $m+1$-qubit register $(\mathsf{M}_i, \mathsf{C}'_i)$, producing a decommitment D_{i+1} and a commitment C_{i+1}; discard C_{i+1}.
     (d) Return $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

The difference between H_1 and H_0 is that in H_1, the decommitments $\mathsf{D}_1, \ldots, \mathsf{D}_k$ are each generated *in parallel*. That is, in hybrid H_1, each D_i the adversary receives is the decommitment resulting from committing to $(\mathsf{M}_{i-1}, \mathsf{C}'_{i-1})$ where C'_{i-1} is initialized to $|0^m\rangle$. In contrast, in hybrid H_0, the decommitments D_i are generated *in sequence*, i.e., D_i is the decommitment resulting from committing to $(\mathsf{M}_{i-1}, \mathsf{C}_{i-1})$, where C_{i-1} is the commitment resulting from committing to $(\mathsf{M}_{i-2}, \mathsf{C}_{i-2})$ in the previous layer, etc.

We prove that H_0 and H_1 are indistinguishable in Claim 5.7 by defining a careful sequence of k sub-hybrids.

- **Hybrid H_2**:
  1. The adversary sends (C, D) to the challenger.
  2. The challenger does the following:
     (a) Apply a binary projective measurement to check that (C, D) is valid.
     (b) Apply $U^\dagger$ to recover the committed message on register $\mathsf{M} = (\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$.
     (c) Initialize k qubits $\mathsf{M}' = (\mathsf{M}'_0, \ldots, \mathsf{M}'_{k-1})$ to $|0^k\rangle$, and apply $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$.

(d) For each $i = 0, \ldots, k-1$:
   i. Initialize an m-qubit register C'_i to $|0^m\rangle$.
   ii. Use Com to commit to the $m+1$-qubit register $(\mathsf{M}_i, \mathsf{C}'_i)$ and obtain $(\mathsf{D}_{i+1}, \mathsf{C}_{i+1})$; discard C_{i+1}.

(e) Return $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

The difference between H_2 and H_1 is that in H_2, the decommitments $\mathsf{D}_1, \ldots, \mathsf{D}_k$ are each decommitments resulting from (independent) commitments to the message $|0^{m+1}\rangle$. Since the k decommitments in both H_1 and H_2 are all independent, indistinguishability of these hybrids follows from the fact that swap-binding composes in parallel (Theorem 4.8); see Claim 5.8.

- **Hybrid H_3**:

1. The adversary sends (C, D) to the challenger.
2. The challenger does the following:
   (a) Apply a binary projective measurement to check that (C, D) is valid.
   (b) Apply $U^\dagger$ to recover the committed message on registers $(\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$.
   (c) Initialize k qubits $\mathsf{M}' = (\mathsf{M}'_0, \ldots, \mathsf{M}'_{k-1})$ to $|0^k\rangle$, and apply $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$.
   (d) Apply U.
   (e) Return $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

Hybrid H_3 corresponds to the $b = 1$ case of the QSC binding security game (Definition 4.2). The differences between H_3 and H_2 are analogous to the difference between H_0 and H_1, except that now each M_i is replaced with $|0\rangle$. We prove $H_3 \approx_c H_2$ in Claim 5.9 using essentially the same arguments as in Claim 5.7.

It remains to prove Claims 5.7 to 5.9. □

Claim 5.7. *Hybrid H_0 is indistinguishable from Hybrid H_1.*

Proof. We prove this claim by a sequence of $k+1$ hybrids $H_0^{(0-1)}, \ldots, H_k^{(0-1)}$. Hybrid $H_0^{(0-1)}$ is identical to H_0, and $H_k^{(0-1)}$ is identical to H_1. For $i = 0, \ldots, k$ we define $H_i^{(0-1)}$ by the following behavior for the challenger:

Hybrid $H_i^{(0-1)}$:

1. The adversary sends (C, D).
2. The challenger does the following:
   (a) Check that the commitment is valid. If not, the experiment aborts.
   (b) If $i > 0$, open the last i commitments using $U_{k-i+1}^\dagger \cdots U_k^\dagger$, initialize a new register $(\mathsf{M}'_{k-i}, \mathsf{C}'_{k-i})$ to $|0\rangle$, and apply $\mathsf{SWAP}[(\mathsf{M}_{k-i}, \mathsf{C}_{k-i}), (\mathsf{M}'_{k-i}, \mathsf{C}'_{k-i})]$.
   (c) If $i > 1$, for $j = k-i, \ldots, k-2$ in sequence, apply Com to $(\mathsf{M}_j, \mathsf{C}_j, \mathsf{Y}_j)$ to produce a decommitment on D_{j+1} and a commitment on C_{j+1}, initialize a new register $(\mathsf{M}'_{j+1}, \mathsf{C}'_{j+1})$ to $|0\rangle$, and apply $\mathsf{SWAP}[(\mathsf{M}_{j+1}, \mathsf{C}_{j+1}), (\mathsf{M}'_{j+1}, \mathsf{C}'_{j+1})]$.

(d) If $i > 0$, apply Com to $(\mathsf{M}_{k-1}, \mathsf{C}_{k-1}, \mathsf{Y}_{k-1})$ to produce a decommitment on D_k and a commitment on C_k.

3. The challenger returns the decommitment $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

Suppose that an adversary A can distinguish hybrids $H_0^{(0-1)}$ and $H_k^{(0-1)}$ with advantage ε. Then for some $i \in \{0, \ldots, k-1\}$, the adversary A can distinguish hybrid $H_i^{(0-1)}$ from hybrid $H_{i+1}^{(0-1)}$ with advantage ε/k. We now construct a reduction $R_i^{(0-1)}$ that uses such an adversary to achieve advantage ε/k in the binding experiment for the commitment scheme Com.

Reduction $R_i^{(0-1)}$:

1. The adversary sends (C, D) to $R_i^{(0-1)}$.
2. The reduction does the following:
   (a) Check that the commitment is valid. If not, the experiment aborts and the reduction makes a random guess $b' \leftarrow \{0,1\}$.
   (b) If $i > 0$, open the last i commitments using $U_{k-i+1}^\dagger \cdots U_k^\dagger$. Initialize an internal register $(\mathsf{M}'_{k-i}, \mathsf{C}'_{k-i})$ to $|0\rangle$ and apply $\mathsf{SWAP}[(\mathsf{M}_{k-i}, \mathsf{C}_{k-i}), (\mathsf{M}'_{k-i}, \mathsf{C}'_{k-i})]$.
   (c) If $i > 1$, for $j = k-i, \ldots, k-2$ in sequence, apply Com to $(\mathsf{M}_j, \mathsf{C}_j, \mathsf{Y}_j)$ to obtain $(\mathsf{D}_{j+1}, \mathsf{C}_{j+1})$. Initialize an internal register $(\mathsf{M}'_{j+1}, \mathsf{C}'_{j+1})$ to $|0\rangle$ and then apply $\mathsf{SWAP}[(\mathsf{M}_{j+1}, \mathsf{C}_{j+1}), (\mathsf{M}'_{j+1}, \mathsf{C}'_{j+1})]$.
   (d) If $i > 0$, apply Com to $(\mathsf{M}_{k-1}, \mathsf{C}_{k-1}, \mathsf{Y}_{k-1})$ to produce a decommitment on D_k and a commitment on C_k. For convenience of notation, let $\mathsf{C}'_k = \mathsf{C}_k$.
   (e) Send $(\mathsf{D}_{k-i}, \mathsf{C}'_{k-i})$ to the challenger. According to the challenge bit b, the challenger applies SWAP^b to an auxiliary $|0\rangle$ state and the message $(\mathsf{M}_{k-i-1}, \mathsf{C}_{k-i-1})$ underlying $(\mathsf{D}_{k-i}, \mathsf{C}'_{k-i})$. Then the challenger returns D_{k-i} to the reduction.
3. The reduction returns the decommitment $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

The view of the adversary in the $b = 0$ world of $R_i^{(0-1)}$ is exactly the same as in $H_i^{(0-1)}$, and the view of the adversary in the $b = 1$ world is exactly the same as in $H_{i+1}^{(0-1)}$. □

Claim 5.8. *Hybrid H_1 is indistinguishable from Hybrid H_2.*

Proof. It is immediate that a QSC is binding to every subset of registers of the message; the same fact for collapse-binding commitments was proven in [Unr16, Lemma 15] and the proof for QSCs is nearly identical. In particular, Com is a binding commitment to M_i for each $i \in \{0, \ldots, k-1\}$. Therefore, H_1 and H_2 are indistinguishable by the following reduction to the security of parallel composition (Theorem 4.8).

- **Reduction R_1:**

1. The adversary sends (C, D) to R.
2. The reduction does the following:
   (a) Apply a binary projective measurement to check that (C, D) is valid. If not, the experiment aborts and the reduction makes a random guess $b' \leftarrow \{0,1\}$.

(b) Apply $U^\dagger$ to recover the committed message on registers $(\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$

(c) For each $i = 0, \ldots, k-1$:

i. Initialize an m-qubit register C'_i to $|0^m\rangle$.

ii. Apply Com to the $m+1$-qubit register $(\mathsf{M}_i, \mathsf{C}'_i, \mathsf{Y}_i)$, producing a decommitment D_{i+1} and a commitment C_{i+1}.

(d) Send $\{\mathsf{D}_i, \mathsf{C}_i\}_{i\in[k]}$ to the challenger.

(e) The challenger samples a random $b \leftarrow \{0,1\}$, applies

$$\bigotimes_{i\in[k]} U_i(\mathsf{SWAP}[\mathsf{M}_i, \mathsf{M}'_i]^b)U_i^\dagger,$$

and returns $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the reduction.

(f) The reduction returns D to the adversary. □

Claim 5.9. *Hybrid H_2 is indistinguishable from Hybrid H_3.*

Proof. The proof is identical to the proof of Claim 5.7, except that we view the commitments as swap-binding commitments to just the registers C_i instead of $(\mathsf{M}_i, \mathsf{C}_i)$. That is, we only swap C_i instead of $(\mathsf{M}_i, \mathsf{C}_i)$, and only make use of the binding property of the commitment scheme on the C_i register.

Formally, we again use a sequence of $k+1$ hybrids $H_0^{(3-2)}, \ldots, H_k^{(3-2)}$. Hybrid $H_0^{(3-2)}$ is identical to H_3, and $H_k^{(3-2)}$ is identical to H_2. For $i = 0, \ldots, k$ we define $H_i^{(3-2)}$ by the following behavior for the challenger:

Hybrid $H_i^{(3-2)}$:

1. The adversary sends (C, D).
2. The challenger does the following:
   (a) Check that the commitment is valid. If not, the experiment aborts.
   (b) Apply $U^\dagger$ to recover the committed message on registers $(\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$.
   (c) Initialize k qubits $\mathsf{M}' = (\mathsf{M}'_0, \ldots, \mathsf{M}'_{k-1})$ to $|0^k\rangle$ and apply $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$.
   (d) Apply U.
   (e) If $i > 0$, open the last i commitments using $U_{k-i+1}^\dagger \cdots U_k^\dagger$, initialize a new register C'_{k-i} to $|0\rangle$, and apply $\mathsf{SWAP}[\mathsf{C}_{k-i}, \mathsf{C}'_{k-i}]$.
   (f) If $i > 1$, for $j = k-i, \ldots, k-2$ in sequence, apply Com to $(\mathsf{M}_j, \mathsf{C}_j, \mathsf{Y}_j)$ to produce a decommitment on D_{j+1} and a commitment on C_{j+1}, initialize a new register C'_{j+1} to $|0\rangle$, and apply $\mathsf{SWAP}[\mathsf{C}_{j+1}, \mathsf{C}'_{j+1}]$.
   (g) If $i > 0$, apply Com to $(\mathsf{M}_{k-1}, \mathsf{C}_{k-1}, \mathsf{Y}_{k-1})$ to produce a decommitment on D_k and a commitment on C_k.
3. The challenger returns the decommitment $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

Suppose that an adversary A can distinguish hybrids $H_0^{(3-2)}$ and $H_k^{(3-2)}$ with advantage ε. Then for some $i \in \{0, \ldots, k-1\}$, the adversary A can distinguish hybrid $H_i^{(3-2)}$ from hybrid $H_{i+1}^{(3-2)}$ with advantage ε/k. We now construct a reduction $R_i^{(3-2)}$ that uses such an adversary to achieve advantage ε/k in the binding experiment for the commitment scheme Com.

Reduction $R_i^{(3-2)}$:

1. The adversary sends (C, D) to $R_i^{(3-2)}$.
2. The reduction does the following:
   (a) Check that the commitment is valid. If not, the experiment aborts and the reduction makes a random guess $b' \leftarrow \{0,1\}$.
   (b) Apply $U^\dagger$ to recover the committed message on registers $(\mathsf{M}_0, \ldots, \mathsf{M}_{k-1})$.
   (c) Initialize k qubits $\mathsf{M}' = (\mathsf{M}'_0, \ldots, \mathsf{M}'_{k-1})$ to $|0^k\rangle$ and apply $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$.
   (d) Apply U.
   (e) If $i > 0$, open the last i commitments using $U_{k-i+1}^\dagger \cdots U_k^\dagger$. Initialize an internal register C'_{k-i} to $|0\rangle$ and apply $\mathsf{SWAP}[\mathsf{C}_{k-i}, \mathsf{C}'_{k-i}]$.
   (f) If $i > 1$, for $j = k-i, \ldots, k-2$ in sequence, apply Com to $(\mathsf{M}_j, \mathsf{C}_j, \mathsf{Y}_j)$ to obtain $(\mathsf{D}_{j+1}, \mathsf{C}_{j+1})$. Initialize an internal register C'_{j+1} to $|0\rangle$ and then apply $\mathsf{SWAP}[\mathsf{C}_{j+1}, \mathsf{C}'_{j+1}]$.
   (g) If $i > 0$, apply Com to $(\mathsf{M}_{k-1}, \mathsf{C}_{k-1}, \mathsf{Y}_{k-1})$ to produce a decommitment on D_k and a commitment on C_k. For convenience of notation, let $\mathsf{C}'_k = \mathsf{C}_k$.
   (h) Send $(\mathsf{D}_{k-i}, \mathsf{C}'_{k-i})$ to the challenger. According to the challenge bit b, the challenger applies SWAP^b to an auxiliary $|0\rangle$ state and the register C_{k-i-1} underlying $(\mathsf{D}_{k-i}, \mathsf{C}'_{k-i})$. Then the challenger returns D_{k-i} to the reduction.
3. The reduction returns the decommitment $\mathsf{D} = (\mathsf{D}_1, \ldots, \mathsf{D}_k)$ to the adversary.

The view of the adversary in the $b = 0$ world of $R_i^{(3-2)}$ is exactly the same as in $H_i^{(3-2)}$, and the view of the adversary in the $b = 1$ world is exactly the same as in $H_{i+1}^{(3-2)}$. □

5.3 Formalizing the folklore: QSCs from QBCs

In this section, we formalize the security of the "folklore" construction of *hiding* QSCs: to commit to an n-qubit quantum state $|\psi\rangle$, sample two random n-bit classical strings r, s, and send $X^r Z^s |\psi\rangle$ together with commitments to the classical string (r, s). The opening is just the opening to the commitment to (r, s), which enables the receiver to recover $|\psi\rangle$ by applying $Z^s X^r$. We instantiate the commitment to (r, s) with any quantum-communication commitment to classical messages satisfying hiding and binding.

Defining quantum bit(string) commitments. The syntax for a quantum bit(string) commitment (QBC) is identical to our syntax for *non-interactive* quantum state commitments (Definition 4.1).

Hiding and binding are defined as follows:

Definition 5.10 (Hiding for QBCs). *For a quantum bit(string) commitment $\mathsf{Com}_{\mathsf{QBC}}$, an interactive adversary A, a challenge bit $b \in \{0,1\}$, and a security parameter λ, define a security experiment $\mathtt{HideExpt}_{\mathsf{QBC},A,b}(\lambda)$ as follows.*

$\mathtt{HideExpt}_{\mathsf{QBC},A,b}(\lambda)$*:*

1. *The adversary A sends a two messages $x_0, x_1 \in \{0,1\}^{|\mathsf{M}|}$ to the challenger.*

2. *The challenger (acting as the honest sender) initializes register M to the standard basis state $|x_b\rangle$ and register W to $|0\rangle$, applies $\mathsf{Com}_{\mathsf{QBC}}$, and sends register C to the adversary.*

3. *The output of the experiment is $b' \leftarrow A$.*

$\mathsf{Com}_{\mathsf{QBC}}$ *is computationally (resp. statistically) hiding if there exists a negligible function $\mu(\lambda)$ such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries A,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{HideExpt}_{\mathsf{QBC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Definition 5.11 (Binding for QBCs [Unr16])**.** *For a quantum bit(string) commitment $\mathsf{Com}_{\mathsf{QBC}}$, an interactive adversary A, a challenge bit $b \in \{0,1\}$, and a security parameter λ, define a binding security experiment $\mathtt{ColBindExpt}_{\mathsf{QBC},A,b}(\lambda)$ as follows. This definition is an adaptation of Unruh's collapse-binding definition [Unr16] to the setting of quantum bit commitments.*

$\mathtt{ColBindExpt}_{\mathsf{QBC},A,b}(\lambda)$*:*

1. *The adversary A (acting as a malicious committer) sends registers (C,D) to the challenger.*

2. *The challenger applies $\mathsf{Com}^{\dagger}_{\mathsf{QBC}}$ and measures $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$; if the decommitment is invalid (the measurement rejects), it aborts the experiment (i.e., outputs a random b').*

3. *The challenger does the following:*

   - *if $b = 0$, the challenger simply applies $\mathsf{Com}_{\mathsf{QBC}}$, and sends the D register to A;*
   - *if $b = 1$, the challenger measures M in the standard basis. It then applies $\mathsf{Com}_{\mathsf{QBC}}$ and sends the D register to A.*

4. *The output of the experiment is $b' \leftarrow A$.*

$\mathsf{Com}_{\mathsf{QBC}}$ *is computationally (resp. statistically) collapse-binding if there exists a negligible function $\mu(\lambda)$ such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries A,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{ColBindExpt}_{\mathsf{QBC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Non-interactive form of the folkore construction. Before proving security of the commitment to an n-qubit state, we write the folklore construction in non-interactive form for a commitment to one qubit. Let $\mathsf{Com}_{\mathsf{QBC}}$ be a quantum bitstring commitment to *two* classical bits. That is, it maps two bits in registers $\mathsf{M}' := (\mathsf{M}'_1, \mathsf{M}'_2)$ and ancilla register W' to commitment register C' and decommitment register D'. Recall that in the folklore construction, the sender commits to two randomly samples bits, r and s. To purify this step, we initialize two additional single-qubit

registers $\mathsf{K} := (\mathsf{K}_1, \mathsf{K}_2)$, and make $(\mathsf{K}_1, \mathsf{M}'_1)$ and $(\mathsf{K}_2, \mathsf{M}'_2)$ each hold one EPR pair. Formally, the construction is as follows:

To commit to a qubit in message register M, a sender will initialize ancillary registers $\mathsf{W} := (\mathsf{K}_1, \mathsf{K}_2, \mathsf{M}'_1, \mathsf{M}'_2, \mathsf{W}')$ to the $|0\rangle$ state. Then the sender applies the following commitment unitary:

$$\mathsf{Com}_{\mathsf{QSC}} = \mathsf{Com}_{\mathsf{QBC}}(\mathsf{ctl}_{\mathsf{M}'_2}\text{-}Z_{\mathsf{M}})(\mathsf{ctl}_{\mathsf{M}'_1}\text{-}X_{\mathsf{M}}) \left((\mathsf{ctl}_{\mathsf{K}_1}\text{-}X_{\mathsf{M}'_1})H_{K_1} \otimes (\mathsf{ctl}_{\mathsf{K}_2}\text{-}X_{\mathsf{M}'_2})H_{\mathsf{K}_2}\right) . \tag{2}$$

The commitment register is $\mathsf{C} := (\mathsf{M}, \mathsf{C}')$ and decommitment register is $\mathsf{D} := (\mathsf{K}_1, \mathsf{K}_2, \mathsf{D}')$. Thus, for any quantum message $|\psi\rangle_{\mathsf{M}}$,

$$\mathsf{Com}_{\mathsf{QSC}}\left(|\psi\rangle_{\mathsf{M}} \otimes |0\rangle_{\mathsf{KM'W'}}\right) = \frac{1}{2} \sum_{r,s\in\{0,1\}} |r,s\rangle_{\mathsf{K}} \otimes \left(\mathsf{Com}_{\mathsf{QBC}} |r,s\rangle_{\mathsf{M}'} \otimes |0\rangle_{\mathsf{W}'}\right) \otimes Z^s X^r |\psi\rangle_{\mathsf{M}} .$$

Theorem 5.12. *The quantum state commitment scheme in Eq. (2),* $\mathsf{Com}_{\mathsf{QSC}}$*, is computationally (resp. statistical) hiding and statistically (resp. computational) binding if the underlying quantum bit commitment scheme,* $\mathsf{Com}_{\mathsf{QBC}}$*, is computationally (resp. statistical) hiding and statistically (resp. computational) binding.*

Proof. Hiding of $\mathsf{Com}_{\mathsf{QSC}}$ follows by a standard hybrid argument. To simplify notation, for $(r,s) \in \{0,1\}^2$, let $\tau_{r,s}$ denote the density matrix corresponding to the commitment to (r,s) under $\mathsf{Com}_{\mathsf{QBC}}$.

- H_0: In this hybrid, the adversary Adv sends a single qubit state $\boldsymbol{\rho}_{\mathsf{M}}$ to the challenger. The challenger returns the mixed state

  $$\frac{1}{4}\sum_{r,s} \tau_{r,s} \otimes (X^r Z^s)\boldsymbol{\rho}_{\mathsf{M}}(Z^s X^r),$$

  corresponding to an honest commitment of $\boldsymbol{\rho}_{\mathsf{M}}$ under $\mathsf{Com}_{\mathsf{QSC}}$. This hybrid is the view of the adversary in the $b = 0$ case of the QSC hiding experiment (Definition 4.3).

- H_1: This hybrid is the same as H_0, except the commitment to (r,s) is replaced by a commitment to $(0,0)$.

  $$\tau_{0,0} \otimes \frac{1}{4}\sum_{r,s}(X^r Z^s)\boldsymbol{\rho}_{\mathsf{M}}(Z^s X^r).$$

  This is indistinguishable from H_1 by the hiding security of $\mathsf{Com}_{\mathsf{QBC}}$: an adversary in the hiding experiment can sample random $r, s \leftarrow \{0,1\}^2$, send (r,s) and $(0,0)$ to the challenger, receive the state $\tilde{\tau}$ (which is either $\tau_{r,s}$ or $\tau_{0,0}$), and prepare the state $\frac{1}{4}\sum_{r,s} \tilde{\tau} \otimes (X^r Z^s)\boldsymbol{\rho}_{\mathsf{M}}(Z^s X^r)$.

- H_2: This hybrid is the same as H_1, except the challenger uses $|0\rangle\langle 0|$ instead of $\boldsymbol{\rho}_{\mathsf{M}}$:

  $$\tau_{0,0} \otimes \frac{1}{4}\sum_{r,s}(X^r Z^s)\,|0\rangle\langle 0|_{\mathsf{M}}\,(Z^s X^r).$$

  This is identical to H_1 since the state in both hybrids is $\tau_{0,0} \otimes (\mathbb{I}/2)$.

- H_3: This hybrid is the same as H_2, except the commitment to $(0,0)$ is replaced by a commitment to (r,s):

$$\frac{1}{4}\sum_{r,s}\tau_{r,s}\otimes(X^rZ^s)\,|0\rangle\langle 0|_{\mathsf{M}}\,(Z^sX^r).$$

This follows by the hiding property of $\mathsf{Com}_{\mathsf{QBC}}$. This completes the proof of hiding for $\mathsf{Com}_{\mathsf{QSC}}$ since H_3 is the view of the adversary in the $b=1$ case of the $\mathsf{Com}_{\mathsf{QSC}}$ hiding experiment (Definition 4.3).

We now show that $\mathsf{Com}_{\mathsf{QSC}}$ is swap binding (Definition 4.2) assuming $\mathsf{Com}_{\mathsf{QBC}}$ is binding (Definition 5.11).

- Hybrid H_0:

  1. The adversary Adv sends a quantum state on registers C,D to the challenger.
  2. The challenger then does the following:
     (a) Apply $\mathsf{Com}^\dagger_{\mathsf{QSC}}$ to (C,D).
     (b) Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I}-|0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort.
     (c) Apply $\mathsf{Com}_{\mathsf{QSC}}$.
     (d) Send the decommitment register D to Adv.

  The view of Adv corresponds to the view of the adversary in the $b=0$ world of the swap-binding experiment for $\mathsf{Com}_{\mathsf{QSC}}$ (Definition 4.2).

- Hybrid H_1:

  1. The adversary Adv sends a quantum state on registers C,D to the challenger.
  2. The challenger then does the following:
     (a) Apply $\mathsf{Com}^\dagger_{\mathsf{QSC}}$ to (C,D).
     (b) Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I}-|0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort.
     (c) Apply $\mathsf{Com}_{\mathsf{QSC}}$.
     (d) Apply $\mathsf{Com}^\dagger_{\mathsf{QBC}}$ to reveal the two-bit committed classical message on register M'.
     (e) Measure M' in the standard basis.
     (f) Apply $\mathsf{Com}_{\mathsf{QBC}}$.
     (g) Send the decommitment register D to Adv.

  The measurement in Step 2b ensures that after applying $\mathsf{Com}_{\mathsf{QSC}}$ in Step 2c, the resulting state includes a valid commitment-decommitment pair for the message in M' under $\mathsf{Com}_{\mathsf{QBC}}$. By the binding security of $\mathsf{Com}_{\mathsf{QBC}}$, the act of measuring the contents of M' in the standard basis is undetectable without access to the commitment C'. Thus, hybrid H_1 is indistinguishable from H_0; otherwise, an adversary in the $\mathsf{Com}_{\mathsf{QBC}}$ binding experiment receives (C,D) from Adv, performs Steps 2a to 2c, sends $(\mathsf{C}',\mathsf{D}')$ to the challenger, receives D' from the challenger, sends D to Adv, and uses Adv's output to win the $\mathsf{Com}_{\mathsf{QBC}}$ binding experiment.

- Hybrid H_2:

  1. The adversary Adv sends registers (C, D) to the challenger.
  2. The challenger then does the following:
     (a) Apply $\mathsf{Com}^{\dagger}_{\mathsf{QSC}}$ to (C, D) to obtain (M, W).
     (b) Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort.
     (c) Replace the contents of M with the state $|0\rangle\langle 0|_{\mathsf{M}}$.
     (d) Apply $\mathsf{Com}_{\mathsf{QSC}}$.
     (e) Apply $\mathsf{Com}^{\dagger}_{\mathsf{QBC}}$ to reveal the two-bit committed classical message on register M'.
     (f) Measure M' in the standard basis.
     (g) Apply $\mathsf{Com}_{\mathsf{QBC}}$.
     (h) Send the decommitment register D to Adv.

  For $(r, s) \in \{0,1\}^2$, let $\gamma_{r,s}$ denote the density matrix corresponding to the decommitment to (r, s) under the quantum commitment to classical messages $\mathsf{Com}_{\mathsf{QBC}}$. By a straightforward computation, regardless of the the contents of M after Step 2b, the state that the challenger sends Adv is

  $$\frac{1}{4}\sum_{r,s} |r,s\rangle\langle r,s|_{\mathsf{K}} \otimes (\gamma_{r,s})_{\mathsf{D}'}.$$

  Thus H_2 is perfectly indistinguishable from H_1.

- Hybrid H_3:

  1. The adversary Adv sends registers (C, D) to the challenger.
  2. The challenger then does the following:
     (a) Apply $\mathsf{Com}^{\dagger}_{\mathsf{QSC}}$ to the (C, D) registers.
     (b) Measure $\{|0\rangle\langle 0|_{\mathsf{W}}, \mathbb{I} - |0\rangle\langle 0|_{\mathsf{W}}\}$. If the measurement rejects, abort.
     (c) Replace the contents of M with the state $|0\rangle\langle 0|_{\mathsf{M}}$.
     (d) Apply $\mathsf{Com}_{\mathsf{QSC}}$.
     ~~Apply $\mathsf{Com}^{\dagger}_{\mathsf{QBC}}$ to reveal the two-bit committed classical message on register M'.~~
     ~~Measure M' in the standard basis.~~
     ~~Apply $\mathsf{Com}_{\mathsf{QBC}}$.~~
     (e) Send the decommitment register D to Adv.

  The difference between hybrids H_2 and H_3 is that challenger measures M' in the former but not in the latter. Again, by the binding security of $\mathsf{Com}_{\mathsf{QBC}}$, the two hybrids are indistinguishable. This completes the proof of binding for $\mathsf{Com}_{\mathsf{QSC}}$ because H_3 is the view of the adversary in the $b = 1$ world of the binding experiment (Definition 4.2).

□

The proof of security above shows that the folklore construction applied to one qubit is a hiding and binding QSC. By appealing to the security of the parallel composition of n such commitment schemes using Theorem 4.8, we get a full QSC for n-qubit quantum messages.

Corollary 5.13. *Let* $\mathsf{Com}_{\mathsf{QSC}}$ *be the QSC defined by Eq.* (2)*. Then for any* $n = \mathsf{poly}(\lambda)$*, the* n*-fold parallel composition* $\bigotimes_{i=1}^{n} \mathsf{Com}_{\mathsf{QSC}}$ *is a hiding and binding QSC if the underlying quantum commitment to classical messages* $\mathsf{Com}_{\mathsf{QBC}}$ *is hiding and binding.*

6 The oracle security lemma

In this section we prove that swap binding is equivalent to a seemingly stronger definition that we call oracle swap binding. As discussed in Section 2.3.2, oracle swap binding plays a crucial role in our rewinding-based security proofs (Section 8).

Recall that in the swap-binding security experiment, the challenger's bit b determines whether the adversary's originally committed message is left on the M register (when $b = 0$) or swapped into the challenger's M' register (when $b = 1$). Of course, the standard swap-binding adversary is unable to access either of these registers: the state on M can only be revealed by applying $\mathsf{Com}^\dagger$ to (C, D) (which the adversary cannot do since the challenger does not return the C register) and the contents of the M' register remain entirely out of the adversary's view. Roughly speaking, the oracle swap binding experiment gives the adversary an oracle that can perform *arbitrary efficient operations on the originally committed message*, i.e., M if $b = 0$ or M' if $b = 1$.

To state the definition formally, we use the following notation:

- For any operator O and commitment unitary Com, we write $\widehat{O} := (\mathsf{Com})O(\mathsf{Com}^\dagger)$.
- Let $\Pi := \mathsf{Com}(\mathbb{I}_\mathsf{M} \otimes |0\rangle\langle 0|_\mathsf{W})\mathsf{Com}^\dagger$ denote the projection onto valid commitment-decommitment states.
- Recall that for a pair of equal-size registers M, M', the unitary $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$ maps $|\psi\rangle_\mathsf{M} |\phi\rangle_{\mathsf{M}'}$ to $|\phi\rangle_\mathsf{M} |\psi\rangle_{\mathsf{M}'}$.

In the oracle swap binding experiment, the adversary specifies an arbitrary unitary $\mathcal{O}$, which acts on M and any number of additional ancilla qubits R. After the adversary sends (C, D) to its challenger and gets back a state on D, it is allowed to perform the following oracle call as many times as desired:

1. The adversary sends registers (R, D) to the oracle.
2. If $b = 0$, the oracle applies the unitary $\widehat{\mathcal{O}}\Pi + (\mathbb{I} - \Pi)$ to the registers $(\mathsf{C}, \mathsf{D}, \mathsf{R})$. If $b = 1$, the oracle applies the unitary $\widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']} \cdot \widehat{\mathcal{O}} \cdot \widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']}\Pi + (\mathbb{I} - \Pi)$.
3. The oracle returns (R, D) to the adversary.

We can write this oracle more compactly as

$$G_b := (\widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']})^b \cdot \widehat{\mathcal{O}} \cdot (\widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']})^b \Pi + (\mathbb{I} - \Pi).$$

When (C, D) contains a valid commitment (i.e., the state is in the image of Π), the oracle behaves as expected: in the $\mathsf{Com}^\dagger$ basis, it applies $\mathcal{O}$ to (M, R) if $b = 0$, or to $(\mathsf{M}', \mathsf{R})$ if $b = 1$. To ensure that the oracle does not enable trivial attacks, we will restrict the oracle's behavior to be independent of b whenever (C, D) is an invalid commitment. In particular, we require that the oracle behave as identity on any state in $(\mathbb{I} - \Pi)$.

Definition 6.1 (Oracle swap-binding for quantum state commitments)**.** Com *is computationally (resp. statistically) oracle swap-binding if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time (resp. unbounded-time) oracle swap-binding adversaries specified by an efficient (resp. inefficient) interactive adversary* A *and any efficient (resp. inefficient) unitary* $\mathcal{O}$*,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{SwapBindExpt}_{\mathsf{QSC},A^{G_b},b}(\lambda)=b]\leq\frac{1}{2}+\mu(\lambda).$$

Theorem 6.2. *If* Com *is computationally (resp. statistically) swap-binding (Definition 4.2), then* Com *is computationally (resp. statistically) oracle swap-binding (Definition 6.1).*

The rest of this section is devoted to proving Theorem 6.2.

6.1 The admissible oracle lemma

Instead of proving Theorem 6.2 directly, we will consider a more abstract distinguishing task that we call the (W,Π)-distinguishing game, parameterized by a binary observable W and a projection Π that commutes with W. We state and prove a lemma we call the *admissible oracle lemma* (Lemma 6.5); in Section 6.2, we show how Theorem 6.2 follows from a special case of the admissible oracle lemma.

Definition 6.3 ((W,Π)-distinguishing game)**.** *Let* (A,B) *be two quantum registers. Let* W *be a binary observable and* Π *be a projector on* (A,B) *such that* Π *commutes with* W*. Consider the following distinguishing game:*

1. *The adversary sends a quantum state on registers* (A,B) *to the challenger.*
2. *The challenger chooses a random bit* $b\leftarrow\{0,1\}$*. Next, it measures measures* $\{\Pi,\mathbb{I}-\Pi\}$*; if the measurement rejects, abort and output a random bit* $b'\leftarrow\{0,1\}$*. Otherwise, the challenger applies* W^b *to* (A,B)*, and returns* B *to the adversary.*
3. *The adversary outputs a guess* b'*.*

We define the distinguishing advantage of the adversary to be $|\Pr[b'=b]-1/2|$*.*

The main result of the section is that, if the (W,Π)-distinguishing game is hard, i.e., the best possible distinguishing advantage is negligible, then it remains hard when the adversary is given oracle access to any *admissible* unitary G, defined as follows.

Definition 6.4. *We say that a unitary* G *is* admissible *if:*

- G *commutes with both* W *and* Π*, and*
- G *acts identically on* $\mathbb{I}-\Pi$*, i.e.,* $G(\mathbb{I}-\Pi)=\mathbb{I}-\Pi$*.*

Informally, the admissible oracle lemma (Lemma 6.5) says the following. Suppose that a quantum adversary with an initial state $|\psi\rangle$ wins the (W,Π)-distinguishing game with advantage ε, making t oracle calls to G. Then there exists a quantum adversary that wins the (W,Π)-distinguishing

game with advantage $\varepsilon^2/8t^2$, but does not make any calls to G. The latter adversary requires an initial state $\boldsymbol{\rho}$ which can be prepared using a circuit Prep^G that runs G as a subroutine. We emphasize that Prep^G does not require special oracle access and can be implemented directly, since it is run before the adversary sends any registers to the challenger.

Lemma 6.5 (Admissible oracle lemma)**.** *Let G be an admissible unitary. Suppose that for some quantum state $|\psi\rangle_{\mathsf{ABR}}$, there is a quantum algorithm D^G (acting only on registers (B,R)) with oracle access to G that achieves a distinguishing advantage ε in the (W,Π)-distinguishing game with initial state $|\psi\rangle$. Then there exists another quantum state $\boldsymbol{\rho}$ and an algorithm E making no queries to G that wins the (W,Π)-distinguishing game with advantage $\frac{\varepsilon^2}{8t^2}$ using initial state $\boldsymbol{\rho}$, where t is an upper bound on the number of queries D makes to G. Moreover,*

- *if D^G is implemented with a size-T oracle circuit (counting oracle queries as size 1), then $\boldsymbol{\rho} \leftarrow \mathsf{Prep}^{G,M}(|\psi\rangle)$, where $\mathsf{Prep}^{G,M}$ is a $\mathsf{poly}(T)$ size oracle circuit with (controlled) queries to G, $G^\dagger$, and $M := (W^+\Pi)\otimes X + (\mathbb{I} - W^+\Pi)\otimes \mathbb{I}$, and*
- *E is $\mathsf{poly}(T)$ size.*

We will prove Lemma 6.5 in two steps:

1. In Lemma 6.6, we show that the distinguishing task is equivalent to the following *mapping task*: given any state in $\mathsf{image}(W^+\Pi)$, output any state in $\mathsf{image}(W^-\Pi)$.
2. In Lemma 6.8, we show that for the *mapping task*, admissible oracles only provide limited help.

For a distinguisher D implemented as a binary projective measurement $D = \{\Pi_D, \mathbb{I} - \Pi_D\}$, the distinguishing advantage on (potentially sub-normalized) $|\psi\rangle$ can be written as

$$\frac{1}{2}|\langle\psi|\,\Pi\Pi_D\Pi\,|\psi\rangle - \langle\psi|\,\Pi W\Pi_D W\Pi\,|\psi\rangle|.$$

For a unitary U and (potentially sub-normalized) state $|\psi\rangle$, we define the *mapping advantage* of U on $|\psi\rangle$ to be

$$\left\|(W^-\Pi)U(W^+\Pi)\,|\psi\rangle\right\|^2$$

The following lemma states that the mapping and distinguishing tasks are equivalent in a setting where the adversary supplies the initial state (up to a quadratic loss in the advantage).

Lemma 6.6. *Let W be a binary observable on a register H, let Π be a projection on H that commutes with W, and let B be a fresh single-qubit register. Fix a state $|\psi\rangle$ on H and let $|\psi_+\rangle := W^+\Pi\,|\psi\rangle$. Then:*

(i) If U has mapping advantage ε on $|\psi\rangle$, then $\{\widetilde{\Pi}_D, \mathbb{I} - \widetilde{\Pi}_D\}$ has distinguishing advantage $\varepsilon/2$ on $\mathsf{ctl}_{\mathsf{B}}U\,|+\rangle_{\mathsf{B}}\,|\psi_+\rangle$, where $\widetilde{\Pi}_D := \mathsf{ctl}_{\mathsf{B}}U\,|+\rangle\langle+|_{\mathsf{B}}\,\mathsf{ctl}_{\mathsf{B}}U^\dagger$.

(ii) If Π_D has distinguishing advantage ε on $|\psi\rangle$, then the unitary $\widetilde{U} = \mathbb{I} - 2\Pi_D$ has mapping advantage at least ε^2 on $|\psi\rangle$.

Proof of Item (i). Let $|\phi\rangle = \mathsf{ctl}_{\mathsf{B}} U |+\rangle_{\mathsf{B}} |\psi_+\rangle$. The proof follows from directly computing the distinguishing advantage

$$\frac{1}{2}\left|\langle\phi| \Pi\widetilde{\Pi}_D\Pi |\phi\rangle - \langle\phi| \Pi W\widetilde{\Pi}_D W\Pi |\phi\rangle\right|.$$

For $b \in \{0,1\}$, we compute

$$\begin{aligned}
\langle\phi| \Pi W^b\widetilde{\Pi}_D W^b\Pi |\phi\rangle &= \left\|\widetilde{\Pi}_D W^b\Pi \mathsf{ctl}_{\mathsf{B}} U |+\rangle_{\mathsf{B}} |\psi_+\rangle\right\|^2 \\
&= \frac{1}{2}\left\||+\rangle\langle+|_{\mathsf{B}} \left(|0\rangle_{\mathsf{B}} W^b |\psi_+\rangle + |1\rangle_{\mathsf{B}} U^\dagger W^b\Pi U |\psi_+\rangle\right)\right\|^2 \\
&= \frac{1}{4}\left\||+\rangle_{\mathsf{B}} \left(|\psi_+\rangle + U^\dagger W^b\Pi U |\psi_+\rangle\right)\right\|^2 \\
&= \frac{1}{4}\left(\langle\psi_+|\psi_+\rangle + \langle\psi_+| U^\dagger\Pi U |\psi_+\rangle + 2\,\mathrm{Re}\,\langle\psi_+| U^\dagger W^b\Pi U |\psi_+\rangle\right).
\end{aligned}$$

Then taking the difference,

$$\begin{aligned}
&\langle\phi| \Pi\widetilde{\Pi}_D\Pi |\phi\rangle - \langle\phi| \Pi W\widetilde{\Pi}_D W\Pi |\phi\rangle \\
&= \frac{1}{2}\left(\mathrm{Re}\,\langle\psi_+| U^\dagger\Pi U |\psi_+\rangle - \mathrm{Re}\,\langle\psi_+| U^\dagger W\Pi U |\psi_+\rangle\right) \\
&= \frac{1}{2}\,\mathrm{Re}\,\langle\psi_+| U^\dagger(I - W)\Pi U |\psi_+\rangle \\
&= \langle\psi_+| U^\dagger W^-\Pi U |\psi_+\rangle \\
&= \varepsilon.
\end{aligned}$$

□

Proof of Item (ii). We compute

$$\begin{aligned}
\left\|W^-\Pi\widetilde{U} |\psi_+\rangle\right\|^2 &= \left\|\Pi\left(\frac{I - W}{2}\right)\widetilde{U}\left(\frac{I + W}{2}\right)\Pi |\psi\rangle\right\|^2 \\
&\geq \left|\langle\psi|\,\Pi\left(\frac{I - W}{2}\right)\widetilde{U}\left(\frac{I + W}{2}\right)\Pi |\psi\rangle\right|^2 \\
&= \frac{1}{4}\left|(\langle\psi| - \langle\psi| W)\widetilde{U}(|\psi\rangle + W |\psi\rangle)\right|^2 \\
&= \frac{1}{4}\left|\langle\psi|\,\widetilde{U} |\psi\rangle - \langle\psi| W\widetilde{U}W |\psi\rangle + 2i\cdot\mathrm{Im}(\langle\psi|\,\widetilde{U}W |\psi\rangle)\right|^2 \\
&\geq \frac{1}{4}\left|\langle\psi|\,\widetilde{U} |\psi\rangle - \langle\psi| W\widetilde{U}W |\psi\rangle\right|^2 \\
&= \frac{1}{2}\cdot\left|\langle\phi| \Pi\widetilde{\Pi}_D\Pi |\phi\rangle - \langle\phi| \Pi W\widetilde{\Pi}_D W\Pi |\phi\rangle\right|^2 \\
&= 2\varepsilon^2.
\end{aligned}$$

□

Remark 6.7. *In [AAS20], a similar statement is proven about the equivalence between distinguishing and* swapping *for the task of distinguishing between two* fixed *states. They also showed that distinguishing and mapping are not in general equivalent for the fixed-state setting. Lemma 6.6 implies that distinguishing, swapping, and mapping are all equivalent when the task is to* provide *a state and detect the application of a fixed operation.*

We now prove that an admissible unitary G cannot help an adversary with the mapping task. Without loss of generality we assume that the adversary repeatedly applies a single unitary U, interleaved with G.[21] Lemma 6.8 bounds the mapping advantage of $(UG)^t$ in terms of the mapping advantage of *just applying* U^q for any $q \le t$, with an initial state that can be prepared with access to a unitary very similar to G.

Lemma 6.8. *Fix a Hilbert space $\mathcal{H}$. Let U be a unitary, $|\psi\rangle$ be a state, and Π_0, Π_1 be a pair of orthogonal projectors, all on $\mathcal{H}$. Let $\Pi = \Pi_0 + \Pi_1$. Let G be a unitary on $\mathcal{H}$ of the form $G = \Pi_0 G_0 + \Pi_1 G_1 + (\mathbb{I} - \Pi)$ where G_0, G_1 are unitaries that commute with Π_0, Π_1 respectively. Define $\widetilde{G}_0 = \Pi_0 G_0 + (\mathbb{I} - \Pi_0)$ and*

$$\varepsilon(t) \coloneqq \max_{q,r,s \le t} \left\| \Pi_1 U^q \Pi_0 (\widetilde{G}_0)^r (U\widetilde{G}_0)^s \Pi_0 |\psi\rangle \right\|$$

for any integer $t \ge 0$. Then for all integers $t \ge 0$,

$$\left\| \Pi_1 (UG)^t \Pi_0 |\psi\rangle \right\| \le 4t^2 \cdot \varepsilon(t).$$

We divide the proof of Lemma 6.8 into Claim 6.9 and Claim 6.10.

Claim 6.9.

$$\left\| \Pi_1 (U\widetilde{G}_0)^t \Pi_0 |\psi\rangle \right\| \le 2t \cdot \varepsilon(t)$$

Proof. Let $H \coloneqq \Pi_0 G_0 + \Pi_1 G_1 - \mathbb{I}$ and write $\widetilde{G}_0 = \Pi_0 G_0 + \Pi_1 + (\mathbb{I} - \Pi) = \mathbb{I} + \Pi_0 H$. We can then expand the state $\Pi_1 (U\widetilde{G}_0)^t \Pi_0 |\psi\rangle$ as

$$\Pi_1 (U\widetilde{G}_0)^t \Pi_0 |\psi\rangle = \Pi_1 (U + U\Pi_0 H)^t \Pi_0 |\psi\rangle = \sum_{r=1}^{t} \Pi_1 U^r F_{t-r} \Pi_0 |\psi\rangle ,$$

where $F_0 \coloneqq \mathbb{I}$, and $F_i \coloneqq U\Pi_0 H (U + U\Pi_0 H)^{i-1}$ for $i \ge 1$. The second equality above uses the fact that $U^r F_{t-r}$ is the sum of the terms in the binomial expansion of $(U + U\Pi_0 H)^t$ that, when going from left to right, consist of exactly r U's before the first $U\Pi_0 H$.

By the triangle inequality, it suffices to show that for each $r \in [t]$,

$$\| \Pi_1 U^r F_{t-r} \Pi_0 |\psi\rangle \| \le 2\varepsilon(t).$$

The $r = t$ case is immediate from assumption. For the $r \le t - 1$ case, rewrite the definition of F_{t-r} using $\Pi_0 H = \Pi_0 G_0 - \Pi_0$ and $\widetilde{G}_0 = \mathbb{I} + \Pi_0 H$:

$$F_{t-r} = U(\Pi_0 G_0 - \Pi_0)(U\widetilde{G}_0)^{t-r-1}.$$

Plugging in this expression for F_{t-r} and invoking the triangle inequality yields

$$\begin{aligned}
&\| \Pi_1 U^r F_{t-r} \Pi_0 |\psi\rangle \| \\
&= \left\| \Pi_1 U^{r+1} \Pi_0 G_0 (U\widetilde{G}_0)^{t-r-1} \Pi_0 |\psi\rangle - \Pi_1 U^{r+1} \Pi_0 (U\widetilde{G}_0)^{t-r-1} \Pi_0 |\psi\rangle \right\| \\
&\le \left\| \Pi_1 U^{r+1} \Pi_0 G_0 (U\widetilde{G}_0)^{t-r-1} \Pi_0 |\psi\rangle \right\| + \left\| \Pi_1 U^{r+1} \Pi_0 (U\widetilde{G}_0)^{t-r-1} \Pi_0 |\psi\rangle \right\|.
\end{aligned}$$

By our initial assumptions, this implies $\| \Pi_1 U^r F_{t-r} \Pi_0 |\psi\rangle \| \le 2\varepsilon(t)$. □

[21] Any adversary can be converted into this form by introducing an additional clock register.

Claim 6.10.

$$\left\|(UG)^t\Pi_0\left|\psi\right\rangle - (U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle\right\| \leq 2t^2\cdot\varepsilon(t)$$

Proof. We prove this claim by induction on t. The base case $t=0$ is immediate. For $t\geq 1$, we have by the inductive hypothesis that

$$\left\|(UG)^t\Pi_0\left|\psi\right\rangle - UG(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle\right\| \leq 2(t-1)^2\cdot\varepsilon(t-1). \tag{3}$$

Using $G=\mathbb{I}+HP$, we can write $UG(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle$ and $(U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle$ as follows:

$$UG(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle = U(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle + UH\Pi(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle \tag{4}$$

$$(U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle = U(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle + UH\Pi_0(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle. \tag{5}$$

By subtracting Eq. (5) from Eq. (4) and taking the norm, we obtain

$$\begin{aligned}\left\|UG(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle - (U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle\right\| &= \left\|UH\Pi(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle - UH\Pi_0(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle\right\| \\ &\leq \left\|UH\right\|_{\mathrm{op}}\left\|\Pi(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle - \Pi_0(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle\right\|.\end{aligned}$$

By Claim 6.9 (and writing $\Pi_1=\Pi-\Pi_0$)

$$\left\|\Pi(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle - \Pi_0(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle\right\| \leq 2(t-1)\cdot\varepsilon(t-1).$$

Since $\left\|UH\right\|_{\mathrm{op}} = \left\|H\right\|_{\mathrm{op}} \leq 2$, it follows that

$$\left\|UG(U\widetilde{G}_0)^{t-1}\Pi_0\left|\psi\right\rangle - (U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle\right\| \leq 4(t-1)\cdot\varepsilon(t-1). \tag{6}$$

Combining Eqs. (3) and (6), we obtain

$$\begin{aligned}\left\|(UG)^t\Pi_0\left|\psi\right\rangle - (U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle\right\| &\leq 2(t-1)^2\cdot\varepsilon(t-1) + 4(t-1)\cdot\varepsilon(t-1) \\ &\leq 2t^2\cdot\varepsilon(t).\end{aligned}$$

□

Proof of Lemma 6.8. By Claim 6.10 and Claim 6.9, we have

$$\begin{aligned}\left\|\Pi_1(UG)^t\Pi_0\left|\psi\right\rangle\right\| &\leq \left\|\Pi_1(U\widetilde{G}_0)^t\Pi_0\left|\psi\right\rangle\right\| + 2t^2\cdot\varepsilon(t) \\ &\leq 2t\cdot\varepsilon(t) + 2t^2\cdot\varepsilon(t) \\ &\leq 4t^2\cdot\varepsilon(t).\end{aligned}$$

□

Proof of the admissible oracle lemma (Lemma 6.5). Let $\{\Pi_{D^G}, \mathbb{I}-\Pi_{D^G}\}$ be the binary projective measurement corresponding to running distinguisher D^G and measuring the output bit that distinguishes $\Pi\left|\psi\right\rangle$ and $W\Pi\left|\psi\right\rangle$ with ε advantage. By Lemma 6.6, Item (ii), the unitary operator $\widetilde{U} := \mathbb{I}-2\Pi_{D^G}$ achieves a mapping advantage at least ε^2 on initial state $\left|\psi\right\rangle$. The unitary $\widetilde{U}$ can be implemented by running D^G, applying Z to the output qubit, then running $(D^G)^\dagger$. This requires access to two oracles G and $G^\dagger$. For convenience, we will introduce the new admissible oracle

$\widetilde{G} := |0\rangle\langle 0|_{\mathsf{X}} \otimes G + |1\rangle\langle 1|_{\mathsf{X}} \otimes G^\dagger$, which can simulate the action of G and $G^\dagger$ using an ancilla qubit X. For this reason, we write the unitary as $\widetilde{U}^{\widetilde{G}}$.

The circuit $\widetilde{U}^{\widetilde{G}}$, which makes $t+1 = O(T)$ calls to $\widetilde{G}$, can be converted to the form $(U\widetilde{G})^t$ for some fixed unitary U making no calls to $\widetilde{G}$ by appending a clock register Y with $\mathrm{poly}(T)$ overhead in overall circuit size and ancilla registers. Then:

$$\left\|(W^-\Pi)(U\widetilde{G})^t(W^+\Pi)(|0\rangle_{\mathsf{X}}|0\rangle_{\mathsf{Y}}|\psi\rangle)\right\|^2 \geq \varepsilon^2.$$

Then by Lemma 6.8, there exist $q, r, s \leq t$ such that

$$\left\|(W^-\Pi)U^q(W^+\Pi)(\widetilde{G}_0)^r(U\widetilde{G}_0)^s(W^+\Pi)(|0\rangle_{\mathsf{X}}|0\rangle_{\mathsf{Y}}|\psi\rangle)\right\|^2 \geq \varepsilon^2/4t^2$$

where $\widetilde{G}_0 = W^+\Pi\widetilde{G} + (\mathbb{I} - W^+\Pi)$. The unitary $\widetilde{G}_0$ can be implemented using $O(1)$ applications of controlled G, $G^\dagger$ and the unitary $M := (W^+\Pi)\otimes X + (\mathbb{I} - W^+\Pi)\otimes\mathbb{I}$.

Let $\mathsf{Prep}^{G,M}$ be the quantum algorithm that on input $|\psi\rangle$ creates the sub-normalized state[22]

$$|\gamma\rangle := (W^+\Pi)(\widetilde{G}_0)^r(U\widetilde{G}_0)^s(W^+\Pi)(|0\rangle_{\mathsf{X}}|0\rangle_{\mathsf{Y}}|\psi\rangle),$$

and then uses $|\gamma\rangle$ to prepare the state

$$|\phi\rangle = \mathsf{ctl}_{\mathsf{Z}}\text{-}U^q\left(|+\rangle_{\mathsf{Z}}|\gamma\rangle\right).$$

By Lemma 6.6, Item (i), if we define the projector $\widetilde{\Pi}_D := \mathsf{ctl}_{\mathsf{Z}}\text{-}U^q|+\rangle\langle+|_{\mathsf{Z}}\,\mathsf{ctl}_{\mathsf{Z}}\text{-}(U^q)^\dagger$, then the binary projective measurement $\{\widetilde{\Pi}_D, \mathbb{I}-\widetilde{\Pi}_D\}$ distinguishes $\Pi|\phi\rangle$ and $W\Pi|\phi\rangle$ with advantage $\varepsilon^2/8t^2$. Thus, the algorithm E that measures the binary projective measurement $\{\widetilde{\Pi}_D, \mathbb{I}-\widetilde{\Pi}_D\}$ wins the (W,Π)-distinguishing game with advantage $\varepsilon^2/8t^2$. □

6.2 Swap binding implies oracle swap binding

We now prove the main theorem of this section, Theorem 6.2, which says that any swap binding commitment scheme is oracle swap binding.

Proof of Theorem 6.2. Suppose that a size-T adversary A^{G_b} for oracle swap-binding (counting queries to G_b as unit cost) achieves distinguishing advantage ε with initial state $|\psi\rangle_{\mathsf{CDR}}$, where R is the adversary's auxiliary registers. We will use Lemma 6.5 to translate this adversary into one that achieves distinguishing advantage $\varepsilon/16T^2$ in the standard swap-binding game.

Let $\mathsf{ctl}_{\mathsf{B}}\text{-}S := |0\rangle\langle 0|_{\mathsf{B}}\otimes\mathbb{I} + |1\rangle\langle 1|_{\mathsf{B}}\otimes\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']}$, where B is a single qubit register and M' is a register of the same dimension as M. Then define the oracle

$$G = (\mathsf{ctl}_{\mathsf{B}}\text{-}S)\widehat{\mathcal{O}}(\mathsf{ctl}_{\mathsf{B}}\text{-}S)\Pi + (\mathbb{I}-\Pi)$$

and operation

$$W = (\mathsf{ctl}_{\mathsf{B}}\text{-}S)X_{\mathsf{B}}(\mathsf{ctl}_{\mathsf{B}}\text{-}S) = X_{\mathsf{B}}\otimes\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']}. \tag{7}$$

[22] By creating a sub-normalized state, we mean that if a measurement $\{W^+\Pi, \mathbb{I}-W^+\Pi\}$ rejects (i.e., apply M on the state and an ancilla qubit and measure the ancilla qubit), abort and do not produce any output. Otherwise, output the resulting state $|\gamma\rangle$.

Observe that A^G is the same quantum algorithm as A^{G_b}, except the calls to G_b are replaced by calls to G. Since A^{G_b} achieves distinguishing advantage ε in oracle swap-binding with initial state $|\psi\rangle_{\mathsf{CDR}}$, A^G achieves distinguishing advantage ε in the (W,Π)-distinguishing game with registers $\mathsf{A} := (\mathsf{B},\mathsf{M}',\mathsf{C})$, $\mathsf{B} := \mathsf{D}$ and initial state $|0\rangle_{\mathsf{B}} \otimes |0\rangle_{\mathsf{M}'} \otimes |\psi\rangle_{\mathsf{CDR}}$.

By Lemma 6.5, there exists another quantum state $\boldsymbol{\rho}$ and efficient algorithm E *without query access to* G such that the following hold:

- there is a size-poly(T) oracle circuit $\mathsf{Prep}^{G,M}$ such that $\boldsymbol{\rho} \leftarrow \mathsf{Prep}^{G,M}(|0\rangle_{\mathsf{BM}'} |\psi\rangle_{\mathsf{CDR}} |0\rangle_{\mathsf{anc}})$ and Prep uses at most T (controlled) queries to G, $G^\dagger$, and $M := (W^+\Pi) \otimes X + (\mathbb{I} - W^+\Pi) \otimes \mathbb{I}$,
- E has size poly(T), and
- E wins the (W,Π)-distinguishing game with advantage $\varepsilon^2/8T^2$ using initial state $\boldsymbol{\rho}$.

If $\mathcal{O}$ is efficiently implementable, then (controlled) G and $G^\dagger$ are also efficiently implementable. In addition, with our given definition of W, the binary projective measurement $\{W^+\Pi, \mathbb{I} - W^+\Pi\}$ is efficiently implementable. This is because

$$W^+\Pi = \left(|+\rangle\langle+|_{\mathsf{B}} \otimes \widehat{\Pi_{\mathsf{Sym}}} + |-\rangle\langle-|_{\mathsf{B}} \otimes \widehat{\Pi_{\mathsf{Antisym}}}\right) \Pi,$$

where Π_{Sym} and Π_{Antisym} are the projectors onto the the symmetric and antisymmetric subspaces of $\mathcal{M} \otimes \mathcal{M}'$, which can be measured using a SWAP test. Thus M is efficiently implementable. For this reason, we now write $\mathsf{Prep} := \mathsf{Prep}^{G,M}$ because Prep can implement the functionality of controlled G, $G^\dagger$, and M without any oracle access.

Thus E can win the (W,Π)-distinguishing game using initial state $\boldsymbol{\rho} \leftarrow \mathsf{Prep}(|0\rangle_{\mathsf{BM}'} |\psi\rangle_{\mathsf{CDR}} |0\rangle_{\mathsf{anc}})$ with advantage $\varepsilon^2/8T^2$. The conclusion follows by an application of Claim 6.11 (which we state and prove below), which gives a distinguisher for swap-binding with advantage $\varepsilon^2/16T^2$. □

Claim 6.11. *Suppose that an efficient distinguisher D wins the (W,Π)-distinguishing game with advantage δ, where W and Π are defined as*

$$W = X_{\mathsf{B}} \otimes \widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']} \text{ and } \Pi = \mathsf{Com}(\mathbb{I}_{\mathsf{M}} \otimes |0\rangle\langle 0|_{\mathsf{W}})\mathsf{Com}^\dagger.$$

Then there is an efficient distinguisher for the swap-binding game (Definition 4.2) with advantage $\delta/2$.

Proof. Suppose that D wins the (W,Π)-distinguishing game with advantage δ using the state $\boldsymbol{\rho}_{\mathsf{BM}'\mathsf{CDR}}$. This means that D can distinguish the states $\mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}(\Pi\boldsymbol{\rho}\Pi)$ and $\mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}(W\Pi\boldsymbol{\rho}\Pi W)$ with advantage δ. Because the B register is traced out in the (W,Π)-distinguishing game, the operation X_{B} in W has no effect on the distinguisher's view, so we may rewrite the latter state as

$$\mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}\left(W\Pi\boldsymbol{\rho}\Pi W\right) = \mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}\left(\left(\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']}\right) \cdot \Pi\boldsymbol{\rho}\Pi \cdot \left(\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']}\right)\right). \tag{8}$$

We use a hybrid argument to show that the distinguisher D yields an adversary for swap-binding with advantage $\delta/2$. We define hybrid games H_0, H_1, and H_2, where the adversary's view in hybrid H_0 corresponds to $\mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}(\Pi\boldsymbol{\rho}\Pi)$ and the adversary's view in hybrid H_2 corresponds to $\mathrm{Tr}_{\mathsf{BM}'\mathsf{C}}(W\Pi\boldsymbol{\rho}\Pi W)$.

- H_0:
  1. The adversary sends a quantum state $\rho_{\mathsf{BM'CDR}}$ to the challenger.
  2. The challenger then does the following:
     (a) Measure $\{\Pi, \mathbb{I} - \Pi\}$ and abort if the measurement rejects.
     (b) Send the (D, R) register to the adversary.
- H_1:
  1. The adversary sends a quantum state $\rho_{\mathsf{BM'CDR}}$ to the challenger.
  2. The challenger then does the following:
     (a) Measure $\{\Pi, \mathbb{I} - \Pi\}$ and abort if the measurement rejects.
     (b) Initialize another register E (of the same dimension as M) to the $|0\rangle$ state.
     (c) Apply $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{E}]$.
     (d) Send the (D, R) register to the adversary.

  This is immediately indistinguishable from H_0 by swap-binding.
- H_2:
  1. The adversary sends a quantum state $\rho_{\mathsf{BM'CDR}}$ to the challenger.
  2. The challenger then does the following:
     (a) Measure $\{\Pi, \mathbb{I} - \Pi\}$ and abort if the measurement rejects.
     (b) Apply $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{M}']$.
     (c) Send the (D, R) register to the adversary.

  We show that H_2 is indistinguishable from H_1 by invoking swap-binding. Given an adversary that distinguishes H_1 from H_2 with advantage δ', we construct the following reduction that distinguishes the $b = 0$ and $b = 1$ worlds of the swap binding game with advantage δ':
  1. Receive $\rho_{\mathsf{BM'CDR}}$ from the adversary.
  2. Measure $\{\Pi, \mathbb{I} - \Pi\}$ and abort if the measurement rejects.
  3. Apply $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{M}']$.
  4. Send registers (C, D) to the swap-binding challenger.
  5. Receive D from the swap-binding challenger.
  6. Forward (D, R) to the adversary.
  7. Use the adversary's output as the guess for the swap-binding game.

  Because the reduction performs the measurement $\{\Pi, \mathbb{I}-\Pi\}$, the challenger's measurement has no effect. In the $b = 0$ world of the swap-binding game, the challenger will do nothing, which corresponds exactly to H_2. In the $b = 1$ world, the challenger will apply $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{E}]$ where E is initialized to $|0\rangle$. Thus, the adversary receives the decommitment after the operation $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{E}] \cdot \widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{M}'] = \mathsf{SWAP}[\mathsf{M}', \mathsf{E}] \cdot \widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{E}]$ has been applied. Because the final $\mathsf{SWAP}[\mathsf{M}', \mathsf{E}]$ has no effect on the adversary's view, this corresponds exactly to H_1.

□

7 A succinct quantum argument protocol

In this section, we construct an interactive succinct quantum argument for any language with quantum PCPs with constant soundness error and polylogarithmic query complexity.

7.1 Preliminaries

7.1.1 Probabilistically checkable proofs

Definition 7.1 (Quantum Probabilistically Checkable Proofs). *A quantum PCP for a language L is parameterized by a completeness parameter c, soundness s, proof length m, randomness complexity ℓ, and query complexity q. We require the following properties:*

- *(Efficient verification) There is a classical* $\mathsf{poly}(n)$ *time procedure that takes as input $x \in \{0,1\}^n$ and $r \in \{0,1\}^\ell$ and outputs the description of circuit for implementing a q-qubit projective measurement $\{\Pi^{\mathsf{PCP}}_{x,r}, \mathbb{I} - \Pi^{\mathsf{PCP}}_{x,r}\}$, which acts on a state of size m.*
- *(Completeness) If $x \in L$, there exists an m-qubit state $\boldsymbol{\pi}$ such that*

$$\mathop{\mathbb{E}}_{r \leftarrow \{0,1\}^\ell} \mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_{x,r} \boldsymbol{\pi}\right) \geq c.$$

- *(Soundness) If $x \notin L$, then for any m-qubit state $\boldsymbol{\pi}$,*

$$\mathop{\mathbb{E}}_{r \leftarrow \{0,1\}^\ell} \mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_{x,r} \boldsymbol{\pi}\right) \leq s.$$

Which languages have good quantum PCPs? It is easy to check that any *classical* PCP is captured by this definition. Define the measurements $\{\Pi^{\mathsf{PCP}}_{x,r}, \mathbb{I} - \Pi^{\mathsf{PCP}}_{x,r}\}$ corresponding to the predicate that the classical PCP verifier checks. Completeness is trivially preserved since any classical proof string π is also a quantum state. Soundness is preserved because the measurements $\{\Pi^{\mathsf{PCP}}_{x,r}, \mathbb{I} - \Pi^{\mathsf{PCP}}_{x,r}\}$ correspond to classical predicates and thus commute with standard basis measurements. If $x \notin L$, the probability that a quantum state $|\psi\rangle = \sum_{\pi \in \{0,1\}^m} \alpha_\pi |\pi\rangle$ is accepted is

$$\sum_{\pi \in \{0,1\}^m} |\alpha_\pi|^2 \Pr[\pi \text{ is accepted}] \leq \sum_{\pi \in \{0,1\}^m} |\alpha_\pi|^2 s = s.$$

The statement for mixed states follows by convexity.

Theorem 7.2 (PCP theorem [BFLS91, FGL+91, AS98, ALM+98]). *There exist constants $0 \leq s < c \leq 1$ and $k = O(1)$ such that every* NP *language has a PCP with completeness c, soundness s, proof length $m = \mathsf{poly}(n)$, randomness complexity $\ell = O(\log n)$, and query complexity $q = O(\log n)$.*

Conjecture 1 (Quantum PCP conjecture [AALV09, AAV13]). *There exist constants $0 \leq s < c \leq 1$ and $k = O(1)$ such that every* QMA *language has a quantum PCP with completeness c, soundness s, proof length $m = \mathsf{poly}(n)$, randomness complexity $\ell = O(\log n)$, and query complexity $q = O(\log n)$.*

Any quantum PCP with constant completeness-soundness gap can be repeated in parallel $\log^2(\lambda)$ times to achieve $1 - \mathsf{negl}(\lambda)$ completeness soundness error. The proof of soundness amplification is the same as the standard proof that QMA can be amplified to exponentially small error by parallel repetition [KSV02].

Claim 7.3 (Soundness amplification). *Suppose that a language L has a quantum PCP with constant completeness-soundness gap $c - s$, proof length m, randomness complexity ℓ, and query complexity q. Then L has a quantum PCP with completeness $1 - \mathsf{negl}(\lambda)$, soundness $\mathsf{negl}(\lambda)$, proof length $m \cdot \mathsf{polylog}(\lambda)$, randomness complexity $\ell \cdot \mathsf{polylog}(\lambda)$, and query complexity $q \cdot \mathsf{polylog}(\lambda)$.*

7.1.2 Interactive argument preliminaries

Quantum interactive protocol syntax. We first establish a syntax for general quantum interactive protocols between a prover and a verifier. We state the definition below for any $(2r-1)$-round protocol, though in this work we will focus on 3-round protocols.

Definition 7.4. *A $(2r-1)$-message quantum interactive protocol $\langle P, V\rangle$ between a prover P and verifier V is specified by two sequences of unitaries $U_{P,1}, \ldots, U_{P,r}$ and $U_{V,1}, \ldots, U_{V,r}$. The prover and verifier have internal registers P and V respectively and also share a message register Z. The interaction proceeds as follows:*

For $i = 1, \ldots, r$, the prover and verifier do the following:

- *The prover applies $U_{P,i}$ to (P, Z) and sends Z to the verifier.*
- *The verifier applies $U_{V,i}$ to (V, Z). If $i < r$, it sends Z back to the prover. If $i = r$, it measures the first qubit of V and accepts if it is 1 and rejects if it is 0.*

We write $\langle P(\boldsymbol{\rho}_{\mathsf{P}}), V(\boldsymbol{\tau}_{\mathsf{V}})\rangle_{\mathrm{OUT}}$ to denote the $0/1$-random variable corresponding to the verifier's decision when the protocol is run with the P register initialized to $\boldsymbol{\rho}_{\mathsf{P}}$ and the V register initialized to $\boldsymbol{\tau}_{\mathsf{V}}$.

The following definition of argument of knowledge is based on [CMSZ21, Definition 3.6] with some modifications to handle extraction of quantum proofs.

Definition 7.5. *Let Ver be a quantum algorithm that takes as input a state $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H})$ and outputs an accept/reject decision. Define $\mathsf{Wit}[\mathsf{Ver}, p] \subseteq \mathbf{D}(\mathcal{H})$ to be the set of all states $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H})$ such that $\Pr[\mathsf{Ver}$ accepts $\boldsymbol{\rho}] \geq p$.*

Definition 7.6 (Argument of knowledge). *Let L be a QMA language and let $\mathsf{Ver}_{L,x}$ be the corresponding QMA verification procedure for statement x. An interactive argument $\langle P, V\rangle$ is an argument of knowledge for L with respect to Ver_L if there exists a quantum algorithm $\mathsf{KnowledgeExt}$ such that for any x, any polynomial-time quantum interactive prover $\tilde{P}$ that convinces the argument verifier with probability $p' := \Pr\left[\langle \tilde{P}(x), V(x)\rangle_{\mathrm{OUT}} = 1\right]$, any $\varepsilon \geq 1/\mathsf{poly}(\lambda)$,*

and any $p \leq p'$,[23]

$$\Pr\left[\boldsymbol{\rho} \in \mathsf{Wit}[\mathsf{Ver}_{L,x}, p-\varepsilon] : \boldsymbol{\rho} \leftarrow \mathsf{KnowledgeExt}^{\tilde{P}}(x, 1^{\lceil 1/p \rceil})\right] \geq \Omega(\varepsilon).$$

The runtime of $\mathsf{KnowledgeExt}^{\tilde{P}}(x, 1^{\lceil 1/p \rceil})$ *is* $\mathsf{poly}(\lambda, 1/\varepsilon)$*, counting calls to* $\tilde{P}$ *as unit time.*

At an intuitive level, this definition states that a protocol is an argument of knowledge with respect to a proof verifier Ver (which will typically correspond to a PCP verifier) if, given access to any efficient malicious prover that convinces the argument verifier with probability p, the extractor can (with noticeable probability) produce a proof string that convinces Ver with probability $\approx p$.

7.2 Quantum tree commitments

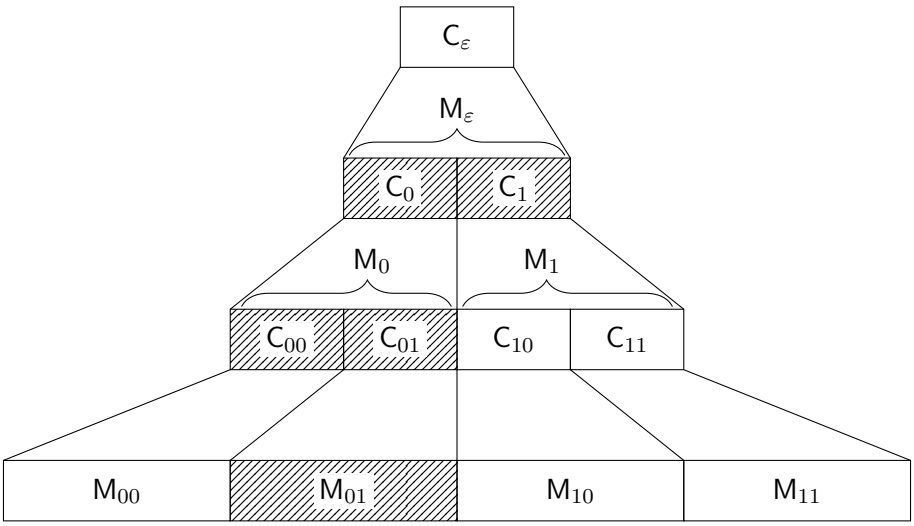

Figure 1: The Merkle tree structure. For simplicity, we assume that $\mathsf{D}_\ell = \mathsf{M}_\ell$ in this picture (as is the case in, for example, the scheme in Theorem 5.3). The shaded boxes correspond to decommitments $\{\mathsf{D}_\varepsilon, \mathsf{D}_0, \mathsf{D}_{01}\}$ for the 01 block.

We now describe the syntax for quantum tree commitments. This syntax is a generalization of a construction that appeared in prior work of Chen and Movassagh [CM21]; in our notation, their construction corresponds to a setting where the ancilla register W and the commitment C must be the same size.

Let sQSC be a succinct QSC with $s(\lambda)$-sized messages and $s(\lambda)/2$-sized commitments and let Com be the unitary implementing the commitment. For any positive integer β, a quantum tree commitment for $(s \cdot 2^\beta)$-size messages M supports the following functionality:

- The tree commitment algorithm $\mathsf{QTreeCom}_{\mathsf{sQSC}}$ does the following:

[23] Following [CMSZ21], we give the extractor a lower bound p on the success probability q as classical advice. It is plausible that this requirement could be removed using an extractor based on techniques of [LMS22b], but we did not attempt to work out the details.

1. Divide M into 2^β registers $\mathsf{M} = \{\mathsf{M}_\ell\}_{\ell\in\{0,1\}^\beta}$ where each M_ℓ is a block of s qubits.
2. For $j = \beta, \beta-1, \ldots, 1$:
   (a) For each $\ell \in \{0,1\}^j$, initialize W_ℓ to $|0\rangle$ and apply Com to (M, W) to obtain $(\mathsf{C}_\ell, \mathsf{D}_\ell)$.
   (b) For each $\ell \in \{0,1\}^{j-1}$, define $\mathsf{M}_\ell := (\mathsf{C}_{\ell,0}, \mathsf{C}_{\ell,1})$ (when $j = 1$, we define $\mathsf{M}_\varepsilon := (\mathsf{C}_0, \mathsf{C}_1)$ where ε is the empty string).
3. Initialize W_ε to $|0\rangle$ and apply Com to $(\mathsf{M}_\varepsilon, \mathsf{W})$ to obtain $(\mathsf{C}_\varepsilon, \mathsf{D}_\varepsilon)$. The commitment is C_ε and $\{\mathsf{D}_\ell\}_{\ell\in\{0,1\}^{\le\beta}}$ is a set of decommitments which will be used later to form the local decommitments.

- For any subset $S \subseteq \{0,1\}^{\le\beta}$, we define the local decommitment decom_S as follows:
  - For any $\ell \in \{0,1\}^{\le\beta}$, let $\mathsf{Path}(\ell)$ be the set of all $|\ell|+1$ prefixes of ℓ (including ℓ itself and the empty string ε).
  - Let $\mathsf{Path}(S) := \bigcup_{\ell\in S} \mathsf{Path}(\ell)$.
  - Finally, define $\mathsf{decom}_S := \{\mathsf{D}_\ell\}_{\ell\in\mathsf{Path}(S)}$.
- The receiver can verify decom_S as follows:
  1. Parse decom_S as $\{\mathsf{D}_\ell\}_{\ell\in\mathsf{Path}(S)}$.
  2. For $j = 0, \ldots, \beta$:
     (a) For each $\ell \in \mathsf{Path}(S) \cap \{0,1\}^j$:
        i. Apply $\mathsf{Com}^\dagger$ to $(\mathsf{C}_\ell, \mathsf{D}_\ell)$ to obtain $(\mathsf{M}_\ell, \mathsf{W}_\ell)$.
        ii. Measure W_ℓ with $(|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|)$ and abort if the outcome is not $|0\rangle$.
        iii. If $j < \beta$, partition M_ℓ into two $(s/2)$-qubit registers $(\mathsf{C}_{\ell,0}, \mathsf{C}_{\ell,1})$.
  3. Output $\bigotimes_{\ell\in S} \mathsf{M}_\ell$.

7.3 The protocol description

In this section we construct a succinct quantum argument from any quantum PCP and succinct QSC. The protocol is parameterized by a choice of the underlying PCP as well as the succinct QSC sQSC, where:

- PCP is a classical or quantum PCP system for a language L.
- sQSC is a succinct QSC with $s(\lambda)$-sized messages and $s(\lambda)/2$-sized commitments.

For any setting of PCP and sQSC satisfying these requirements, we denote the resulting succinct quantum argument as SQUARG[PCP, sQSC] (for Succinct QUantum ARGument). The protocol is specified as follows.

SQUARG[PCP, sQSC]:

Prover input: x and a corresponding PCP $|\pi\rangle$. We assume the PCP has length $m = s \cdot 2^\beta$ qubits for some integer β, which is without loss of generality (by padding with 0's).

Verifier input: x.

1. The prover applies $\mathsf{QTreeCom}_{\mathsf{sQSC}}$ to $|\pi\rangle$ and obtains $\mathsf{C}_\varepsilon, \{\mathsf{D}_\ell\}_{\ell\in\{0,1\}^{\leq\beta}}$. It sends the root commitment C_ε to the verifier.
2. The verifier sends random coins $r \leftarrow \{0,1\}^{\mathsf{rc}}$ to the prover.
3. The prover sends decom_{S_r} where $S_r \subseteq \{0,1\}^\beta$ is the set of block labels corresponding to the positions that $\{\Pi_r, \mathbb{I} - \Pi_r\}$ checks (i.e., those that contain Q_r).
4. Finally, the verifier checks decom_{S_r}. If the decommitment is valid, it also obtains the qubits Q_r of the PCP corresponding to r and measures $\{\Pi_r, \mathbb{I} - \Pi_r\}$. It accepts if the measurement accepts.

The following is immediate by construction.

Claim 7.7 (Succinctness)**.** *Suppose that for an instance of size n, we have a* PCP *with proofs of length $m = \mathsf{poly}(n,\lambda)$, query complexity $\mathsf{poly}(\lambda)$, randomness complexity $\mathsf{poly}(\lambda)$, and completeness-soundness gap $1 - \mathsf{negl}(\lambda)$. Moreover, assume that the PCP verifier runs in time $t(n,\lambda)$. Then the protocol* $\mathsf{SQUARG}[\mathsf{PCP}, \mathsf{sQSC}]$ *satisfies the following properties:*

- *The quantum prover runs in time* $\mathsf{poly}(m,\lambda)$.
- *The quantum verifier runs in time* $t(n,\lambda) + \mathsf{poly}(\lambda, \log(m))$.
- *The total quantum communication is* $\mathsf{poly}(\lambda, \log(m))$ *qubits.*

8 Security of the quantum succinct argument protocol

In this section we prove security of the quantum succinct argument protocol described in Section 7.3.

Theorem 8.1. *Let* PCP *be a probabilistically checkable proof for a language L and let* sQSC *be a succinct QSC with s-qubit messages and $s/2$-qubit commitments. If* sQSC *is swap binding, then* SQUARG[PCP, sQSC] *is an argument of knowledge for L with respect to the* PCP *verifier (Definition 7.6).*

By Claim 7.7, we obtain the following corollaries of Theorem 8.1.

Corollary 8.2. *Assuming the existence of succinct QSCs, there is a three-message quantum succinct argument of knowledge for* NP.[24]

Corollary 8.3. *If the quantum PCP conjecture holds [AALV09, AAV13], then assuming the existence of succinct QSCs, there is a three-message quantum succinct argument (of knowledge) for* QMA.

Remark 8.4. *Technically, our definition of argument of knowledge (Definition 7.6) states that if the argument prover convinces the argument verifier with probability p, then the knowledge extractor outputs a PCP (with noticeable probability) that would convince the PCP verifier with probability $\approx p$. In the setting of arguments/proofs of knowledge for* NP*, our notion is slightly different from the traditional notion, which requires that the extractor output an* NP *witness. (PCPs are not* NP *witnesses since they are not guaranteed to be deterministically verifiable.)*

However, this issue can easily be resolved assuming the PCP is "witness-extractable" (see [Val08] for a discussion). Moreover, this distinction does not arise in the QMA *setting, since the definition of* QMA *only requires probabilistic verification, so quantum PCPs are already valid* QMA *witnesses.*

The rest of this section is organized as follows:

- In Section 8.1, we recall the "one-bit" extraction procedure from [CMSZ21].
- In Section 8.2, we describe a knowledge extraction procedure that outputs a quantum PCP given black-box access to any malicious prover for the quantum succinct argument.
- In Section 8.3, we prove that our knowledge extractor satisfies the conditions of Definition 7.6, establishing Theorem 8.1.

8.1 CMSZ preliminaries

Recap: estimating success probability. A key component of the [CMSZ21] extraction procedure is a subroutine called Est defined with respect to a set of projectors $\{\Pi_r\}_r$. At a high level, Est is a quantum measurement on a state $\boldsymbol{\rho}$ that estimates the probability that, for a random

[24] This follows by combining Theorem 1.4 with the classical PCP theorem [BFLS91, FGL+91, AS98, ALM+98], since quantum PCPs encompass classical PCPs.

r, the measurement $\{\Pi_r, \mathbb{I} - \Pi_r\}$ accepts. In [CMSZ21], this Est procedure is implemented using the alternating projectors trick of [MW05] and incorporates additional modifications proposed by [Zha20].

To make this concrete, fix a Hilbert space $\mathcal{H}$, a corresponding register H, and a state $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H})$. Let $\Pi := \{\Pi_r\}_{r \in R}$ denote a set of projections acting on H. In our setting, each Π_r will correspond to the projection that checks whether the prover outputs an accepting response on the challenge $r \in R$. Thus, the "success probability" of $\boldsymbol{\rho}$ is defined to be $\mathbb{E}_{r \leftarrow R}[\mathrm{Tr}(\Pi_r \boldsymbol{\rho})]$.

The Est algorithm is also parameterized by an additive error bound $0 < \varepsilon < 1$ and a failure probability $0 < \delta < 1$.[25] We give a formal description below, following [CMSZ21]:

$\mathsf{Est}^{\Pi}_{\varepsilon,\delta}$:

Input: A state $\boldsymbol{\rho}$ on register H.

Setup: Define $t = \lceil 2\log(2/\delta)/\varepsilon^2 \rceil$.

1. Initialize a register Q to the state $|+_R\rangle := \frac{1}{2}|\top\rangle + \frac{1}{2}|\bot\rangle + \frac{1}{\sqrt{2|R|}}\sum_{r \in R}|r\rangle$.
2. Define $M_{\mathrm{unif}} := \{|+_R\rangle\langle+_R| \otimes \mathbb{I}_{\mathsf{H}}, \mathbb{I} - |+_R\rangle\langle+_R| \otimes \mathbb{I}_{\mathsf{H}}\}$ and $M_{\mathrm{win}} := \{\Pi_{\mathrm{win}}, \mathbb{I} - \Pi_{\mathrm{win}}\}$ where
$$\Pi_{\mathrm{win}} = \sum_{r \in R} |r\rangle\langle r|_{\mathsf{Q}} \otimes \Pi_r + |\top\rangle\langle\top|_{\mathsf{Q}} \otimes \mathbb{I}.$$
For $i = 1, \ldots, t$:
   (a) Measure M_{unif}, obtaining outcome $b_{2i-1} \in \{0, 1\}$.
   (b) Measure M_{win}, obtaining outcome $b_{2i} \in \{0, 1\}$.
3. If $b_{2t} = 1$, skip to Step 4. Otherwise, apply $M_{\mathrm{win}}, M_{\mathrm{unif}}$ in an alternating fashion until $M_{\mathrm{unif}} \to 1$, or a further $2t$ measurements have been applied.
4. Discard Q and output
$$\tilde{p} := 2\left(\frac{\sum_{i \in [t]}(1 - (b_i - b_{i-1})^2)}{t}\right) - \frac{1}{2}.$$

Remark 8.5. *Following [CMSZ21], the additional symbols $\bot$ and $\top$ are used to ensure that* Est *runs in strict (rather than expected) polynomial time. In particular, the outcomes obtained in Step 2 correspond to the success probability of the prover in the following game:*

- *with $1/2$ probability, we sample a random $r \leftarrow R$ and test whether the $\{\Pi_r, \mathbb{I} - \Pi_r\}$ accepts and*
- *with $1/2$ probability, we sample a random "challenge" from $\{\top, \bot\}$, and output success if the challenge is $\top$ and failure if the challenge is $\bot$.*

If the prover has "real" success probability $p = \mathbb{E}_{r \leftarrow R}[\mathrm{Tr}(\Pi_r \boldsymbol{\rho})]$, its success probability in this modified game is $1/4 + p/2 \in [1/4, 3/4]$. This rescaling enables running in strict polynomial time and is corrected in the last step of Est *(Step 4).*

[25] In [Zha20, CMSZ21], this is captured by the notion of (ε, δ)-almost-projectivity defined in [Zha20]: for any initial state $\boldsymbol{\rho}$, if $\mathsf{Est}_{\Pi,\varepsilon,\delta}$ is applied *twice in a row*, the probability that the resulting outcomes p, p' are more than ε apart is at most δ.

Recap: repairing the state after measurement. [CMSZ21] use Est as a subroutine for their "state repair" procedure. To describe state repair, we introduce some additional definitions:

- Without loss of generality, the procedure $\mathsf{Est}^{\Pi}_{\varepsilon,\delta}$ can be efficiently implemented by (1) initializing an ancilla register V to $|0\rangle$, (2) applying some unitary to (H,V), (3) performing some projective measurement $\{\Pi'_p\}_p$ on V to obtain the estimate $\widetilde{p}$, and (4) tracing out V. We define $\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta}$ to be the unitary applied in step (2).[26]
- For any $p \in [0,1]$, we define a projection Π_p on (H,V) as

$$\Pi_p := \sum_{p' \geq p} (\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta})^\dagger \Pi'_{p'} (\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta}).$$

  In other words, Π_p is the projection that corresponds to running $\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta}$ and obtaining an estimate $p' \geq p$.

With these definitions in hand, we now describe the [CMSZ21] repair procedure. State repair is applied to a (normalized) state $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{H})$ that was just "disturbed" by some projective measurement $\{D, \mathbb{I} - D\}$; we will assume the outcome of the measurement was D, i.e., $\boldsymbol{\rho}$ satisfies $\mathrm{Tr}(D\boldsymbol{\rho}) = 1$. Given a target success probability p, the procedure will output a "repaired" state whose success probability is p, assuming that certain conditions on $\boldsymbol{\rho}$ hold.

$\underline{\mathsf{Repair}^{\Pi,D}_{\varepsilon,\delta,p}}$:

Input: A state $\boldsymbol{\rho}$ on register H such that $\mathrm{Tr}(D\boldsymbol{\rho}) = 1$.

1. Initialize the ancilla register V to $|0\rangle$.
2. Perform the measurement $A := \{\Pi_p, \mathbb{I} - \Pi_p\}$. If the measurement accepts, skip to Step 4.
3. Define $B := \{D_{\mathsf{H}} \otimes |0\rangle\langle 0|_{\mathsf{V}}, \mathbb{I}_{\mathsf{H},\mathsf{V}} - D_{\mathsf{H}} \otimes |0\rangle\langle 0|_{\mathsf{V}}\}$. Perform the measurements $B, A, B, A, \ldots$ in alternating fashion until either an A measurement accepts, or $\lceil 1/\sqrt{\delta} \rceil$ total applications of (B, A) have occurred.
4. Apply $\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta}$ and discard the V registers.

We refer the reader to [CMSZ21, Section 4.3] for additional details and intuition behind the procedure.

A one-bit extractor. Using Est and Repair, we define a procedure BitExtract that takes a single copy of a state $\boldsymbol{\rho}$ and tries to obtain many accepting outcomes for many $\{\Pi_r, \mathbb{I} - \Pi_r\}$ measurements. At a very high level, BitExtract works by (1) running Est to estimate the initial success probability of $\boldsymbol{\rho}$ and (2) repeatedly applying $\{\Pi_r, \mathbb{I} - \Pi_r\}$ for random $r \leftarrow R$ followed by an application of Repair to restore the success probability.

Formally, the BitExtract procedure is parameterized by an error tolerance $\gamma > 0$, a lower bound q on the success probability, and a runtime parameter T.

[26] In [CMSZ21, Section 4.3], $\mathsf{CoherentEst}^{\Pi}_{\varepsilon,\delta}$ corresponds to the unitary U_{M}.

$\underline{\mathsf{BitExtract}^{\Pi}_{\gamma,q,T}}$:

Input: A state $\boldsymbol{\rho}$ on register H.

Setup: Define $\varepsilon = \gamma/4T$ and $\delta = (\gamma/16T)^2$.

1. For $i = 0, \ldots, T-1$:
   (a) Measure $\mathsf{Est}^{\Pi}_{\varepsilon,\delta} \to p_i$. If $p_i < q$, abort and output failure.
   (b) Sample a random $r \leftarrow R$ and apply the measurement $\{\Pi_r, \mathbb{I} - \Pi_r\}$.
   (c) If the measurement accepts, set $D = \Pi_r$. If the measurement rejects, set $D = \mathbb{I} - \Pi_r$.
   (d) Apply $\mathsf{Repair}^{\Pi,D}_{\varepsilon,\delta,p_i}$.
2. Measure $\mathsf{Est}^{\Pi}_{\varepsilon,\delta} \to p_T$.
3. Output success if $p_T \geq q$; otherwise output failure.

We require the following guarantee on BitExtract which we prove using [CMSZ21, Claim 4.14].

Claim 8.6. *Let* $p = \mathbb{E}_{r \leftarrow R}[\mathrm{Tr}(\Pi_r \boldsymbol{\rho})]$. *Then*

$$\Pr\left[\mathsf{BitExtract}^{\Pi}_{\gamma,q,T}(\boldsymbol{\rho}) \to \mathsf{success}\right] \geq p - q - \gamma.$$

Proof. If $p_i \geq q + \gamma/2 - i\gamma/2T$ for all $i \in \{0, \ldots, T\}$, then the extractor will succeed. We therefore take a union bound over the probability that $p_0 < q + \gamma/2$ and the probability that $p_i < p_{i-1} - \gamma/2T$ for some $i \in [T]$:

- By Markov's inequality, the probability that $p_0 < q + \gamma/2$ is at most $1 - (p - q - \gamma/2)$.
- By [CMSZ21, Claim 4.14], the probability that $p_i < p_{i-1} - \gamma/2T$ for some $i \in [T]$ is at most $8T\sqrt{\delta} \leq \gamma/2$.

The claim follows by a union bound over these probabilities. □

8.2 Description of the knowledge extractor

In this subsection, we describe our knowledge extractor KnowledgeExt. We first establish some useful notation:

- Recall the following notation for the tree commitments to PCP proof strings introduced in Sections 7.2 and 7.3. The number of blocks (leaves) of the tree commitment is 2^β. Each node of the tree is indexed by a string $\ell \in \{0,1\}^{\leq \beta}$, where node ℓ is distance $|\ell|$ away from the root. For a challenge r, the set $S_r \subseteq \{0,1\}^\beta$ is the set of (block) indices that contain the positions the PCP verifier checks on randomness r. For any tree node $\ell \in \{0,1\}^{\leq \beta}$, $\mathsf{Path}(\ell)$ is the set of all $|\ell|+1$ prefixes of ℓ (including ℓ itself and the empty string ε). For $S \subseteq \{0,1\}^{\leq \beta}$, $\mathsf{Path}(S) := \bigcup_{\ell \in S} \mathsf{Path}(\ell)$. The tree commitment is the root C_ε. For any $\ell \in \{0,1\}^{\leq \beta}$, the commitment-decommitment pair $(\mathsf{C}_\ell, \mathsf{D}_\ell)$ can be opened by applying $\mathsf{Com}^\dagger$ to obtain the corresponding message-ancilla registers $(\mathsf{M}_\ell, \mathsf{W}_\ell)$. If $|\ell| < \beta$, M_ℓ can be split into two $s/2$ qubit registers $(\mathsf{C}_{\ell,0}, \mathsf{C}_{\ell,1})$ (see Fig. 1).

- Let H denote the register containing the malicious prover's state after it sends the tree commitment C_ε.
- Let $\mathsf{M}' := (\mathsf{M}'_\ell)_{\ell\in\{0,1\}^{\leq\beta}}$ denote a set of s-qubit registers. Looking ahead, these will eventually contain the extracted messages from the tree commitment and the extracted PCP will be the state on the registers $(\mathsf{M}'_\ell)_{\ell\in\{0,1\}^{\beta}}$ corresponding to the leaves. Similarly to M, if $|\ell| < \beta$, then M'_ℓ can also be split into two $s/2$ qubit registers $(\mathsf{C}'_{\ell,0}, \mathsf{C}'_{\ell,1})$.
- Let U_r denote the unitary that the malicious prover applies to H to generate its response. We will abuse notation slightly and write U as shorthand for $\{U_r\}_{r\in R}$.[27]

Next, we define subroutines SwapRecover and SwapDiff, which will simplify the description of our extractor. Both of the subroutines are unitaries that can be implemented given oracle access to the prover unitary U_r for some challenge r.

- $\mathsf{SwapRecover}^{U_r}_{S,E}$ is implemented with oracle access to the prover unitary U_r for some r and is parameterized by two sets of tree indices $E \subseteq \{0,1\}^{\leq\beta}$ and $S \subseteq \mathsf{Path}(S_r)$, where S_r is the set of PCP (block) indices for challenge r. At any stage of the extraction procedure, E will denote the set of tree indices whose messages have already been extracted (i.e., swapped out) into M'. Roughly speaking, $\mathsf{SwapRecover}^{U_r}_{S,E}$ is the unitary that (1) runs the prover unitary U_r to generate its response to challenge r, (2) opens the decommitments to reveal the messages corresponding to indices in S, and (3) swaps in any messages in $E \cap S$.

  In our extraction procedure, calls to $\mathsf{SwapRecover}^{U_r}_{S,E}$ will use $S = \mathsf{Path}(S_r)$, where S_r is the set of indices of PCP blocks containing all the indices that the PCP verifier checks on challenge r. However, in order to prove correctness of our extractor, we will need to define hybrid extractors that run SwapRecover for subsets $S \subseteq \mathsf{Path}(S_r)$.

  Formally, $\mathsf{SwapRecover}^{U_r}_{S,E}$ acts on the prover's response register $(\mathsf{D}_\ell)_{\ell\in S}$, the tree commitment root C_ε, and the extracted registers $(\mathsf{M}'_\ell)_{\ell\in\{0,1\}^{\leq\beta}}$.

  $\underline{\mathsf{SwapRecover}^{U_r}_{S,E}}$:

  1. Apply U_r.
  2. For $j = 0, \ldots, \beta$:

     For each $\ell \in S \cap \{0,1\}^j$:

     (a) Apply $\mathsf{Com}^\dagger$ to $(\mathsf{C}_\ell, \mathsf{D}_\ell)$ to obtain $(\mathsf{M}_\ell, \mathsf{W}_\ell)$.[28]

     (b) If $\ell \in E$, apply $\mathsf{SWAP}[\mathsf{M}_\ell, \mathsf{M}'_\ell]$.

     (c) If $j < \beta$, split M_ℓ into two $s/2$-qubit registers $(\mathsf{C}_{\ell,0}, \mathsf{C}_{\ell,1})$.

  We emphasize that the outer loop must be performed in the order $j = 0, \ldots, \beta$. However, the operations in the inner loop commute for all $\ell \in S \cap \{0,1\}^j$.

[27] Even when R is an exponential-size set, the collection of unitaries $\{U_r\}$ can be compactly described by a single unitary U acting on $|r\rangle\langle r| \otimes \boldsymbol{\rho}_{\mathsf{H}}$.

[28] Technically, there are choices of S for which $\mathsf{SwapRecover}^{U_r}_{S,E}$ is not well-defined (in particular, this may occur if this C_ℓ has not been revealed in a previous loop). However, we will avoid such choices of S.

- We additionally define a unitary $\mathsf{SwapDiff}^{U_r}_{S,E}$ (for the same U_r, S, E as in SwapRecover) as follows:

  $\underline{\mathsf{SwapDiff}^{U_r}_{S,E}}$:

  1. Apply $\mathsf{SwapRecover}^{U_r}_{S,E}$.
  2. Apply $(\mathsf{SwapRecover}^{U_r}_{S,E\cup S})^\dagger$.

  In our extraction procedure, if the prover answers correctly on challenge r, we will apply $\mathsf{SwapDiff}^{U_r}_{S,E}$ with $S = \mathsf{Path}(S_r)$ in order to swap out any messages at tree indices in $\mathsf{Path}(S_r) \setminus E$. These indices correspond to tree nodes that (a) the prover opens on challenge r and (b) were not extracted in an earlier query.

For any set of extracted nodes $E \subseteq \{0,1\}^{\leq\beta}$, we define a set of projections $\Pi^U_E := \{\Pi^{U_r}_{r,E}\}_{r\in R}$ as

$$\Pi^{U_r}_{r,E} := (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E})^\dagger \left(\Pi^{\mathsf{PCP}}_r \otimes \bigotimes_{\ell\in\mathsf{Path}(S_r)} |0\rangle\langle 0|_{\mathsf{W}_\ell}\right) \mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E},$$

where Π^{PCP}_r is the projection the verifier applies to $\bigotimes_{j\in S_r} \mathsf{M}_j$ to check the proof on random coins r (see Definition 7.1; here we leave the dependence on the instance x implicit). In words, the projective measurement $\{\Pi^{U_r}_{r,E}, \mathbb{I} - \Pi^{U_r}_{r,E}\}$ measures whether the prover gives an accepting response on challenge r, where the response is computed by swapping in any previously-extracted messages from the registers $(\mathsf{M}'_i)_{i\in E}$ (whenever necessary).

We now define $\mathsf{Extract}^U_{\gamma,q,T}$, the primary subroutine in our full extractor KnowledgeExt. Extract essentially behaves identically to BitExtract, except that whenever it sends a challenge $r \leftarrow R$ and receives openings for messages in $\mathsf{Path}(S_r)$, it swaps out the messages in $\mathsf{Path}(S_r)$ that it has not yet extracted. $\gamma > 0$ is an error tolerance, q is a lower bound on the allowable success probability, and T is a runtime parameter (these parameters will be set when we define KnowledgeExt).

$\underline{\mathsf{Extract}^U_{\gamma,q,T}}$:

Input: A prover state on register H and a tree commitment register C_ε.

Setup: Define $\varepsilon = \gamma/4T$ and $\delta = (\gamma/16T)^2$. Initialize registers $\mathsf{M}' := (\mathsf{M}'_i)_{i\in[2^{\beta+1}-1]}$.[29] Initialize $E = \emptyset$.

1. For $t = 0, \ldots, T-1$:
   (a) Measure $\mathsf{Est}^{\Pi^U_E}_{\varepsilon,\delta} \to p_t$. If $p_t < q$, abort and output failure.
   (b) Sample a random $r \leftarrow R$ and apply the measurement $\{\Pi^{U_r}_{r,E}, \mathbb{I} - \Pi^{U_r}_{r,E}\}$.
   (c) If the measurement accepts, set $D := \Pi^{U_r}_{r,E}$. If the measurement rejects, set $D := \mathbb{I} - \Pi^{U_r}_{r,E}$. Perform the following steps:
      i. Apply $\mathsf{SwapDiff}^{U_r}_{\mathsf{Path}(S_r),E}$.

[29] We identify each tree node $\ell \in \{0,1\}^{\leq\beta}$ with an integer $i \in [2^{\beta+1}-1]$ following a level-order traversal, i.e., the root is $i = 1$, its children are $2, 3$, etc.

ii. Update $E \leftarrow E \cup \mathsf{Path}(S_r)$.

(d) Apply $\mathsf{Repair}^{\Pi_E^U, D}_{\varepsilon,\delta,p_t}$.

2. Measure $\mathsf{Est}^{\Pi_E^U}_{\varepsilon,\delta} \to p_T$. Return "success" if $p_T \geq q$ and "failure" otherwise. Additionally, output the registers $(\mathsf{M}'_i)_{i\in[2^{\beta+1}-1]\setminus[2^\beta-1]}$.

We now define KnowledgeExt.

$\underline{\mathsf{KnowledgeExt}^{\widetilde{P}}(x, 1^{\lceil 1/p\rceil}, 1^{\lceil 1/\gamma\rceil})}$:

1. Run the malicious prover $\widetilde{P}$ to obtain a tree commitment C.
2. Let $\boldsymbol{\tau}_{\mathsf{H},\mathsf{C}}$ denote the global state on the prover's registers H and the commitment register C. Sample a random $T \leftarrow \{1, 2, \ldots, \lceil 2^{\beta+3}/\gamma^2\rceil\}$ and run $\mathsf{Extract}^U_{\gamma/4,q+\gamma/4,T}(\boldsymbol{\tau}_{\mathsf{H},\mathsf{C}}) \to (b, \boldsymbol{\pi})$ where b is a bit indicating success/failure. Discard b and output $\boldsymbol{\pi}$.

8.3 Correctness of our extractor

In this subsection, we will complete the proof of our main theorem (Theorem 8.1) by showing that KnowledgeExt satisfies the conditions of Definition 7.6.

The main technical steps are captured by the following claim about the Extract subroutine.

Claim 8.7 (Success probability of the Extract subroutine). *For any state $\boldsymbol{\tau}_{\mathsf{H},\mathsf{C}}$ and efficient prover unitary U, let p denote the initial success probability, i.e., $p := \mathbb{E}_{r\leftarrow R}[\mathrm{Tr}(\Pi_r \boldsymbol{\tau})]$ where $\Pi_r := \Pi^{U_r}_{r,E=\emptyset}$. If* sQSC *is swap-binding, then for any $\gamma = 1/\mathsf{poly}(\lambda)$, any $q < p - \gamma - 1/\mathsf{poly}(\lambda)$, and any $T = \mathsf{poly}(\lambda)$, the subroutine $\mathsf{Extract}^U_{\gamma,q,T}$ succeeds with probability at least $p - q - \gamma - \mathsf{negl}(\lambda) = 1/\mathsf{poly}(\lambda)$.*

Proof outline. At a high level, our proof of Claim 8.7 consists of the following steps:

1. First, we write down a sequence of hybrid extraction procedures $\mathsf{HybExtract}[j]$, one for each tree node $[2^{\beta+1}-1]$. $\mathsf{HybExtract}[j]$ is defined similarly to Extract, except that the extractor only swaps out nodes in $[j]$. Thus, $\mathsf{HybExtract}[0]$ will correspond to BitExtract for which the [CMSZ21] statistical guarantees apply, and $\mathsf{HybExtract}[2^{\beta+1}-1]$ corresponds to Extract.
2. Next, we show that for all j, if an adversary distinguishes between $\mathsf{HybExtract}[j-1]$ and $\mathsf{HybExtract}[j]$, this implies a reduction that violates the *oracle swap binding* security (defined in Definition 6.1) of sQSC. We refer the reader to Section 2.3.2 for discussion on the necessity of oracle swap binding. Crucially, by Theorem 6.2, oracle swap binding security is *implied* by standard swap binding.

Proof of Claim 8.7. For $j \in \{0, \ldots, 2^{\beta+1}-1\}$, define $\mathsf{HybExtract}[j]$ as follows.

$\underline{\mathsf{HybExtract}^U_{\gamma,q,T}[j]}$ (differences from Extract highlighted in red):

Input: A prover state on register H and a tree commitment register C_ε.

Setup: Define $\varepsilon = \gamma/4T$ and $\delta = (\gamma/16T)^2$. Initialize registers $\mathsf{M}' := (\mathsf{M}'_i)_{i\in[2^{\beta+1}-1]}$. Initialize $E = \emptyset$.

1. For $t = 0, \ldots, T-1$:
   (a) Measure $\mathsf{Est}^{\Pi^U_E}_{\varepsilon,\delta} \to p_t$. If $p_t < q$, abort and output failure.
   (b) Sample a random $r \leftarrow R$ and apply the measurement $\{\Pi^{U_r}_{r,E}, \mathbb{I} - \Pi^{U_r}_{r,E}\}$.
   (c) If the measurement accepts, set $D := \Pi^{U_r}_{r,E}$. If the measurement rejects, set $D := \mathbb{I} - \Pi^{U_r}_{r,E}$. Perform the following steps:
      i. Apply $\mathsf{SwapDiff}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E}$.
      ii. Update $E \leftarrow E \cup (\mathsf{Path}(S_r)\cap[j])$.
   (d) Apply $\mathsf{Repair}^{\Pi^U_E, D}_{\varepsilon,\delta,p_t}$.
2. Measure $\mathsf{Est}^{\Pi^U_E}_{\varepsilon,\delta} \to p_T$. Return "success" if $p_T \geq q$ and "failure" otherwise. Additionally, output the registers $(\mathsf{M}'_i)_{i\in[2^{\beta+1}-1]\setminus[2^\beta-1]}$.

Observe that:

- $\mathsf{HybExtract}^U_{\gamma,q,T}[0]$ is equivalent to running $\mathsf{BitExtract}^{\Pi}_{\gamma,q,T}$ for $\Pi = \{\Pi_r\}$ (where $\Pi_r := \Pi^{U_r}_{r,E=\emptyset}$) and additionally outputting registers $(\mathsf{M}'_i)_{i\in[2^{\beta+1}-1]\setminus[2^\beta-1]}$ set to 0.
- $\mathsf{HybExtract}^U_{\gamma,q,T}[2^{\beta+1}-1]$ is identical to $\mathsf{Extract}^U_{\gamma,q,T}$.

We will prove that for all $j \in [2^{\beta+1}-1]$ and any efficient adversary (specified by some state $\boldsymbol{\tau}_{\mathsf{H},\mathsf{C}}$ and a unitary U),

$$\left|\Pr\left[\mathsf{HybExtract}^U_{\gamma,q,T}[j-1](\boldsymbol{\tau}) \to \text{success}\right] - \Pr\left[\mathsf{HybExtract}^U_{\gamma,q,T}[j](\boldsymbol{\tau}) \to \text{success}\right]\right| = \mathsf{negl}(\lambda).$$

Suppose otherwise. We will give a reduction $\mathsf{Reduction}[j]$ that wins the oracle swap binding game (Definition 6.1) using the commitment at index j. Before we state $\mathsf{Reduction}[j]$, we define the oracle G_b that we provide the reduction in the oracle swap binding game:

- Recall from Definition 6.1 that the adversary in the oracle swap binding game has access to a unitary G_b with the form

  $$G_b = (\widehat{\mathsf{SWAP}[\mathsf{M}_j, \mathsf{M}'_j]})^b \cdot \widehat{\mathcal{O}} \cdot (\widehat{\mathsf{SWAP}[\mathsf{M}_j, \mathsf{M}'_j]})^b \Pi + (\mathbb{I} - \Pi),$$

  for some unitary $\mathcal{O}$ chosen by the adversary and register M''_j that the challenger initializes to $|0\rangle$. Recall that $\Pi := (\mathsf{Com})(\mathbb{I}_{\mathsf{M}_j} \otimes |0\rangle\langle 0|_{\mathsf{W}_j})(\mathsf{Com}^\dagger)$ denotes the projection onto valid commitment-decommitment pairs, and that for any operation U, we write $\widehat{U} := (\mathsf{Com})U(\mathsf{Com}^\dagger)$.
- Let B be a one-qubit ancilla register. We will instantiate G_b for the following choice of $\mathcal{O} := \mathcal{O}_j$ for $\mathcal{O}_j$ defined as

  $$\mathcal{O}_j = \sum_{\substack{E\subseteq[2^{\beta+1}-1]\\ r\in R}} |E\rangle\langle E|_{\mathsf{E}} \otimes |r\rangle\langle r|_{\mathsf{Q}} \otimes \mathcal{O}_{r,E,j},$$

where the unitary $\mathcal{O}_{r,E,j}$ is defined as

$$\mathcal{O}_{r,E,j} \coloneqq \mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E} \cdot (X_{\mathsf{B}} \otimes \Pi^{U_r}_{r,E,j} + \mathbb{I}_{\mathsf{B}} \otimes (\mathbb{I} - \Pi^{U_r}_{r,E,j})) \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E})^\dagger$$

where

$$\Pi^{U_r}_{r,E,j} \coloneqq (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E})^\dagger \left(\Pi^{\mathsf{PCP}}_r \otimes \bigotimes_{\ell \in \mathsf{Path}(S_r)\setminus\{j\}} |0\rangle\langle 0|_{\mathsf{W}_\ell} \right) \mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E}.$$

This operation $\mathcal{O}$ is carefully defined to enable the reduction in a swap-binding security game to (a) run the (swap-augmented) prover on any classical challenge r and (b) run Est and Repair on the prover (which requires running the swap-augmented prover on a superposition of challenges r), *even after the challenger is given the* C_j *register*. Formally, $\mathcal{O}$ allows the adversary to pick any choice of r and E and apply a unitary we call $\mathcal{O}_{r,E,j}$. $\mathcal{O}_{r,E,j}$ is defined so that both of the following are true:

- The unitary

  $$(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E})^\dagger \cdot (\widehat{\mathcal{O}_{r,E,j}}\Pi + (\mathbb{I} - \Pi)) \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E})$$

  is equivalent to checking (and recording the outcome onto B) whether the swap-augmented prover in $\mathsf{HybExtract}[j-1]$ would make the verifier accept when run on challenge r when the extracted set is E.

- The unitary

  $$\begin{aligned}(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E})^\dagger \cdot (\widehat{\mathsf{SWAP}[\mathsf{M}_j, \mathsf{M}'_j]}) \cdot (\widehat{\mathcal{O}_{r,E,j}}\Pi + (\mathbb{I} - \Pi)) \\ \cdot (\widehat{\mathsf{SWAP}[\mathsf{M}_j, \mathsf{M}'_j]}) \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E})\end{aligned}$$

  is equivalent to checking (and recording the outcome onto B) whether the swap-augmented prover in $\mathsf{HybExtract}[j]$ would make the verifier accept when run on challenge r when the extracted set is E.

For $\mathcal{O}$ to be a valid oracle for the oracle swap binding game for the commitment at index j, it must be implemented as a circuit that acts only on M_j and other registers independent of the jth commitment (i.e., $\mathcal{O}$ should not be defined with respect to $\mathsf{C}_j, \mathsf{D}_j$, or W_j). An implementation of $\mathcal{O}$ that naively implements all the unitaries that appear in the description of $\mathcal{O}_{r,E,j}$ and $\Pi^{U_r}_{r,E,j}$ would *not* satisfy this guarantee, since, e.g., $(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E})^\dagger$ in the definition $\mathcal{O}_{r,E,j}$ acts on registers $\mathsf{C}_j, \mathsf{D}_j$ if $j \in \mathsf{Path}(S_r)$. However, we observe that to implement $\mathcal{O}_{r,E,j}$, it suffices to implement $(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E}) \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E})^\dagger$ and its inverse, and this operation can be implemented to satisfy the syntactic requirement. In particular, $(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r),E}) \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E})^\dagger$ is equivalent to the following procedure:

For each $\ell \in \mathsf{Path}(S_r)\setminus[j]$ in increasing order:

(a) Apply $\mathsf{Com}^\dagger$ to $(\mathsf{C}_\ell, \mathsf{D}_\ell)$ to obtain $(\mathsf{M}_\ell, \mathsf{W}_\ell)$.

(b) If $\ell \in E$, apply $\mathsf{SWAP}[\mathsf{M}_\ell, \mathsf{M}'_\ell]$.

At a high level, Reduction$[j]$ works as follows:

- It runs HybExtract$[j-1]$ as usual until the first time the adversary gives an accepting response containing a decommitment to node j (up to this point, there is no difference between running HybExtract$[j-1]$ or HybExtract$[j]$). In particular, it simply implements operations involving the projections $\Pi^{U_r}_{r,E}$ as normal.
- Once the adversary gives an accepting response containing a decommitment to node j, the reduction sends the valid commitment-decommitment pair to the oracle swap binding challenger corresponding to the commitment at node j. Since the challenger does not return the commitment register, the reduction can no longer implement operations involving the projections $\Pi^{U_r}_{r,E}$ properly. However, using the oracle G_b defined as above, the reduction can implement operations involving a projection $\Pi^{U_r,G_b}_{r,E}$ that we define as

$$\Pi^{U_r,G_b}_{r,E} := \langle 1|_{\mathsf{B}} \langle E|_{\mathsf{E}} \langle r|_{\mathsf{Q}} \widetilde{G}_b |0\rangle_{\mathsf{B}} |E\rangle_{\mathsf{E}} |r\rangle_{\mathsf{Q}},$$

  where

$$\widetilde{G}_b := (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E})^\dagger \cdot G_b \cdot (\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E}).$$

  By following the definitions, it can be verified that $\Pi^{U_r,G_b}_{r,E}$ has exactly the same behavior as $\Pi^{U_r}_{r,E}$, except that operations requiring the committed message in the node j commitment are implemented with the oracle G_b.

To simplify the description of Reduction$[j]$, we introduce a "flag" bit $\mathsf{flag} \in \{0,1\}$ that is initialized to 0 and flips to 1 after the reduction interacts with the oracle swap binding challenger. Define

$$\Pi^{U_r,G_b}_{r,E,\mathsf{flag}} := \begin{cases} \Pi^{U_r}_{r,E} & \text{if } \mathsf{flag} = 0 \\ \Pi^{U_r,G_b}_{r,E} & \text{if } \mathsf{flag} = 1, \end{cases} \tag{9}$$

and let $\Pi^{U,G_b}_{E,\mathsf{flag}} := \{\Pi^{U_r,G_b}_{r,E,\mathsf{flag}}\}_{r\in R}$.

We also need to modify the definitions of SwapRecover and SwapDiff, since the original syntax of these subroutines will not be applicable after the reduction sends commitment C_j to the oracle swap binding challenger.

- We define $\mathsf{SwapRecover}^{U_r}_{S,E,j}$, to behave the same as $\mathsf{SwapRecover}^{U_r}_{S,E}$ except that it does not swap out C_j. Formally, $\mathsf{SwapRecover}^{U_r}_{S,E,j}$ does the following:

  $\underline{\mathsf{SwapRecover}^{U_r}_{S,E,j}}$:

  - If $j = 1$ (i.e., j is the root), apply U_r.
  - Else:
    1. Apply $\mathsf{SwapRecover}_{S\cap[j-1]\setminus\{\mathsf{sib}(j),\mathsf{par}(j)\},E}$, where $\mathsf{sib}(j)$ is the sibling and $\mathsf{par}(j)$ is the parent of node j.

2. Apply $\mathsf{Com}^{\dagger}_{\mathsf{par}(j)}$. If $\mathsf{par}(j) \in E$, apply $\mathsf{SWAP}[\mathsf{C}_{\mathsf{sib}(j)}, \mathsf{C}'_{\mathsf{sib}(j)}]$.
3. If $\mathsf{sib}(j) \in S \cap [j-1]$, apply $\mathsf{Com}^{\dagger}_{\mathsf{sib}(j)}$. If $\mathsf{sib}(j) \in E$, apply $\mathsf{SWAP}[\mathsf{M}_{\mathsf{sib}(j)}, \mathsf{M}'_{\mathsf{sib}(j)}]$.

- Next, we define

  $\underline{\mathsf{SwapDiff}^{U_r}_{S,E,j}}$:

  1. Apply $\mathsf{SwapRecover}^{U_r}_{S,E,j}$.
  2. Apply $(\mathsf{SwapRecover}^{U_r}_{S,E\cup S,j})^{\dagger}$.

- Finally, we define $\mathsf{SwapDiff}^{U_r}_{S,E,j,\mathsf{flag}}$ as

$$\mathsf{SwapDiff}^{U_r}_{S,E,j,\mathsf{flag}} := \begin{cases} \mathsf{SwapDiff}^{U_r}_{S,E} & \text{if } \mathsf{flag} = 0 \\ \mathsf{SwapDiff}^{U_r}_{S,E,j} & \text{if } \mathsf{flag} = 1. \end{cases} \tag{10}$$

We now state the reduction $\mathsf{Reduction}[j]$ which plays the oracle swap-binding game and outputs a bit b'. The reduction has the same input and setup as $\mathsf{HybExtract}[j-1]$ and $\mathsf{HybExtract}[j]$ and uses the same values of ε and δ.

$\underline{\mathsf{Reduction}[j]}$:

1. Initialize $\mathsf{flag} = 0$.
2. For $t = 0, \ldots, T-1$:
   (a) Measure $\mathsf{Est}^{\Pi^{U,G_b}_{E,\mathsf{flag}}}_{\varepsilon,\delta} \to p_t$. If $p_t < q$, abort.
   (b) Sample a random $r \leftarrow R$ and apply the measurement $\{\Pi^{U_r,G_b}_{r,E,\mathsf{flag}}, \mathbb{I} - \Pi^{U_r,G_b}_{r,E,\mathsf{flag}}\}$.
   (c) If the measurement rejects, set $D := \mathbb{I} - \Pi^{U_r,G_b}_{r,E,\mathsf{flag}}$. If the measurement accepts, set $D := \Pi^{U_r,G_b}_{r,E,\mathsf{flag}}$ and perform the following steps:
      i. If $j \in \mathsf{Path}(S_r)$ and $\mathsf{flag} = 0$:
         A. Apply $\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E}$.
         B. Apply $\mathsf{SWAP}[\mathsf{C}_j, \mathsf{C}'_j]$ and send $\mathsf{D}_j, \mathsf{C}'_j$ to the oracle swap binding challenger.
         C. The challenger returns D_j.
         D. Update $E \leftarrow E \cup (\mathsf{Path}(S_r) \cap [j])$.
         E. Apply $(\mathsf{SwapRecover}^{U_r}_{\mathsf{Path}(S_r)\cap[j-1],E,j})^{\dagger}$.
         F. Set $\mathsf{flag} = 1$.
      ii. Otherwise:
         A. Apply $\mathsf{SwapDiff}^{U_r}_{\mathsf{Path}(S_r)\cap[j],E,j,\mathsf{flag}}$.
         B. Update $E \leftarrow E \cup (\mathsf{Path}(S_r) \cap [j])$.
   (d) Apply $\mathsf{Repair}^{\Pi^{U,G_b}_{E,\mathsf{flag}},D}_{\varepsilon,\delta,p_t}$.
3. Measure $\mathsf{Est}^{\Pi^{U,G_b}_{E,\mathsf{flag}}}_{\varepsilon,\delta} \to p_T$. Guess $b' = 0$ if $p_T \geq q$ and guess $b' = 1$ otherwise.

If the oracle swap binding challenger's bit is $b = 0$, then $\Pr[\mathsf{Reduction}[j] \to 0]$ is equal to $\Pr[\mathsf{HybExtract}[j-1] \text{ succeeds}]$. If the challenger's bit is $b = 1$, then $\Pr[\mathsf{Reduction}[j] \to 0]$ is equal to $\Pr[\mathsf{HybExtract}[j] \text{ succeeds}]$. Thus, by oracle swap binding security, we have

$$\Pr[\mathsf{HybExtract}[j] \text{ succeeds}] \geq \Pr[\mathsf{HybExtract}[j-1] \text{ succeeds}] - \mathsf{negl}(\lambda).$$

Since $2^\beta = \mathsf{poly}(\lambda)$, it follows from Claim 8.6 that

$$\Pr\left[\mathsf{Extract}^U_{\gamma,q,T}\right] \geq p - q - \gamma - \mathsf{negl}(\lambda). \qquad \square$$

Why does KnowledgeExt produce a valid PCP? Claim 8.7 shows that after running Extract, the swap-augmented prover still has high success probability. To complete the proof of Theorem 8.1, it remains to show that when T is sampled uniformly at random from $\{1, 2, \ldots, \lceil 2^{\beta+3}/\gamma^2 \rceil\}$ in the KnowledgeExt procedure, with noticeable probability the subroutine $\mathsf{Extract}^U_{\gamma/4,q+\gamma/4,T}$ outputs a good PCP string.[30]

Claim 8.8. *For any state $\boldsymbol{\tau}_{\mathsf{H},\mathsf{C}}$ and efficient prover unitary U, let p denote the initial success probability, i.e., $p := \mathbb{E}_{r\leftarrow R}[\mathrm{Tr}(\Pi_r \boldsymbol{\tau})]$ where $\Pi_r := \Pi^{U_r}_{r,E=\emptyset}$. If sQSC is swap-binding, then for any $\gamma = 1/\mathsf{poly}(\lambda)$ and any $q < p - \gamma - 1/\mathsf{poly}(\lambda)$, with probability at least $p - q - \gamma - \mathsf{negl}(\lambda) \geq 1/\mathsf{poly}(\lambda)$ the procedure $\mathsf{KnowledgeExt}^{\widetilde{P}}(x, 1^{\lceil 1/p \rceil}, 1^{\lceil 1/\gamma \rceil})$ outputs an extracted proof $\boldsymbol{\pi}$ such that*

$$\mathbb{E}_{r\leftarrow R}[\mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_r \boldsymbol{\pi}\right)] \geq q - \mathsf{negl}(\lambda).$$

Proof. Let $\boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}}$ be the joint state of both the prover and extractor after running KnowledgeExt (not to be confused with the state $\boldsymbol{\pi}$ output by KnowledgeExt). By Claim 8.7, with probability at least $p - q - \gamma/2 - \mathsf{negl}(\lambda)$ the subroutine $\mathsf{Extract}^U_{\gamma/4,q+\gamma/4,T}$ outputs success and

$$\mathbb{E}_{r\leftarrow R}[\mathrm{Tr}\left(\Pi^{U_r}_{r,E} \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}}\right)] \geq q + \gamma/4.$$

We divide the M' register into $\mathsf{M}' = (\mathsf{M}'_{\text{internal}}, \mathsf{M}'_{\text{leaves}})$ so that the extracted proof corresponds to $\mathsf{M}'_{\text{leaves}}$. For any state $\boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}}$, the extracted proof $\boldsymbol{\pi} = \mathrm{Tr}_{\mathsf{C},\mathsf{H},\mathsf{M}'_{\text{internal}}}(\boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}})$ satisfies

$$\begin{aligned}
\mathbb{E}_{r\leftarrow R}[\mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_r \boldsymbol{\pi}\right)] &= \Pr_{r\leftarrow R}[\Pi^{\mathsf{PCP}}_r \text{ accepts } \boldsymbol{\pi}] \\
&\geq \Pr_{r\leftarrow R}[\Pi^{\mathsf{PCP}}_r \text{ accepts } \boldsymbol{\pi} \text{ and } S_r \subseteq E] \\
&\geq \Pr_{r\leftarrow R}[\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}} \text{ and } S_r \subseteq E] \\
&= \Pr_{r\leftarrow R}[\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}}] - \Pr_{r\leftarrow R}[\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}} \text{ and } S_r \not\subseteq E] \\
&= \mathbb{E}_{r\leftarrow R}[\mathrm{Tr}\left(\Pi^{U_r}_{r,E} \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}}\right)] - \Pr_{r\leftarrow R}[\{\Pi^{U_r}_{r,E}, \mathbb{I} - \Pi^{U_r}_{r,E}\} \text{ accepts on } \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}} \text{ and } S_r \not\subseteq E] \\
&\geq q + \gamma/4 - \Pr_{r\leftarrow R}[\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho}_{\mathsf{M}',\mathsf{C},\mathsf{H}} \text{ and } S_r \not\subseteq E].
\end{aligned}$$

[30] A similar claim is proved in [CMSZ21, LMS22b], but their argument is specialized for extracting classical proof strings and works by taking a union bound over all possible proofs. We therefore require a different proof for this final step.

The second inequality follows because the measurement $\{\Pi^{U_r}_{r,E}, \mathbb{I} - \Pi^{U_r}_{r,E}\}$ checks that both the measurement $\{\Pi^{\mathsf{PCP}}_r, \mathbb{I} - \Pi^{\mathsf{PCP}}_r\}$ accepts $\boldsymbol{\pi}$ and all commitment-decommitment pairs along the root to leaf path are valid.

For any state $\boldsymbol{\rho}$ on registers $(\mathsf{M}', \mathsf{C}, \mathsf{H})$, define

$$x(\boldsymbol{\rho}) = \Pr_{r \leftarrow R}[\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho} \text{ and } S_r \not\subseteq E].$$

We will show that $\Pr_{\boldsymbol{\rho} \leftarrow \mathsf{KnowledgeExt}}[x(\boldsymbol{\rho}) > \gamma/4] \leq \gamma/2$, which will complete the proof.

Recall that $\mathsf{KnowledgeExt}^{\widetilde{P}}(x, 1^{\lceil 1/p \rceil}, 1^{\lceil 1/\gamma \rceil})$ runs $\mathsf{Extract}_{q+\gamma/4,\gamma/4,T}$, stopping at a random time $t \leftarrow [T]$. Consider an execution of $\mathsf{Extract}_{q+\gamma/4,\gamma/4,T+1}$, and let e_t denote the event that both $b_t = 1$ (i.e., the measurement $\{\Pi^{U_r}_{r,E_t}, \mathbb{I} - \Pi^{U_r}_{r,E_t}\}$ accepts for the randomly chosen r at step t) and $S_t \not\subseteq E_t$, where S_t and E_t are the sets S_r and E, respectively, at step t. Observe that

$$\begin{aligned}
\mathop{\mathbb{E}}_{\boldsymbol{\rho} \leftarrow \mathsf{KnowledgeExt}}[x(\boldsymbol{\rho})] &= \mathop{\mathbb{E}}_{t \leftarrow [T]} \mathop{\mathbb{E}}_{\boldsymbol{\rho} \leftarrow \mathsf{Extract}_{q+\gamma/4,\gamma/4,t}} [x(\boldsymbol{\rho})] && (11)\\
&= \mathop{\mathbb{E}}_{t \leftarrow [T]} \mathop{\Pr}_{\substack{\boldsymbol{\rho} \leftarrow \mathsf{Extract}_{q+\gamma/4,\gamma/4,t} \\ r \leftarrow R}} [\Pi^{U_r}_{r,E} \text{ accepts } \boldsymbol{\rho} \text{ and } S_r \not\subseteq E] && (12)\\
&= \mathop{\mathbb{E}}_{t \leftarrow \{2,\ldots,T+1\}} \Pr[e_t] && (13)
\end{aligned}$$

Since there are only 2^β blocks in the proof and each time e_t happens the extractor will swap out a new block, we can have at most 2^β events e_t, i.e.,

$$\frac{1}{T} \sum_{t=2}^{T+1} \mathbf{1}_{e_t} \leq \frac{2^\beta}{T},$$

where $\mathbf{1}_{e_t}$ is the indicator for the event e_t.

Taking expectations and applying Eq. (11),

$$\mathop{\mathbb{E}}_{\boldsymbol{\rho}}[x(\boldsymbol{\rho})] \leq \frac{2^\beta}{T}.$$

Finally, by Markov's inequality we have

$$\Pr_{\boldsymbol{\rho}}[x(\boldsymbol{\rho}) > \gamma/4] \leq \frac{2^{\beta+2}}{T \cdot \gamma} \leq \frac{\gamma}{2}. \qquad \square$$

9 Computationally sound quantum sigma protocols

In this section, we use our techniques to prove computational soundness for *quantum sigma protocols* instantiated with our hiding and binding QSCs. This gives a quantum analogue of a standard template for building classical sigma protocols from any *zero-knowledge PCP* (zk-PCP) and any hiding-binding commitment. While quantum sigma protocols were previously considered by Broadbent and Grilo [BG20], the results in this section differ from [BG20] in several respects:

- As discussed in Section 2.4, the [BG20] analysis of computational soundness was incomplete because it was missing the rewinding analysis.
- As an additional contribution, we also place their protocol in a general framework using our new abstraction of hiding and binding QSCs. We also consider a slightly more general definition of quantum zero-knowledge PCPs than [BG20].

We begin by defining quantum zk-PCPs, a quantum analogue of classical zk-PCPs.[31] Since we are not concerned with succinctness in this section, it suffices to consider quantum PCPs with soundness $1 - 1/\mathsf{poly}(n)$, which are known for all of QMA [BG20].

Remark 9.1. *Our definition of quantum zk-PCPs differs slightly from the the definition of* locally simulatable proofs *due to [BG20]. In particular, we require the simulator generates the mixed state corresponding to the honest PCP verifier's view, rather than (a classical description of) the reduced density matrix* for all subsets *of qubits up to a certain size as in [BG20]. Our definition captures both the locally simulatable proofs of [BG20] as well as classical zk-PCPs such as [GMW86] 3-coloring.*[32]

Definition 9.2 (Quantum zero-knowledge PCPs)**.** *A quantum zero-knowledge PCP for a language L is parameterized by a completeness parameter c, soundness s, proof length m, randomness complexity ℓ, and query complexity q. We require the following properties:*

- *(Efficient verification) There is a classical* $\mathsf{poly}(n)$ *time procedure that takes as input $x \in \{0,1\}^n$ and $r \in \{0,1\}^\ell$ and outputs the description of circuit for implementing a q-qubit projective measurement $\{\Pi^{\mathsf{PCP}}_{x,r}, \mathbb{I} - \Pi^{\mathsf{PCP}}_{x,r}\}$, which acts on a state of size m.*
- *(Honest-verifier zero knowledge) Let $Q_r \subset [m]$ denote the size-q subset of indices that $\{\Pi^{\mathsf{PCP}}_{x,r}, \mathbb{I} - \Pi^{\mathsf{PCP}}_{x,r}\}$ checks. There exists an efficient quantum algorithm* $\mathsf{PCPSim}(x)$ *that outputs a mixed state of the form*

$$\frac{1}{2^\ell} \sum_{r \in \{0,1\}^\ell} |r\rangle\langle r| \otimes \boldsymbol{\rho}_{Q_r},$$

such that for any $x \in L$, there exists an m-qubit locally simulatable proof $\boldsymbol{\pi}$ satisfying completeness, i.e.,

$$\mathop{\mathbb{E}}_{r \leftarrow \{0,1\}^\ell} \mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_{x,r} \boldsymbol{\pi}\right) \geq c,$$

[31] We refer the reader to Ishai's blog post [Ish20] for background on classical zk-PCPs.

[32] The 3-coloring zk-PCP does not satisfy the [BG20] definition of a locally simulatable proof, since it is not possible to generate the verifier's view for pairs of vertices that do not share an edge.

and the following zero-knowledge property:

$$\left\| \frac{1}{2^\ell} \sum_{r\in\{0,1\}^\ell} |r\rangle\langle r| \otimes \mathrm{Tr}_{\overline{Q_r}}(\boldsymbol{\pi}) - \frac{1}{2^\ell} \sum_{r\in\{0,1\}^\ell} |r\rangle\langle r| \otimes \boldsymbol{\rho}_{Q_r} \right\|_1 = \mathsf{negl}(n).$$

- *(Soundness) If $x \notin L$, then for any m-qubit state $\boldsymbol{\pi}$,*

$$\mathop{\mathbb{E}}_{r\leftarrow\{0,1\}^\ell} \mathrm{Tr}\left(\Pi^{\mathsf{PCP}}_{x,r}\boldsymbol{\pi}\right) \leq s.$$

Lemma 9.3 ([BG20]). *Every language in* QMA *has a zero-knowledge quantum PCP with completeness $c(n) \geq 1 - \mathsf{negl}(n)$ and soundness $s(n) \leq 1 - 1/\mathsf{poly}(n)$.*

We remark that because our definition allows mixed state PCPs, any classical zk-PCP is also a quantum zk-PCP. In particular, this means the zk-PCPs underlying the [GMW86] 3-coloring protocol and the [IKOS07] MPC-in-the-head protocols fall into our framework.

A generic quantum sigma protocol. We now describe a generic template for quantum sigma protocols. The protocol is parameterized by a choice of zk-PCP, zkPCP, and quantum commitment scheme, QSC, where:

- zkPCP is quantum zk-PCP for a language L, and
- QSC is a hiding-binding QSC for one-qubit messages.

For any zkPCP and QSC satisfying the above, we denote the resulting quantum sigma protocol by $\mathsf{QSigma}[\mathsf{zkPCP}, \mathsf{QSC}]$.[33]

$\mathsf{QSigma}[\mathsf{zkPCP}, \mathsf{QSC}]$:

Prover input: x and a corresponding zk-PCP $\boldsymbol{\rho}$ on $m = \mathsf{poly}(n)$ qubits $\bigotimes_{i=1}^m \mathsf{M}_i$.

Verifier input: x.

1. The prover applies $\mathsf{Com}_{\mathsf{QSC}}(\mathsf{M}_i) \to (\mathsf{C}_i, \mathsf{D}_i)$ and sends C_i for all $i \in [m]$ to the verifier.
2. The verifier sends random coins r to the prover.
3. The prover sends D_i for all $i \in Q_r$ to the verifier.
4. The verifier applies $\mathsf{Com}^\dagger_{\mathsf{QSC}}(\mathsf{C}_i, \mathsf{D}_i) \to (\mathsf{M}_i, \mathsf{W}_i)$ and measures W_i with $\{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$ for all $i \in Q_r$; it aborts if the outcome is not $|0\rangle\langle 0|$. Finally, the verifier measures $\{\Pi^{\mathsf{PCP}}_r, \mathbb{I} - \Pi^{\mathsf{PCP}}_r\}$. It accepts if the measurement accepts.

[33] Broadbent and Grilo [BG20] presented this protocol in the special case that the QSC was defined by the folklore construction (see Section 5.3).

9.1 Argument of knowledge

Theorem 9.4. *Let* zkPCP *be a probabilistically checkable proof for a language L. Then* QSigma[zkPCP, QSC] *is a computational (resp. statistical) proof of knowledge for L with respect to the* zkPCP *verifier (Definition 7.6) if* QSC *satisfies computational (resp. statistical) binding.*

Proof. The proof strategy is nearly identical to the proof of Theorem 8.1 and we will avoid repeating it in full. The only difference arises from the fact that instead of a Merkle tree of commitments to the PCP, we have independent commitments for every block of the zkPCP proof string. Thus, we can use the same proof as Theorem 8.1 with the following (syntactic) simplifications:

- There are now m explicit commitments instead of $2^{\beta+1}-1$ (implicit) commitments. Correspondingly, the $\mathsf{HybExtract}_j$ procedures only need to be defined for $j \in \{0,1,\ldots,m\}$.
- For any $S \subseteq [m]$, replace $\mathsf{Path}(S)$ with S.
- For any $S \subseteq [m]$ and $E \subseteq [m]$, replace $\mathsf{SwapRecover}^{U_r}_{S,E}$ with the algorithm that applies $\mathsf{Com}^{\dagger}_{\mathsf{QSC}}$ to $(\mathsf{C}_i, \mathsf{D}_i)$ to obtain $(\mathsf{M}_i, \mathsf{W}_i)$ for all $i \in S$. If $i \in E$, apply $\mathsf{SWAP}[\mathsf{M}_i, \mathsf{M}'_i]$ as well.
- For any $S \subseteq [m]$, $E \subseteq [m]$, $j \in [m]$, replace $\mathsf{SwapRecover}^{U_r}_{S,E,j}$ with $\mathsf{SwapRecover}^{U_r}_{S\cap[j-1],E}$ (for $\mathsf{SwapRecover}^{U_r}_{S,E}$ as defined in the previous bullet). □

9.2 Zero knowledge

We now prove that our quantum sigma protocol is zero knowledge against malicious verifiers whenever the size of the challenge space $\{0,1\}^\ell$ for the underlying zkPCP is at most $\mathsf{poly}(\lambda)$. This is a standard application of Watrous's zero-knowledge rewinding lemma, which we recall below. We refer the reader to [Wat09, Section 3] for a definition of zero-knowledge against quantum attacks.[34]

Lemma 9.5 ([Wat09])**.** *Let Q be a quantum circuit that acts on $n+k$ qubits, where the first n qubits may take an arbitrary state $|\psi\rangle$ as input and the remaining k qubits are initially set to the state $|0^k\rangle$. For any n-qubit state $|\psi\rangle$, define $p(\psi) \in [0,1]$ and (normalized) states $|\phi_0(\psi)\rangle, |\phi_1(\psi)\rangle$ so that*

$$Q|\psi\rangle\,|0^k\rangle = \sqrt{p(\psi)}\,|0\rangle\,|\phi_0(\psi)\rangle + \sqrt{1-p(\psi)}\,|1\rangle\,|\phi_1(\psi)\rangle\,. \tag{14}$$

Suppose that there exists $p_0, q \in (0,1)$ and $\varepsilon \in (0,1/2)$ such that for all n-qubit states $|\psi\rangle$,

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

[34] Even though Watrous's paper focuses on post-quantum security of classical protocols, Watrous's zero-knowledge definitions only refer to the admissible super-operators (CPTP maps) induced by the interaction with the prover (or the simulator) and thus apply equally well to quantum-communication protocols.

Then there exists a general quantum circuit R of size

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

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

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

Theorem 9.6. *Suppose* zkPCP *has randomness complexity $\ell = O(\log\lambda)$. Then if* QSC *satisfies computational (resp. statistical) hiding,* QSigma[zkPCP, QSC] *satisfies computational (resp. statistical) zero knowledge.*

Proof. For any instance x and security parameter λ, we will define a unitary $U_{x,\lambda}$ acting on the following registers:

- An m-qubit register $\mathsf{M} = (\mathsf{M}_1, \ldots, \mathsf{M}_m)$ where each M_i is a one-qubit register.
- A dm-qubit register $\mathsf{W} = (\mathsf{W}_1, \ldots, \mathsf{W}_m)$ where each W_i is a $d(\lambda)$-qubit register.
- A pair of ℓ-qubit registers R, R'. When we run the simulator, R will be a register containing uniform PCP verifier randomness r and R' will contain the string output by the malicious verifier $\widetilde{V}$.
- An ancilla register S corresponding to the workspace of PCPSim_x.
- A register V corresponding to the auxiliary quantum input for the malicious verifier $\widetilde{V}$.
- A one-qubit register B.

All registers are initialized to $|0\rangle$ except V, which is initialized to the auxiliary quantum input of $\widetilde{V}$.

For a size-q subset $Q_r \subset [m]$ and q-qubit state $\boldsymbol{\rho}_{Q_r}$, let $\boldsymbol{\rho}_{Q_r,\overline{Q}_r}$ denote the m-qubit state whose reduced density matrix on indices Q_r is $\boldsymbol{\rho}_{Q_r}$ and is all-zero everywhere else. We will assume that PCPSim_x is a unitary acting on $\mathsf{R}, \mathsf{M}, \mathsf{S}$ that, when applied to the all-zero state, produces a state whose reduced density matrix on (R, M) has the form

$$\frac{1}{2^\ell}\sum_{r\in\{0,1\}^\ell} |r\rangle\langle r|_{\mathsf{R}} \otimes \boldsymbol{\rho}_{Q_r,\overline{Q}_r}.$$

We now define the unitary $U_{x,\lambda}$ to perform the following steps:

1. Apply the unitary PCPSim_x to $(\mathsf{R}, \mathsf{M}, \mathsf{S})$.
2. For each index $i \in [m]$, apply $\mathsf{Com}_{\mathsf{QSC}}$ to $(\mathsf{M}_i, \mathsf{W}_i)$ to obtain $(\mathsf{C}_i, \mathsf{D}_i)$.
3. Apply the malicious verifier unitary $\widetilde{V}$ on $(\mathsf{C}_1, \ldots, \mathsf{C}_m, \mathsf{V}, \mathsf{R}')$.

4. Apply the unitary

$$\mathbb{I}_{\mathsf{B}} \otimes \sum_{r\in\{0,1\}^\ell} |r\rangle\langle r|_{\mathsf{R}} \otimes |r\rangle\langle r|_{\mathsf{R}'} + X_{\mathsf{B}} \otimes \sum_{r,s\in\{0,1\}^\ell, r\neq s} |r\rangle\langle r|_{\mathsf{R}} \otimes |s\rangle\langle s|_{\mathsf{R}'}\,,$$

which applies the bit-flip operator X to B controlled on the contents of R and R' being different (in the standard basis).

Fix any $x \in L$. We will show that there exists a negligible function $\mu(\lambda)$ and a constant λ_0 such that for all $\lambda \geq \lambda_0$, the unitary $U_{x,\lambda}$ satisfies the conditions of Lemma 9.5 for $p_0 = 1/2^{\ell+1}$, $q = 1/2^\ell$, and $\varepsilon = \mu(\lambda)$ where:

- $|\psi\rangle$ in Lemma 9.5 corresponds to the auxiliary input state on V,
- $|0^k\rangle$ in Lemma 9.5 corresponds to the initial state on registers $(\mathsf{M}, \mathsf{W}, \mathsf{R}, \mathsf{R}', \mathsf{S}, \mathsf{B})$,
- and the left-most qubit on the right-hand side of Eq. (14) corresponds to B.

Suppose otherwise. Then there exists a family of states $\{|\psi_\lambda\rangle_{\mathsf{V}}\}_{\lambda\in\mathbb{N}}$ and a constant $c > 0$ such that for infinitely many λ, we have $\left|p(\psi_\lambda) - 1/2^\ell\right| > 1/\lambda^c$. We will give a reduction breaking hiding (Definition 4.3) of the m-qubit QSC corresponding to the m different 1-qubit QSCs in the protocol; this can be turned into a reduction breaking the hiding of one of the 1-qubit QSCs by a standard hybrid argument. The reduction works as follows:

1. Run PCPSim_x to obtain the mixed state

$$\frac{1}{2^\ell} \sum_{r\in\{0,1\}^\ell} |r\rangle\langle r|_{\mathsf{R}} \otimes (\boldsymbol{\rho}_{Q_r,\overline{Q}_r})_{\mathsf{M}}.$$

2. Send $(\mathsf{M}_1, \ldots, \mathsf{M}_m)$ to the QSC hiding challenger.
3. The challenger returns commitments $(\mathsf{C}_1, \ldots, \mathsf{C}_m)$. Run $\widetilde{V}$ on C to obtain a challenge register R'.
4. Measure R and R' in the standard basis. If the two outcomes are different, output a random bit b'. If the outcome is the same, then:
   - if $p(\psi_\lambda) > 1/2^\ell + 1/\lambda^c$, guess $b' = 0$, and
   - if $p(\psi_\lambda) < 1/2^\ell - 1/\lambda^c$, guess $b' = 1$.

If the challenger's bit is $b = 1$ in the hiding experiment (Definition 4.3), then the challenger returns commitments $(\mathsf{C}_1, \ldots, \mathsf{C}_m)$ where the underlying messages are all-zero states. In this case, the probability that the reduction obtains the same challenge when it measures R and R' is $1/2^\ell$, since R is a uniform mixture over $r \in \{0,1\}^\ell$ independent of the malicious verifier's view.

If the challenger's bit is $b = 0$, then the probability that the reduction obtains the same challenge when it measures R and R' is $p(\psi_\lambda)$. An elementary calculation shows that the reduction guesses b with probability $1/2 + O(1/\lambda^c)$, which violates the hiding of the QSC.

Thus, by applying Lemma 9.5, we can efficiently generate a state $\boldsymbol{\rho}(\psi_\lambda)$ such that $\boldsymbol{\rho}(\psi_\lambda)$ is within $1 - \mathsf{negl}(\lambda)$ trace distance from $\phi_0(\psi_\lambda)$, the (normalized) state obtained by applying $U_{x,\lambda}$ on $|\psi_\lambda\rangle_{\mathsf{V}} |0\rangle_{\mathsf{M,W,R,R',S,B}}$ and post-selecting on $\mathsf{B} = 0$. With one additional step, we can generate the simulated view: controlled on the value r in R', generate a response register Z containing D_i for each $i \in Q_r$. The resulting simulated malicious verifier view is $(\mathsf{C}_1, \ldots, \mathsf{C}_m, \mathsf{R}', \mathsf{Z}, \mathsf{V})$. We omit the formal details of the rest of the proof, which are a direct analogue of Watrous's proof [Wat09]. In short, the hiding of the commitments together with the zero-knowledge property of the quantum zk-PCP implies that the simulated view is indistinguishable from a view in which the m committed qubits are a valid zk-PCP state $\boldsymbol{\pi}$. The latter corresponds to the view of the malicious verifier in the interaction with the honest prover initialized with $\boldsymbol{\pi}$. □

Acknowledgments

We thank Prabhanjan Ananth, James Bartusek, Andrea Coladangelo, Yael Kalai, Willy Quach, Nicholas Spooner, Vinod Vaikuntanathan and Umesh Vazirani for helpful discussions. We thank Luowen Qian for his detailed feedback on an early draft of this manuscript. We thank Alex Lombardi for many insightful conversations throughout the course of this project and countless suggestions that greatly improved the paper.

References

[AALV09] Dorit Aharonov, Itai Arad, Zeph Landau, and Umesh Vazirani. The detectability lemma and quantum gap amplification. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, page 417–426, New York, NY, USA, 2009. Association for Computing Machinery. 4, 5, 51, 56

[AAS20] Scott Aaronson, Yosi Atia, and Leonard Susskind. On the hardness of detecting macroscopic superpositions, 2020. 44

[AAV13] Dorit Aharonov, Itai Arad, and Thomas Vidick. The quantum PCP conjecture. *CoRR*, abs/1309.7495, 2013. 4, 5, 51, 56

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Preliminary version in FOCS '92. 4, 5, 51, 56

[AQY22] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. Cryptography from pseudorandom quantum states. 2022. https://ia.cr/2021/1663. 4, 10, 88

[ARU14] Andris Ambainis, Ansis Rosmanis, and Dominique Unruh. Quantum attacks on classical proof systems: The hardness of quantum rewinding. In *55th FOCS*, pages 474–483. IEEE Computer Society Press, October 2014. 6, 14

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Preliminary version in FOCS '92. 4, 5, 51, 56

[BB21] Nir Bitansky and Zvika Brakerski. Classical binding for quantum commitments. In Kobbi Nissim and Brent Waters, editors, *TCC 2021, Part I*, volume 13042 of *LNCS*, pages 273–298. Springer, Heidelberg, November 2021. 18

[BCG+02] Howard Barnum, Claude Crépeau, Daniel Gottesman, Adam Smith, and Alain Tapp. Authentication of quantum messages. In *43rd FOCS*, pages 449–458. IEEE Computer Society Press, November 2002. 18, 19

[BCQ22] Zvika Brakerski, Ran Canetti, and Luowen Qian. On the computational hardness needed for quantum cryptography. Cryptology ePrint Archive, Paper 2022/1181, 2022. https://eprint.iacr.org/2022/1181. 4

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *23rd ACM STOC*, pages 21–31. ACM Press, May 1991. 4, 5, 51, 56

[BG20] Anne Broadbent and Alex B. Grilo. QMA-hardness of consistency of local density matrices with applications to quantum zero-knowledge. In *61st FOCS*, pages 196–205. IEEE Computer Society Press, November 2020. 1, 5, 17, 69, 70

[BG22] Anne Broadbent and Alex Bredariol Grilo. Qma-hardness of consistency of local density matrices with applications to quantum zero-knowledge. *SIAM Journal on Computing*, 51(4):1400–1450, 2022. 17

[BJ15] Anne Broadbent and Stacey Jeffery. Quantum homomorphic encryption for circuits of low T-gate complexity. In Rosario Gennaro and Matthew J. B. Robshaw, editors, *CRYPTO 2015, Part II*, volume 9216 of *LNCS*, pages 609–629. Springer, Heidelberg, August 2015. 3, 9, 23

[BJSW16] Anne Broadbent, Zhengfeng Ji, Fang Song, and John Watrous. Zero-knowledge proof systems for QMA. In Irit Dinur, editor, *57th FOCS*, pages 31–40. IEEE Computer Society Press, October 2016. 1

[BKL+22] James Bartusek, Yael Tauman Kalai, Alex Lombardi, Fermi Ma, Giulio Malavolta, Vinod Vaikuntanathan, Thomas Vidick, and Lisa Yang. Succinct classical verification of quantum computation. CRYPTO '22, 2022. 18, 20

[BKP18] Nir Bitansky, Yael Tauman Kalai, and Omer Paneth. Multi-collision resistance: a paradigm for keyless hash functions. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *50th ACM STOC*, pages 671–684. ACM Press, June 2018. 5

[CCY20] Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. Classical verification of quantum computations with efficient verifier. In Rafael Pass and Krzysztof Pietrzak, editors, *TCC 2020, Part III*, volume 12552 of *LNCS*, pages 181–206. Springer, Heidelberg, November 2020. 18

[CCY21] Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. A black-box approach to post-quantum zero-knowledge in constant rounds. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part I*, volume 12825 of *LNCS*, pages 315–345, Virtual Event, August 2021. Springer, Heidelberg. 14

[CM21] Lijie Chen and Ramis Movassagh. Quantum merkle trees. *CoRR*, abs/2112.14317, 2021. 1, 2, 4, 13, 17, 53

[CMSZ21] Alessandro Chiesa, Fermi Ma, Nicholas Spooner, and Mark Zhandry. Post-quantum succinct arguments: Breaking the quantum rewinding barrier. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2021, Denver, CO, USA, February 7-10, 2022*, pages 49–58. IEEE, 2021. 2, 5, 13, 14, 15, 16, 17, 52, 53, 56, 57, 58, 59, 62, 67

[CVZ20] Andrea Coladangelo, Thomas Vidick, and Tina Zhang. Non-interactive zero-knowledge arguments for QMA, with preprocessing. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part III*, volume 12172 of *LNCS*, pages 799–828. Springer, Heidelberg, August 2020. 1

[DS22] Marcel Dall'Agnol and Nicholas Spooner. On the necessity of collapsing. Cryptology ePrint Archive, Report 2022/786, 2022. https://eprint.iacr.org/2022/786. 88

[FGL+91] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating clique is almost NP-complete (preliminary version). In *32nd FOCS*, pages 2–12. IEEE Computer Society Press, October 1991. 4, 5, 51, 56

[GMW86] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity and a methodology of cryptographic protocol design (extended abstract). In *27th FOCS*, pages 174–187. IEEE Computer Society Press, October 1986. 5, 69, 70

[GMW87] Oded Goldreich, Silvio Micali, and Avi Wigderson. How to prove all NP-statements in zero-knowledge, and a methodology of cryptographic protocol design. In Andrew M. Odlyzko, editor, *CRYPTO'86*, volume 263 of *LNCS*, pages 171–185. Springer, Heidelberg, August 1987. 7

[Har05] Aram Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, 2005. 91

[Har13] Aram W. Harrow. The church of the symmetric subspace, 2013. 91

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999. 11, 29

[IKOS07] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Zero-knowledge from secure multiparty computation. In David S. Johnson and Uriel Feige, editors, *39th ACM STOC*, pages 21–30. ACM Press, June 2007. 5, 70

[Ish20] Yuval Ishai. Zero-knowledge proofs from information-theoretic proof systems - part i. https://zkproof.org/2020/08/12/information-theoretic-proof-systems/, 2020. 69

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. Pseudorandom quantum states. In Hovav Shacham and Alexandra Boldyreva, editors, *CRYPTO 2018, Part III*, volume 10993 of *LNCS*, pages 126–152. Springer, Heidelberg, August 2018. 12, 91

[Kil92] Joe Kilian. A note on efficient zero-knowledge proofs and arguments (extended abstract). In *24th ACM STOC*, pages 723–732. ACM Press, May 1992. 1, 4, 12, 13

[Kit97] A Yu Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191, 1997. 20

[KO09] Takeshi Koshiba and Takanori Odaira. Statistically-hiding quantum bit commitment from approximable-preimage-size quantum one-way function. In *Workshop on Quantum Computation, Communication, and Cryptography*, pages 33–46. Springer, 2009. 18

[KO11] Takeshi Koshiba and Takanori Odaira. Non-interactive statistically-hiding quantum bit commitment from any quantum one-way function, 2011. 18

[Kre21] William Kretschmer. Quantum pseudorandomness and classical complexity. In Min-Hsiu Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2021, July 5-8, 2021, Virtual Conference*, volume 197 of *LIPIcs*, pages 2:1–2:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. 4, 10, 12, 29

[KSV02] Alexei Y. Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, Providence, RI, 2002. 20, 52

[KSW08] Dennis Kretschmann, Dirk Schlingemann, and Reinhard F. Werner. The information-disturbance tradeoff and the continuity of stinespring's representation. *IEEE Transactions on Information Theory*, 54(4):1708–1717, 2008. 88, 89

[LMS22a] Russell W. F. Lai, Giulio Malavolta, and Nicholas Spooner. Quantum rewinding for many-round protocols. TCC '22, 2022. 14

[LMS22b] Alex Lombardi, Fermi Ma, and Nicholas Spooner. Post-quantum zero knowledge, revisited (or: how to do quantum rewinding undetectably). In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, FOCS '22, 2022. 14, 53, 67

[Mah18] Urmila Mahadev. Classical verification of quantum computations. In Mikkel Thorup, editor, *59th FOCS*, pages 259–267. IEEE Computer Society Press, October 2018. 17

[MW05] Chris Marriott and John Watrous. Quantum arthur-merlin games. *CoRR*, abs/cs/0506068, 2005. 57

[MY22] Tomoyuki Morimae and Takashi Yamakawa. Quantum commitments and signatures without one-way functions. 2022. https://ia.cr/2021/1691. 4, 10

[Sti55] William F. Stinespring. Positive functions on *-algebras. 1955. 88

[Unr12] Dominique Unruh. Quantum proofs of knowledge. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 135–152. Springer, Heidelberg, April 2012. 5, 14, 15

[Unr16] Dominique Unruh. Computationally binding quantum commitments. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT 2016, Part II*, volume 9666 of *LNCS*, pages 497–527. Springer, Heidelberg, May 2016. 1, 3, 5, 6, 9, 11, 14, 15, 23, 24, 25, 33, 36, 84

[Val08] Paul Valiant. Incrementally verifiable computation or proofs of knowledge imply time/space efficiency. In Ran Canetti, editor, *TCC 2008*, volume 4948 of *LNCS*, pages 1–18. Springer, Heidelberg, March 2008. 56

[vdG97] Jeroen van de Graaf. *Towards a formal definition of security for quantum protocols*. PhD thesis, University of Montreal, 1997. 14

[Wat09] John Watrous. Zero-knowledge against quantum attacks. *SIAM J. Comput.*, 39(1):25–58, 2009. 14, 71, 74

[Yan22] Jun Yan. General properties of quantum bit commitment. ASIACRYPT '22, 2022. 2, 4, 18, 24, 25, 81

[YWLQ15] Jun Yan, Jian Weng, Dongdai Lin, and Yujuan Quan. Quantum bit commitment with application in quantum zero-knowledge proof. In *International Symposium on Algorithms and Computation*, pages 555–565. Springer, 2015. 18

[Zha12] Mark Zhandry. How to construct quantum random functions. In *53rd FOCS*, pages 679–687. IEEE Computer Society Press, October 2012. 11, 29

[Zha20] Mark Zhandry. Schrödinger's pirate: How to trace a quantum decoder. In Rafael Pass and Krzysztof Pietrzak, editors, *TCC 2020, Part III*, volume 12552 of *LNCS*, pages 61–91. Springer, Heidelberg, November 2020. 57

A Removing interaction from any QSC

In this section, we define the syntax for an interactive QSC and define two security definitions for interactive QSCs called "honest swap binding" and "honest hiding." We show that interactive QSCs satisfying these security definitions can be compiled to *non-interactive* QSCs that are swap-binding (Definition 4.2) and hiding (Definition 4.3) in Theorem A.5. This transformation is similar to the round-collapse compiler of [Yan22].

We begin by defining syntax for interactive QSCs.

Definition A.1 (Interactive QSC syntax)**.** *An interactive quantum state commitment* QSC *between a quantum sender S and a quantum receiver R consists of an interactive commit phase $\langle S, R\rangle_{\mathrm{Com}}$ and unitary* Recover*.*

- *(Commitment phase) The sender commits to a state on message register* M *by engaging in a quantum interactive protocol $\langle S, R\rangle_{\mathsf{Com}}$ with the receiver.*
- *(Opening phase) To open the commitment, the sender engages in a quantum interactive protocol $\langle S, R\rangle_{\mathsf{Open}}$ with the receiver. At the end of the protocol, the receiver applies a binary outcome projective measurement to decide whether to accept or reject the opening. If it accepts, the receiver recovers the originally committed message on some register* M*.*

For completeness, we require that the map induced by performing an honest commitment followed by an honest opening is the identity map on the message space $\mathcal{M}$.

Definition A.2 (Completeness)**.** *For any quantum state commitment* QSC*, let $\Phi_{\mathsf{QSC}} : \mathbf{S}(\mathcal{M}) \to \mathbf{S}(\mathcal{M})$ be the CPTP map induced by (1) performing an honest commitment to* M *(2) performing an honest opening, and (3) replacing* M *with $|0\rangle$ if the receiver rejects. We say that a QSC scheme* QSC *is complete if*

$$\left\|\Phi_{\mathsf{QSC}} - \mathbb{I}_{\mathbf{S}(\mathcal{M})}\right\|_{\diamond} = 0.$$

In particular, this definition implies that if the sender honestly commits to a state $|\psi\rangle$ and later opens honestly, the receiver will accept with probability 1 and recovers $|\psi\rangle$. It is easy to check that the syntax for a non-interactive QSC (Definition 4.1) is a special case of an interactive QSC.

Next, we define honest hiding and honest swap binding.

Honest hiding and honest swap binding. We define a security experiment $\mathtt{HonHideExpt}_{\mathsf{QSC},A,b}(\lambda)$ parameterized by a quantum commitment scheme QSC, an interactive adversary A, a challenge bit $b \in \{0,1\}$, and a security parameter λ.

Definition A.3 (Honest hiding)**.** *For a quantum commitment scheme* QSC*, an interactive adversary A, a challenge bit $b \in \{0,1\}$, and a security parameter λ, define a security experiment $\mathtt{HonHideExpt}_{\mathsf{QSC},A,b}(\lambda)$ as follows.*

$\mathtt{HonHideExpt}_{\mathsf{QSC},A,b}(\lambda)$*:*

1. *The adversary A prepares a message* M *and sends it to the challenger.*

2. *The challenger locally executes an entire commit phase between an honest sender and an honest receiver, where the honest sender commits to* M *if* $b = 0$*, or* $|0\rangle$ *if* $b = 1$*. The challenger sends the internal state* C *of the honest receiver to* A*.*

3. *The output of the experiment is* $b' \leftarrow A$*.*

QSC *is computationally (resp. statistically) honest-hiding if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries* A*,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{HonHideExpt}_{\mathsf{QSC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Definition A.4 (Honest swap binding)**.** *For a quantum commitment scheme* QSC*, an interactive adversary* A*, a challenge bit* $b \in \{0,1\}$*, and a security parameter* λ*, define the security experiment* $\mathtt{HonSwapBindExpt}_{\mathsf{QSC},A,b}(\lambda)$ *as follows.*

$\mathtt{HonSwapBindExpt}_{\mathsf{QSC},A,b}(\lambda)$*:*

1. *The adversary* A *sends the challenger a register* M*.*

2. *The challenger locally executes an entire commit phase between an honest sender and an honest receiver, where the honest sender commits to* M *if* $b = 0$*, or* $|0\rangle$ *if* $b = 1$*. The challenger sends the internal state* D *of the honest sender to* A*.*

3. *The output of the experiment is* $b' \leftarrow A$*.*

QSC *is computationally (resp. statistically) honest swap binding if there exists a negligible function* $\mu(\lambda)$ *such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries* A*,*

$$\Pr_{b\leftarrow\{0,1\}}[\mathtt{HonSwapBindExpt}_{\mathsf{QSC},A,b}(\lambda) = b] \leq \frac{1}{2} + \mu(\lambda).$$

Compiling an interactive QSC into a non-interactive QSC. Consider the following non-interactive commitment scheme, compiled from a (potentially) interactive quantum commitment scheme QSC:

- Suppose the commit phase of QSC is k rounds, where in round $i \in [k]$, the receiver applies a unitary R_i, sends a register X_i to the sender, who applies a unitary S_i and sends back a register Y_i.
- Set

$$\mathsf{Com} = S_k R_k S_{k-1} R_{k-1} \cdots S_1 R_1. \tag{15}$$

Theorem A.5. *Let* QSC *be computationally (resp. statistical) honest swap binding and statistically (resp. computationally) honest hiding. Then the non-interactive commitment scheme* Com *defined by Equation* (15) *is computationally (resp. statistical) swap binding and statistically (resp. computationally) hiding.*

Proof. The honest-hiding experiment for the interactive scheme is equivalent to the hiding experiment for the non-interactive scheme from the adversary's point of view because the challenger locally executes the entire commit phase.

Similarly, by Property 4.5, the swap binding experiment for non-interactive commitments can be equivalently phrased by the adversary sending a message register M, and receiving an honest decommitment to it in register D. Thus, the honest-binding experiment for the interactive QSC is equivalent to the swap binding experiment for the non-interactive QSC. □

B Equivalences between binding definitions

B.1 Pauli binding

In this subsection, we present an alternative binding definition for QSCs we call *Pauli binding*, which applies to non-interactive and interactive QSCs (Definition A.1). Pauli binding is a generalization of sum-binding (see [Unr16]) in which the adversary picks an arbitrary Pauli operator P, receives a random bit b from the challenger specifying either the $+1$ or -1 eigenspace of P, and wins if it can open to a message (that is measured to be) in that eigenspace. We will prove that this notion is equivalent to swap binding for non-interactive QSCs.

Definition B.1 (Pauli binding)**.** *For a quantum commitment scheme* QSC*, an interactive adversary A, and a security parameter λ, define a security experiment* $\texttt{PauliBindExpt}_{\mathsf{QSC},A}(\lambda)$ *as follows.*

$\texttt{PauliBindExpt}_{\mathsf{QSC},A}(\lambda)$*:*

1. *The adversary A (acting as a malicious sender) engages in the commit phase of* QSC *with the challenger (acting as the receiver). Then, A sends the description of an n-qubit Pauli operator P to the challenger.*
2. *The challenger samples a random $b \leftarrow \{-1, 1\}$ and sends b to the adversary.*
3. *A engages in the opening phase of* QSC *with the challenger. If the challenger rejects the opening as invalid, then the experiment aborts and outputs* fail*.*
4. *Next, the challenger measures $\{P^+, P^-\}$ on the opened message register* M*. It outputs* win *if the outcome is b, and otherwise outputs* fail*.*

QSC *is computationally (resp. statistically) Pauli binding if there exists a negligible function $\mu(\lambda)$ such that for all polynomial-time (resp. unbounded-time) quantum interactive adversaries A,*

$$\Pr\left[\texttt{PauliBindExpt}_{\mathsf{QSC},A}(\lambda) \to \text{win}\right] \leq \frac{1}{2} + \mu(\lambda).$$

Remark B.2. *Strictly speaking, Pauli binding assumes the messages have "qubit structure," i.e., the dimension of the Hilbert space must be a power of 2. It is possible to generalize this definition to qudits where d is* $\mathsf{poly}(\lambda)$*, but it is not immediately obvious how to extend this definition to message spaces of arbitrary dimension. Thus, we only show an equivalence with swap-binding when the message space has qubit structure. We note that the swap binding definition is agnostic to the dimension of the message space.*

Theorem B.3. *A non-interactive quantum state commitment scheme* Com *is swap binding if and only if it is Pauli binding.*

Proof. We first show that if Com is Pauli binding, then it is also swap binding. Suppose that an adversary A distinguishes between the $b = 0$ and $b = 1$ worlds of the swap binding game with advantage ε. Consider the following three hybrids:

- H_0: In this hybrid, an adversary A sends a commitment-decommitment pair (C, D) to the challenger. The challenger applies $\mathsf{Com}^\dagger$, measures $\{|0\rangle\langle 0|, \mathbb{I} - |0\rangle\langle 0|\}$ on register W (and aborts if the measurement rejects), applies Com, and sends D to the adversary.
- H_1: This hybrid is the same as H_0, except before applying Com, the challenger applies a uniformly random n-qubit Pauli P to M (thus making it maximally mixed).
- H_2: This hybrid is the same as H_0, except before applying Com, the challenger replaces the contents of M with $|0\rangle$, i.e., initializes ancillary register M' to $|0\rangle$ and applies $\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']$.

Hybrids H_0 and H_2 are exactly the $b = 0$ and $b = 1$ worlds of the swap binding experiment for Com, so A distinguishes either H_0 and H_1 or H_1 and H_2 with advantage $\varepsilon/2$. Let's assume that A distinguishes H_0 and H_1 with $\varepsilon/2$ advantage as the proof for H_1 and H_2 follows almost identically.

We purify the actions of the adversary A, so that right after sending (C, D) to the challenger, A and the challenger jointly hold a pure quantum state $|\psi\rangle_{\mathsf{CDR}}$, where R is held by A. Then after receiving D back from the challenger, A measures a projector $\widetilde{\Pi}$ on (D, R) to determine its output for the swap binding game. Recall that for any operator O, we define the notation $\widehat{O} := (\mathsf{Com})O(\mathsf{Com}^\dagger)$. For any Pauli P acting on register M define the quantities

$$\mathsf{DistAdv}_P := \left| \left\| \widetilde{\Pi} \widehat{P} \widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\psi\rangle \right\|^2 - \left\| \widetilde{\Pi} \widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\psi\rangle \right\|^2 \right|$$

and

$$\mathsf{MapAdv}_P := \left\| \widehat{P^-} \widehat{|0\rangle\langle 0|}_{\mathsf{W}} (\mathbb{I} - 2\widetilde{\Pi}) \widehat{P^+} \widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\psi\rangle \right\|^2.$$

Then

$$\begin{aligned} \mathbb{E}_{P \leftarrow \{I,X,Y,Z\}^{\otimes n}} [\mathsf{MapAdv}_P] &\geq \mathbb{E}_P \left[(\mathsf{DistAdv}_P)^2 \right] \\ &\geq \left(\mathbb{E}_P \left[\mathsf{DistAdv}_P \right] \right)^2 \\ &= (\varepsilon/2)^2, \end{aligned}$$

where the first inequality is by Lemma 6.6, Item (ii), the second inequality is Jensen's inequality, and the last equality is because A distinguishes H_0 and H_1 with advantage $\varepsilon/2$.

Then consider the Pauli binding adversary A' that does the following:

1. Receive commitment-decommitment registers (C, D) from A.
2. Sample a random n-qubit Pauli P and measure $\{\widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}, \mathbb{I} - \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\}$ on (C, D). Then:
   - If the measurement accepts, send register C and Pauli P to the challenger. Receive $b \in \{-1, 1\}$ from the challenger. If $b = -1$, then apply $(\mathbb{I} - 2\widetilde{\Pi})$, which can be done by coherently applying A, applying Z to its output, and applying A in reverse. If $b = 1$, do nothing to the state. Then, send register D to the challenger.
   - If the measurement rejects, prepare another commitment to a $+1$ eigenstate of P, i.e., initialize M' to a $+1$ eigenstate of P and W' to $|0\rangle$, apply Com to $(\mathsf{M}', \mathsf{W}')$ (resulting in $(\mathsf{C}', \mathsf{D}')$), and send C' and P to the challenger. Receive b from the challenger. Then, send decommitment register D' to the challenger.

We analyze the success probability of A' in the Pauli binding game. Let win indicate the event in which A' wins the Pauli binding game and let accept indicate the event that the measurement $\{\widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}, \mathbb{I} - \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\}$ accepts. Then

$$\begin{aligned}\Pr[\mathrm{win}] &= \Pr[\mathrm{win} \wedge \mathrm{accept} \wedge (b=1)] + \Pr[\mathrm{win} \wedge \mathrm{accept} \wedge (b=-1)] + \Pr[\mathrm{win} \wedge \bar{\mathrm{accept}}] \\ &\geq \frac{1}{2}\Pr[\mathrm{accept}] + \frac{1}{2}\mathbb{E}[\mathsf{MapAdv}_P] + \frac{1}{2}(1 - \Pr[\mathrm{accept}]) \\ &\geq \frac{1}{2} + \frac{\varepsilon^2}{8}.\end{aligned}$$

Thus A' wins the Pauli binding game with noticeable advantage.

We now show that swap binding implies Pauli binding. Let A be an adversary that wins the Pauli binding game for Com with probability $1/2 + \varepsilon$ using Pauli operator P. Let us purify its actions so that it sends the C register of a pure quantum state $|\psi\rangle_{\mathsf{CDR}}$ to the challenger, and applies unitary operations U_- and U_+ upon receiving $b = -1$ and $b = 1$, respectively. Then letting $U := U_- U_+^\dagger$ and $|\phi\rangle := U_+ |\psi\rangle$, adversary A winning the Pauli binding game with probability $1/2+\varepsilon$ is equivalent to:

$$1 + 2\varepsilon \leq \left\| \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U_+ |\psi\rangle \right\|^2 + \left\| \widehat{P^-}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U_- |\psi\rangle \right\|^2 \tag{16}$$

$$= \left\| \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\phi\rangle \right\|^2 + \left\| \widehat{P^-}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U |\phi\rangle \right\|^2. \tag{17}$$

Let us define (sub-normalized) quantum states

- $|\alpha\rangle := \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\phi\rangle$.
- $|a\rangle := \widehat{P^-}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\phi\rangle$.
- $|b\rangle := \widehat{P^-}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U (\mathbb{I} - \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}) |\phi\rangle$.
- $|c\rangle := (\mathbb{I} - \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}) |\phi\rangle$

Then

$$\begin{aligned}1 + 2\varepsilon &\leq \||\alpha\rangle\|^2 + \||a\rangle + |b\rangle\|^2 \\ &\leq \||\alpha\rangle\|^2 + (\||a\rangle\| + \||b\rangle\|)^2 \\ &= \||\alpha\rangle\|^2 + \||a\rangle\|^2 + 2\||a\rangle\|\||b\rangle\| + \||b\rangle\|^2 \\ &\leq \||\alpha\rangle\|^2 + \||a\rangle\|^2 + 2\||a\rangle\|\||b\rangle\| + \||c\rangle\|^2 \\ &\leq 1 + \||a\rangle\|^2 + 2\||a\rangle\|\||b\rangle\| \\ &\leq 1 + \||a\rangle\|^2 + 2\||a\rangle\| \\ &\leq 1 + 3\||a\rangle\|,\end{aligned}$$

where the first inequality is a restatement of Equation (17), the second inequality is triangle inequality, the third inequality is because $\||b\rangle\| \leq \||c\rangle\|$, the fourth inequality is because $|\alpha\rangle \perp |c\rangle$

and $\||\alpha\rangle + |c\rangle\|^2 \leq 1$, the fifth inequality is because $\||b\rangle\| \leq 1$, and the last inequality is because $\||a\rangle\| \leq 1$.

Thus

$$(2\varepsilon/3)^2 \leq \||a\rangle\|^2 = \left\|\widehat{P^-}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} U \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}} |\phi\rangle\right\|^2.$$

By Lemma 6.6, Item (i), if we define $\Pi := \mathsf{ctl}_{\mathsf{B}}\text{-}U\,|+\rangle\langle+|_{\mathsf{B}}\,\mathsf{ctl}_{\mathsf{B}}\text{-}U^\dagger$ and $|\gamma\rangle := \mathsf{ctl}_{\mathsf{B}}\text{-}U(|+\rangle_{\mathsf{B}} \otimes (\widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}})\,|\phi\rangle)$, then

$$\left|\left\|\Pi\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\,|\gamma\rangle\right\|^2 - \left\|\Pi\widehat{P}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\,|\gamma\rangle\right\|^2\right| \geq 2\varepsilon^2/9. \tag{18}$$

In words, the measurement $\{\Pi, \mathbb{I} - \Pi\}$ distinguishes whether $\widehat{P}$ is applied or not to $\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\,|\gamma\rangle$.

Then consider the following three hybrids:

- H_0: In this hybrid, the adversary sends commitment and decommitment registers (C, D) to the challenger. The challenger measures $\{\widehat{|0\rangle\langle 0|}_{\mathsf{W}}, \mathbb{I} - \widehat{|0\rangle\langle 0|}_{\mathsf{W}}\}$. If the measurement rejects, then abort. Otherwise, send D back to the adversary.
- H_1: This hybrid is the same as H_0, except before sending D back to the adversary, the challenger initializes a new register E (of the same dimension as M) to $|0\rangle$, then applies $\widehat{\mathsf{SWAP}}[\mathsf{M}, \mathsf{E}]$.
- H_2: This is the same as H_0, except before sending D back to the adversary, the challenger applies $\widehat{P}$ to (C, D).

Equation (18) implies that the adversary A' that does the following achieves a distinguishing advantage $2\varepsilon^2/9$ between hybrids H_0 and H_2:

1. Prepare the state $\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\,|\gamma\rangle$ from $|\psi\rangle_{\mathsf{CDR}}$ by applying U_+, then measuring $\{\widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}, \mathbb{I} - \widehat{P^+}\widehat{|0\rangle\langle 0|}_{\mathsf{W}}\}$, initializing a qubit register B to $|+\rangle$, applying $\mathsf{ctl}_{\mathsf{B}}\text{-}U$, and measuring $\{\widehat{|0\rangle\langle 0|}_{\mathsf{W}}, \mathbb{I} - \widehat{|0\rangle\langle 0|}_{\mathsf{W}}\}$.
   - If both measurements accept, then the resulting state is (normalized) $|\gamma\rangle$. Send C and D registers to the challenger, and receive D back from the challenger. Then output the result of measuring $\{\Pi, \mathbb{I} - \Pi\}$.
   - If either of the measurements reject, just output a random bit at the end of the hybrids. That is, prepare any valid commitment and decommitment registers $(\mathsf{C}', \mathsf{D}')$ by initializing new registers $(\mathsf{M}', \mathsf{W}')$ to $|0\rangle$ and applying Com to the registers. Send $(\mathsf{C}', \mathsf{D}')$ to the challenger, receive D', then output a random bit.

Thus A' achieves a distinguishing advantage $\varepsilon^2/9$ between hybrids H_0 and H_1 or H_1 and H_2. If A' distinguishes H_0 and H_1, then it wins the swap binding game with the same advantage by definition of the swap binding game. If A' distinguishes H_1 and H_2, then a modified A' that applies $\widehat{P}$ to the commitment-decommitment registers right before sending them to the challenger will win the swap binding game with the same advantage.

□

Remark B.4. *We remark that a nearly identical proof to the above can be used to show that* collapse-binding *for non-interactive QBCs is equivalent to a restricted version of Pauli binding where the adversary is only allowed to send Pauli operators from* $\{\mathbb{I}, Z\}^{\otimes n}$*. A very similar definitional equivalence (for collapse-binding) was also proved in a concurrent and independent work of [DS22]. [DS22] show that collapse binding is equivalent to "chosen-bit binding," which corresponds to a version of our Pauli binding definition where the operators are from the set* $\{Z_1, \ldots, Z_n\}$*.*

B.2 A statistical extraction-based definition

In the classical setting, a defining feature of statistically binding commitments is that the commitment string information-theoretically fixes the committed message. Equivalently, there is an inefficient procedure to *extract* the committed message from the commitment (up to negligible error). In the quantum setting, [AQY22] proposed defining statistical binding for quantum commitments to *classical messages* (QBCs) in a similar fashion: a QBC is statistically binding if, given the commitment register C, it is possible to extract the committed message bit (in an appropriate sense).

In this subsection, we present (a) a statistical extraction definition for commitments to *quantum messages* and (b) a proof that this definition is equivalent to statistical swap binding. We emphasize that this alternative characterization of swap binding only applies in the *statistical* binding setting. To state the statistical extraction definition, recall the following notation, where Com is the commitment unitary for a non-interactive quantum state commitment:

- $\widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']} := (\mathsf{Com})\mathsf{SWAP}[\mathsf{M}, \mathsf{M}'](\mathsf{Com}^\dagger)$ and
- $\Pi := (\mathsf{Com})(\mathbb{I}_{\mathsf{M}} \otimes |0\rangle\langle 0|_{\mathsf{W}})(\mathsf{Com}^\dagger)$.

Definition B.5 (Statistical extractability)**.** *A non-interactive quantum commitment scheme* QSC *with commitment unitary* Com *is* statistically extractable *if there exists a CPTP map* $\mathsf{Ext} : (\mathsf{C}, \mathsf{M}') \to (\mathsf{C}, \mathsf{M}')$ *such that, for all states* $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{C} \otimes \mathcal{D} \otimes \mathcal{H})$ *and all unitaries* U *acting only on* (D, H)*,*

$$\left\| (\Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})} \circ \mathsf{Ext}_{(\mathsf{C},\mathsf{M}')})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) - (\widehat{\mathsf{SWAP}[\mathsf{M}, \mathsf{M}']} \circ \Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) \right\|_1 \leq \mathsf{negl}(\lambda).$$

The fact that statistical extractability is equivalent to statistical binding follows from the "information-disturbance tradeoff" [KSW08, Theorem 3], restated in Theorem B.6. To state the theorem, we require the notion of a *complementary channel*. For any CPTP map $\Phi : \mathbf{S}(\mathcal{X}) \to \mathbf{S}(\mathcal{Y})$, Stinespring's dilation thoerem [Sti55] implies that there exists a unitary U that maps registers (X, W) to (Y, Z) (where W and Z are defined implicitly by U) such that $\Phi(\boldsymbol{\rho}) = \mathrm{Tr}_{\mathsf{Z}}(U(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{W}}))$[35]. We say that Φ^c is a *complementary channel* to Φ if there exists a unitary U such that $\Phi^c(\rho) = \mathrm{Tr}_{\mathsf{Y}}(U(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{W}}))$ and $\Phi(\rho) = \mathrm{Tr}_{\mathsf{Z}}(U(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{W}}))$.

[35] Here we use the notation that $U(\boldsymbol{\rho}) = U\boldsymbol{\rho}U^\dagger$ for a unitary U because the proof of the main theorem in this subsection is stated in terms of quantum channels.

Theorem B.6 (Information-disturbance tradeoff [KSW08, Theorem 3])**.** *Let $\mathcal{M} \otimes \mathcal{A} \otimes \mathcal{B}$ be a finite-dimensional Hilbert space. Suppose that $T : \mathcal{M} \to \mathcal{A}$ is a quantum channel with complementary channel $T^c : \mathcal{M} \to \mathcal{B}$. Then there exists a decoding channel $\mathsf{Dec} : \mathcal{A} \to \mathcal{M}$ such that*

$$\|\mathsf{Dec} \circ T - \mathbb{I}_M\|_\diamond \leq \min_{\boldsymbol{\sigma} \in \mathbf{D}(\mathcal{B})} 2\|T^c - S_{\boldsymbol{\sigma}}\|_\diamond$$

where $S_{\boldsymbol{\sigma}}$ is the completely depolarizing channel that always outputs $\boldsymbol{\sigma} \in \mathbf{D}(\mathcal{B})$.

Theorem B.7. *A non-interactive quantum commitment scheme QSC is statistically swap-binding if and only if it is statistically extractable.*

Proof. Suppose that QSC is statistically extractable. Then invoking Definition B.5 with $U = \mathbb{I}$ and initial state $\Pi\boldsymbol{\rho}\Pi$,

$$\left\|(\Pi_{(\mathsf{C},\mathsf{D})} \circ \mathsf{Ext}_{(\mathsf{C},\mathsf{M}')} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) - (\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})\right\|_1 \leq \mathsf{negl}(\lambda).$$

Since $\Pi_{(\mathsf{C},\mathsf{D})}$ is a projection, it follows that

$$\left\|(\mathsf{Ext}_{(\mathsf{C},\mathsf{M}')} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) - (\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})\right\|_1 \leq \mathsf{negl}(\lambda).$$

In particular, in the $b = 1$ world of the swap-binding security experiment (Definition 4.2) the adversary receives the state

$$\begin{aligned}
&\mathrm{Tr}_{(\mathsf{C},\mathsf{M}')}[(\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})] \\
&\approx_s \mathrm{Tr}_{(\mathsf{C},\mathsf{M}')}[(\mathsf{Ext}_{(\mathsf{C},\mathsf{M}')} \circ \Pi_{(\mathsf{C},\mathsf{D})})(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})] \\
&= \mathrm{Tr}_{(\mathsf{C},\mathsf{M}')}[\Pi_{(\mathsf{C},\mathsf{D})}(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})].
\end{aligned}$$

Since $\mathrm{Tr}_{(\mathsf{C},\mathsf{M}')}[\Pi_{(\mathsf{C},\mathsf{D})}(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{M}'})]$ is the state the adversary receives in the $b = 0$ world, swap-binding follows.

For the other direction, suppose that QSC is statistically swap-binding. Let $T : \mathsf{M} \to \mathsf{C}$ be defined by

$$T(\boldsymbol{\rho}) = \mathrm{Tr}_{\mathsf{D}}(\mathsf{Com}(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{W}})).$$

The complementary channel is

$$T^c(\boldsymbol{\rho}) = \mathrm{Tr}_{\mathsf{C}}(\mathsf{Com}(\boldsymbol{\rho} \otimes |0\rangle\langle 0|_{\mathsf{W}})),$$

and since QSC is statistically swap-binding we have

$$T^c(\boldsymbol{\rho}) \approx_s T^c(|0\rangle\langle 0|_{\mathsf{M}})$$

for all $\boldsymbol{\rho} \in \mathbf{D}(\mathcal{M})$. Applying Theorem B.6 with $\boldsymbol{\sigma} = T^c(|0\rangle\langle 0|_{\mathsf{M}})$ yields a decoding channel $\mathsf{Dec} : \mathsf{C} \to \mathsf{M}$ such that

$$\|\mathsf{Dec} \circ T - \mathbb{I}_M\|_\diamond \leq \mathsf{negl}(\lambda).$$

Let $\widetilde{\mathsf{Dec}}$ be a Stinespring dilation of Dec, i.e., $\widetilde{\mathsf{Dec}} : (\mathsf{C},\mathsf{E}) \to (\mathsf{M},\mathsf{F})$ is a unitary such that $\mathrm{Tr}_{\mathsf{F}}(\widetilde{\mathsf{Dec}}(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{\mathsf{E}})) = \mathsf{Dec}(\boldsymbol{\tau})$. We can now define our extractor $\mathsf{Ext} : (\mathsf{C},\mathsf{M}') \to (\mathsf{C},\mathsf{M}')$ for $\boldsymbol{\tau} \in \mathbf{D}(\mathcal{C})$ as

$$\mathsf{Ext}(\boldsymbol{\tau}) = \mathrm{Tr}_{\mathsf{E}}[(\widetilde{\mathsf{Dec}}^\dagger \circ \mathsf{SWAP}[\mathsf{M},\mathsf{M}'] \circ \widetilde{\mathsf{Dec}})(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{(\mathsf{E},\mathsf{M}')})].$$

Since $\|\mathsf{Ext} \circ \Pi - \Pi \circ \mathsf{Ext}\|_\diamond \leq \mathsf{negl}(\lambda)$, it follows that

$$
\begin{aligned}
&(\Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})} \circ \mathsf{Ext}_{(\mathsf{C},\mathsf{M}')})(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) \\
&\approx_s (\mathsf{Ext}_{(\mathsf{C},\mathsf{M}')} \circ \Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})})(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{\mathsf{M}'}) \\
&= \mathrm{Tr}_{\mathsf{E}}[(\widetilde{\mathsf{Dec}}^\dagger \circ \mathsf{SWAP}[\mathsf{M},\mathsf{M}'] \circ \widetilde{\mathsf{Dec}} \circ \Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})})(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{\mathsf{M}'} \otimes |0\rangle\langle 0|_{\mathsf{E}})] \\
&\approx_s (\widehat{\mathsf{SWAP}[\mathsf{M},\mathsf{M}']} \circ \Pi_{(\mathsf{C},\mathsf{D})} \circ U_{(\mathsf{D},\mathsf{H})})(\boldsymbol{\tau} \otimes |0\rangle\langle 0|_{\mathsf{M}'}).
\end{aligned}
$$

□

C Quantum encryption with short keys from PRUs

Pseudo-random unitaries (PRUs) are families of efficient unitaries that are indistinguishable from Haar random unitaries under black-box access [JLS18]. In Theorem C.1, we show that the Hilbert space of a PRU can be expanded at the cost of restricting the security to a single use — effectively converting a pseudo-superpolynomial-design to a pseudo-1-design on a larger space.

The construction is based on the Schur transform $\mathbf{U}_{\mathsf{Sch}}$ [Har05], which maps between the standard representation of $(\mathbb{C}^p)^{\otimes \ell}$ and the "Schur-Weyl basis," in which both ℓ-fold products of unitaries $U^{\otimes \ell}$ and permutations act naturally. In particular, the Schur transform maps between the symmetric subspace on ℓ qu*p*its and an $\binom{\ell+p-1}{\ell}$-dimensional Hilbert space on $\left\lceil \log_2 \binom{\ell+p-1}{\ell} \right\rceil$ qubits [Har05].

Given a PRU family $\{U_k\}_{k\in\{0,1\}^{d(\lambda)}}$ acting on p dimensions, we build a one-time quantum encryption scheme $\{\mathsf{Expand}(U_k,\ell)\}_{k\in\{0,1\}^{d(\lambda)}}$ with $\binom{\ell+p-1}{\ell}$-dimensional messages $|\psi\rangle$ as follows:

1. Initialize the appropriate registers to select the symmetric subspace in the Schur basis and the remaining registers to $|\psi\rangle$. In the notation of [Har05], this is $|\Lambda = 0, p_\Lambda = 0, q_\Lambda = \psi\rangle$.
2. Apply $\mathbf{U}_{\mathsf{Sch}}(U_k)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger$ to $|\Lambda = 0, p_\Lambda = 0, q_\Lambda = \psi\rangle$.
3. Return the q_Λ register as the output of $\mathsf{Expand}(U_k,\ell)$.

Theorem C.1 (PRU Expansion). *If $\{U_k\}_{k\in\{0,1\}^{d(\lambda)}}$ is a PRU family on p dimensions, then $\{\mathsf{Expand}(U_k,\ell)\}_{k\in\{0,1\}^{d(\lambda)}}$ is a secure one-time quantum encryption scheme with messages of dimension $\binom{\ell+p-1}{\ell} \geq \left(1+\frac{p-1}{\ell}\right)^\ell$ for any $\ell = \mathsf{poly}(\lambda)$.*

Proof. Since the symmetric subspace is invariant under $(U_k)^{\otimes \ell}$, the auxiliary registers Λ, p_Λ are unaffected by $\mathbf{U}_{\mathsf{Sch}}(U_k)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger$. Letting ρ be the state $|\Lambda = 0, p_\Lambda = 0, q_\Lambda = \psi\rangle$,

$$\begin{aligned}\mathbf{U}_{\mathsf{Sch}}(U_k)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger \rho \mathbf{U}_{\mathsf{Sch}}(U_k^\dagger)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger &\approx_c \mathbf{U}_{\mathsf{Sch}}\left(\int_{U\leftarrow\mu(\mathbb{C}^p)} (U)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger \rho \mathbf{U}_{\mathsf{Sch}}(U^\dagger)^{\otimes \ell}\, dU\right)\mathbf{U}_{\mathsf{Sch}}^\dagger \\ &= \binom{\ell+p-1}{\ell}^{-1} |\Lambda = 0, p_\Lambda = 0\rangle\langle\Lambda = 0, p_\Lambda = 0|\,.\end{aligned}$$

The first line follows from the security of the PRU family $\{U_k\}_{k\in\{0,1\}^{d(\lambda)}}$. The equality follows from Schur's lemma, since the symmetric subspace on $(\mathbb{C}^p)^{\otimes \ell}$ is an irreducible representation of the unitary group under the action $U \mapsto U^{\otimes \ell}$ [Har13]. Schur's lemma tells us that the integral is maximally mixed over the symmetric subspace, and in the Schur-Weyl basis the maximally mixed state over the symmetric subspace is the state where q_Λ is random and $\Lambda = 0, p_\Lambda = 0$. □

Similarly, we note that a many-time secure PRS family $\{|\phi_k\rangle\}_{k\in\{0,1\}^{d(\lambda)}}$ on p dimensions may be converted to a one-time secure PRS family on $\binom{\ell+p-1}{\ell}$ dimensions via $\mathrm{Tr}_{\Lambda,p_\Lambda}(\mathbf{U}_{\mathsf{Sch}}(|\phi_k\rangle\langle\phi_k|)^{\otimes \ell}\mathbf{U}_{\mathsf{Sch}}^\dagger)$.

Theorem C.2 (PRS Expansion). *If $\{|\phi_k\rangle\}_{k\in\{0,1\}^{d(\lambda)}}$ is an ℓ-time secure PRS family on p dimensions, then $\{\mathsf{Expand}(|\phi_k\rangle,\ell)\}_{k\in\{0,1\}^{d(\lambda)}}$ is a one-time secure PRS family on $\binom{\ell+p-1}{\ell} \geq \left(1+\frac{p-1}{\ell}\right)^\ell$ dimensions.*

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il