Unclonable Commitments and Proofs

Vipul Goyal[1,2], Giulio Malavolta[3], and Justin Raizes[1]

[1]Carnegie Mellon University
[2]NTT Research
[3]Bocconi University

Abstract

Non-malleable cryptography, proposed by Dolev, Dwork, and Naor (SICOMP '00), has numerous applications in protocol composition. In the context of proofs, it guarantees that an adversary who receives a proof cannot maul it into another valid proof. However, non-malleable cryptography (particularly in the non-interactive setting) suffers from an important limitation: An attacker can always copy the proof and resubmit it to another verifier (or even multiple verifiers).

In this work, we *prevent even the possibility of copying the proof as it is*, by relying on quantum information. We call the resulting primitive *unclonable proofs*, making progress on a question posed by Aaronson. We also consider the related notion of *unclonable commitments*. We introduce formal definitions of these primitives that model security in various settings of interest. We also provide a near tight characterization of the conditions under which these primitives are possible, including a rough equivalence between unclonable proofs and public-key quantum money.

1 Introduction

Non-malleable cryptography studies cryptographic primitives that (provably) cannot be mauled in a "controlled" way. For instance, non-malleable zero-knowledge [DDN00] considers protocols between a prover and a verifier, where the prover wants to convince the verifier of the validity of a certain NP-statement x. At the same time, non-malleability guarantees that the verifier cannot act as a *man-in-the-middle*, i.e., it cannot use the interaction with the prover to convince a *different* verifier (unless they were able to create their proof from scratch). However, traditional non-malleable cryptographic protocols (particularly the non-interactive ones) suffer from an important limitation: A man-in-the-middle can always *copy* the messages of the prover and forward them to the new verifier. This attack is clearly unavoidable, and it is typically ruled out by weakening the non-malleability guarantee (for instance, by only requiring non-malleability across different NP-statements, or by associating each participant with a *tag* and defining non-malleability to hold only across different tags). In this work, we ask the following question:

Can we construct proofs that (provably) cannot be cloned?

While this question is clearly hopeless classically, quantum information behaves in a fundamentally different manner. The no-cloning theorem [WZ82] offers the tantalizing possiblity that one may be able to overcome the above limitation leveraging quantum information. In fact, recent years have seen a surge of success in constructing cryptographic primitives with *unclonability* guarantees, such as quantum money [Wie83, Aar09, AC12, FGH$^+$12, Zha21], unclonable encryption [BL20, AK21, AKL$^+$22], signature tokens [BS17,

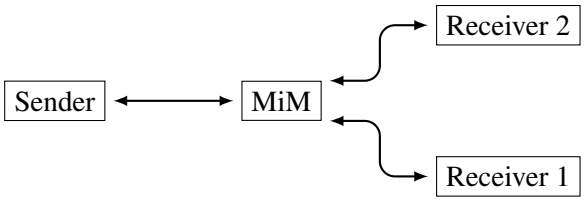

Figure 1: The Commitment/Proof Cloning Experiment

CLLZ21], revocable decryption keys [KN22, AKN$^+$23, APV23, BGG$^+$23], unclonable pseudorandom functions [CLLZ21], to mention a few. However, to the best of our knowledge, the notion of unclonable proofs (and even the more basic notion of unclonable commitments) has not been formally studied. In fact, this same question was posed by Aaronson as open problem.[1] The purpose of this work is to make progress on this front.

Unclonable Commitments and Proofs. In this work we initiate the formal study of unclonable commitments and unclonable proofs.[2] More specifically, we consider a man-in-the-middle (MiM) interacting in a *left* session with an honest sender, and in two *right* sessions with two honest receivers. This is illustrated in Figure 1. In these settings we require that:

- (Commitments) The message committed to in one of the right sessions is independent of the message committed to in the left session. In particular, consider two (non-communicating) distinguishers which each receive one of the messages that were committed to in the right sessions, then attempt to guess whether the left session is a commitment to m_0 or to m_1. We require that they cannot *simultaneously* succeed with negligible advantage over random guessing, i.e. with probability $1/2 + \mathrm{negl}(\lambda)$.
- (Proofs: Simulation-Extractability) If the right receivers both accept statements $\widetilde{x}_1$ and $\widetilde{x}_2$, then MiM must know a witness for at least one of the two statements. This is formalized using a simulator-extractor which simultaneously simulates MiM's view *without* the witness for the left session and, if both right sessions accept, extracts a witness for either $\widetilde{x}_1$ and $\widetilde{x}_2$.
- (Proofs: Simulation-Soundness) Even if MiM receives a simulated proof for any statement x (potentially not in the language $\mathcal{L}$) in the left session, it cannot convince two honest verifiers to accept statements $\widetilde{x}_1 \notin \mathcal{L}$ and $\widetilde{x}_2 \notin \mathcal{L}$ in the right sessions.

We explicitly mention here that the notion of unclonability is in principle meaningful for proofs that do not satisfy any zero-knowledge guarantee. However, there must be a sense in which the witness is hidden, as otherwise the verifier could violate the above properties by simply recovering the witness and computing any number of honest proofs.

Application: Unclonable Credentials. As an interesting application, we mention that unclonable proofs/arguments can be used to convert any classical credential into an *unclonable credential*. Starting from any publicly-verifiable credential (such as a digital signature), one can simply commit to such a credential and produce an unclonable proof of the validity of such a credential. The resulting state can be then publicly verified and it can (provably) not be cloned. This can be used, for instance, to building

[1] `https://scottaaronson.blog/?p=2903`

[2] We refer to both computationally and statistically sound protocols as proofs in this work.

uncloneable smartcards or tokens to grant access to a restricted area, or to log into a particular device. Contrary to existing unclonable smartcards, unclonable credentials provide provably secure guarantees against an adversary that can even tamper with the hardware.

1.1 Our Results

As the main contributions of this work, we lay the definitional groundwork for unclonable commitments and proofs, and we study under which computational assumptions we can build this primitives. We consider two settings: Same-protocol unclonability and strong unclonability.

Same-Protocol Unclonability. In these settings, we require that the above unclonability guarantees hold when the same verification protocol is performed in all three sessions. For commitments, we obtain three different constructions with tradeoffs between their assumptions and properties.

Theorem 1.1 (Unclonable Commitments - Informal)**.** *We obtain the following constructions of unclonable commitments.*

- *Assuming public-key quantum money and non-interactive non-malleable commitments, there exist non-interactive unclonable commitments.*
- *Assuming post-quantum non-malleable commitments, there exist interactive post-quantum unclonable commitments.*
- *Assuming non-interactive post-quantum non-malleable commitments for one left session and four right sessions, there exist non-interactive unclonable commitments.*

The assumption of non-malleable commitments is somewhat minimal, since unclonability is a strengthening of non-malleability. On the other hand, we show that non-interactive unclonable proofs require (seemingly) stronger assumptions.

Theorem 1.2 (Non-Interactive Unclonable Proofs - Informal)**.** *Assuming non-interactive zero knowledge and either one-way functions or public key encryption, non-interactive unclonable proofs exist* if and only if *public key quantum money exists.*

If we are willing to allow interaction, then we can construct an unclonable proof with only classical computation.

Theorem 1.3 (Post-Quantum Unclonable Proofs - Informal)**.** *Assuming post-quantum one-way functions, there exist (interactive) post-quantum unclonable proofs.*

Additionally, we study the setting where the adversary receives k commitments/proofs in the left sessions and attempts to clone them to produce $k + r$ commitments/proofs in the right sessions, which we call *many-many unclonability*. Our constructions also obtain many-many unclonability, if instantiated with a commitment/proof where non-malleability holds even for many left and right sessions (concurrent non-malleability).

Theorem 1.4 (Many-Many Unclonability - Informal)**.** *Assuming concurrent non-malleable commitments (respectively, proofs), there exist many-many unclonable commitments (respectively, proofs).*

Strong Unclonability. In these settings, we require that the above unclonability guarantees hold for commitments/proofs with respect to *any verification procedure*. This is a strong notion of unclonability that effectively captures the intuition that the information vanishes after the verification is performed. Unfortunately, we show that unclonable proofs do not exist in these settings, assuming the existence of commitments. In these settings we show that unclonable proofs are impossible to achieve.

Theorem 1.5 (Impossibility of Strongly Unclonable Proofs - Informal)**.** *Assuming classical commitments, there do not exist strongly unclonable proofs, even for interactive protocols.*

The proof of the impossibility result makes use of a secure multiparty computation protocol to construct an explicit attack. To the best of our knowledge, this is the first time that techniques from secure multiparty computation have been used to analyze unclonability.

We complement the impossibility result by mentioning several possible modifications to the model that may allow sidestepping the impossibility. We explore one of them to construct strongly unclonable commitments in the quantum random oracle model (QROM), and leave the rest as open problems.

Theorem 1.6 (Strongly Unclonable Commitments - Informal)**.** *In the QROM, there unconditionally exist strongly unclonable commitments against right protocols which are statistically binding.*

2 Technical Overview

Since unclonability is closely related to non-malleability, we begin by recalling this notion and discussing its relation with unclonability. In a non-malleability experiment, an adversarial MiM interacts with a sender in a left session and with a receiver in the right session. This is similar to the scenario illustrated in Figure 1, except that there is only one right receiver. Since it is impossible to prevent the MiM from directly forwarding the messages between the two sessions, non-malleability is commonly defined with respect to identities, or tags, to force differences between the two sessions. If the MiM uses a different tag in the right session than in the left session, then we may require one of the following guarantees:

- (Commitments [DDN00, PR05b]) The message committed to in the right session is independent of the message committed to in the left session. In particular, the joint view of the MiM and the value committed to in the right session cannot be used to distinguish whether the left session is a commitment to m_0 or to m_1.
- (Proofs: Simulation-Extraction [PR05b]) The MiM must know a witness for the statement $\widetilde{x}$ in the right session. This is formalized using a simulator-extractor which simultaneously simulates MiM's view *without* the witness for the left session and, if the right session accepts, extracts a witness for $\widetilde{x}$.
- (Proofs: Simulation-Soundness [Sah99]) Even if the MiM receives a simulated proof for any statement x (potentially not in the language $\mathcal{L}$) in the left session, it cannot convince an honest verifier to accept a false statement $\widetilde{x} \notin \mathcal{L}$ in the right session.

We emphasize that these guarantees *only* hold if the left and right sessions use different tags.

2.1 Definitions: Unclonable Commitments

Unclonable commitments remove the different-tag restriction by considering *two* right sessions, instead of one. Although the MiM can still forward one of the right sessions, we require that the messages $\widetilde{m_1}$ and $\widetilde{m_2}$

committed to in the right sessions cannot simultaneously be dependent on the message m committed to in the left session. We formalize this by requiring that no adversary wins the following game with probability greater than $1/2 + \mathrm{negl}$.

1. The challenger commits to either m_0 or m_1 in the left session, and the MiM commits in the right sessions. Afterwards, the MiM splits its internal state into two registers MiM_1 and MiM_2.
2. Two non-communicating distinguishers D_1 and D_2 each receive a residual register MiM_1 (respectively MiM_2) and the value $\widetilde{m}_1$ (respectively, $\widetilde{m}_2$) of the commitment in right session one (respectively, two). The adversary wins if both distinguishers correctly guess which message the challenger committed to.

To ensure that $\widetilde{m_1}$ and $\widetilde{m_2}$ are well-defined, we define unclonability for statistically binding commitments. This allows us to define the value of a commitment using a computationally-*un*bounded extractor that acts on the receiver's state. The case of computationally bounded unclonable commitments is less straightforward. Even in the less demanding setting of classical non-malleability, definitions with respect to computational binding are nuanced since it is difficult to define a meaningful "value committed in the right session". Nevertheless, our paradigm can in principle be applied to any notion of non-malleability for computational binding, such as non-malleability with respect to opening [DIO98] or non-malleability with respect to replacement [Goy11], simply by using the corresponding notion of a commitment's value. For example, non-malleability with respect to opening defines a commitment's value to be the receiver's output at the end of the opening phase.

Relation to Unclonable Encryption. The definition of unclonable commitments has many similarities to unclonable encryption [BL20], but has a few key differences. In unclonable encryption, the adversary splits a ciphertext $\mathsf{Enc}(k, m_b)$ into two registers $\mathcal{B}$ and $\mathcal{C}$. Then, adversaries B and C are given $\mathcal{B}$ and $\mathcal{C}$, respectively, along with the secret key k, and try to guess the bit b simultaneously without communicating. Philosophically, unclonable encryption captures the scenario where an eavesdropping adversary wants to collect information about a ciphertext in the hopes of later learning the key.

In contrast, unclonable commitments philosophically captures a scenario where an eavesdropping adversary not only wishes to leak information about the committed message, but to actually make use of the leaked information before it is revealed. Because the adversary is attempting a more difficult task, we can hope to construct unclonable commitments from weaker assumptions. Indeed, indistinguishable-unclonable encryption is currently only known in the random oracle model [AKL+22]. In contrast, we show that when the same protocol is used in all three sessions, it is possible to construct non-interactive unclonable commitments from any non-interactive non-malleable commitment (see Section 5.2). In this sense, unclonable commitments are weaker than unclonable encryption.

However, the two notions are technically incomparable. We note that unclonable encryption may in principle be unconditionally possible, whereas unclonable commitments necessarily require computational assumptions. Even with computational assumptions, it is not clear how to transform an unclonable encryption scheme into an unclonable commitment. Indeed, existing techniques for adding computational assumptions to unclonable encryption (e.g. to create unclonable public-key encryption) make crucial use of a "fake-key" property that allows the encrypter to open a ciphertext to any message it wants [AK21]. This technique is fundamentally incompatible with the binding requirement of a commitment scheme.

Relation to [BC23]. Recently, Broadbent and Culf proposed a different definition of unclonable commitments. In their definition, the committer Alice interacts with a receiver Bob, while an eavesdropper Eve sees

(and can interfere with) their messages. After the commitment phase, Alice and Bob run a check phase. During the opening phase, which happens after the check phase, Bob and Eve cannot directly communicate. Their definition guarantees that if Alice accepts in the check phase, then at the end of the opening phase Eve has no information about Alice's message. Here, Eve plays a similar role to the one that MiM plays in our definition. There are several core differences between our definitions.

- The first difference is that their definition takes place in the interactive setting due to the extra check phase, whereas our definition also makes sense for non-interactive settings.
- The second is that their definition is for statistically hiding (and hence computationally binding) commitments, whereas ours is for computationally hiding, statistically binding commitments.
- A third difference is that their scheme provides no guarantees if Alice rejects during the check phase, or if Eve can communicate with Bob during the opening phase. In this case, it is possible that Eve and Bob both hold information about Alice's message. In contrast, our guarantee always holds (even if the left session aborts during one of the two phases).

To exemplify the difference between the two approaches, consider the following toy scenario: Alice has found a proof that P=NP and wants to commit to the proof to the Clay institute (Bob), to later claim the 1M$ prize. Note that our constructions allow Alice to make a non-interactive commitment. Eve intercepts the commitment, and she tries to clone it and submit as her own commitment to Bob (who is distributing the prize). Our definition prevents this attack since Eve's commitment is not allowed to depend on that of Alice. On the other hand, Broadbent and Culf's definition does not necessarily defend against this attack: This is because their definition only applies if Eve and Bob do not communicate at all during the opening phase. In this example, Eve can intercept Alice's opening, create her own opening based on that, and submit to Bob.

2.2 Definitions: Unclonable Proofs

We now turn our attention to defining unclonable proofs. Existing notions of unclonability, such as unclonable encryption [BL20] or even copy-protection [Aar09] attempt to prevent an adversary in possession of a cloned state from learning new information. *However, these notions do not make sense in the context of proofs, since it is always possible to generate a fresh proof using the witness.* To fill this gap, we present two incomparable notions of unclonable proofs which are inspired by the non-malleable proof literature.

Extraction-Unclonability. The first notion aims to realize the intuition that the *only* way for the MiM to generate two accepting proofs is to generate one of them using the witness. In other words, the MiM must "know" at least one witness. In the classical setting of non-malleability, this is formalized by the notion of simulation-extraction.

In the setting of unclonability, we consider a cloning experiment where the MiM receives a proof of a statement x in the left session and sends proofs in two right sessions (see Figure 1). This experiment outputs MiM's final view, as well as the statements $\widetilde{x}_1$ and $\widetilde{x}_2$ in the two right sessions and whether the verifiers accept in those sessions. Extraction-unclonability requires the existence of a simulator $\mathcal{S}(x)$ which receives as input the statement x to be proved in the left session (but not a witness) and outputs two things. First, it must output a view τ which is indistinguishable from adversary's view at the end of the real cloning experiment. Second, it must output two potential witnesses $\widetilde{w}_1$ and $\widetilde{w}_2$. If the verifiers accept in both right sessions of τ, then at least one of $\widetilde{w}_1$ and $\widetilde{w}_2$ must be a valid witness for $\widetilde{x}_1$ and $\widetilde{x}_2$, respectively. We note that since witness relations are efficiently checkable, it is possible to efficiently determine which of the two right sessions extraction succeeded in, and which session was potentially forwarded.

Soundness-Unclonability. A more nuanced adversary might attempt to directly break soundness of both right sessions. Although extraction-unclonability prevents this by extracting at one least valid witness if the left session proves a true statement $x \in \mathcal{L}$, it does not provide any guarantees if the left session proves a *false* statement $x \notin \mathcal{L}$. This nuance is reminiscent of the difference between simulation-extraction and simulation-soundness in the setting of classical non-malleability, which [JP14] shows are incomparable.

In fact, this scenario arises naturally in cryptographic proofs. Consider a "paired commitment" where the left prover commits twice to the same value v and proves in zero knowledge that the two commitments are consistent. A simple strategy to prove security of this paired commitment is to use a hybrid argument where first we simulate the zero-knowledge proof, then switch the commitments one at a time to a new value v'. Observe that in the middle hybrid, the left commitments are to *different* values v and v', yet the zero-knowledge proof is still being simulated. If the receiver (the MiM) were to engage as a prover in a second proof during this hybrid, then the results could be catastrophic, since there would be no guarantee of soundness. Thus, we could not simultaneously rely on the hiding of the paired commitment and the soundness of a second proof.

To address this issue, we extend simulation-soundness to the setting of unclonability. Consider the same proof cloning experiment as for extraction-unclonability. Simulation-soundness requires the existence of a simulator $\mathcal{S}(x)$ which receives as input the statement x to be proved in the left session (but not a witness), then outputs a view τ which is indistinguishable from the adversary's view at the end of the proof cloning experiment. Furthermore, if $\widetilde{x}_1 \notin \mathcal{L}$ and $\widetilde{x}_2 \notin \mathcal{L}$ in τ, then at most one of the receivers accept in the right sessions of τ.

Application: Unclonable Credentials. We expand on the previously mentioned application of unclonable credentials and briefly discuss why both extraction-unclonability and soundness-unclonability suffice for it. Recall that to construct an unclonable credential (say for an employee's badge), one can classically commit to a signature of the badge number b, then produce an unclonable proof that the commitment contains a valid signature of b.

If the proof were to satisfy extraction-unclonability, then any copier would be able to produce a valid signature without receiving one in the clear. This immediately violates the unforgeability of the signature scheme.

On the other hand, if the proof were to satisfy soundness-unclonability, then we could consider a series of hybrid experiments where b becomes an invalid credential in general. First, the unclonable proof is simulated. Second, the badge holds a commitment to $\perp$, instead of a commitment to a signature on b. Finally, the signing key is "punctured", so that no valid signature on b exists [BSW16]. In the final hybrid, any proof of a signature for b must be breaking soundness, so soundness-unclonability guarantees that the adversary cannot create two accepting badges for b. Since this experiment is indistinguishable from the real world, it also cannot create an acceptable second badge for b in the real world.

2.3 Same-Protocol Unclonability

In same-protocol unclonability, we consider the case where the right sessions use the same commitment/proof protocol as the left session. For example, this scenario could be useful when many users are interacting with the same central entity, who uses a standardized cryptographic suite.

Unclonable Tag-Generation. A non-malleable commitment has similar security to an unclonable one, except that its security guarantee only holds when the right and left sessions use different tags. Thus, if there

was a way to guarantee that one of the two right sessions uses a different tag than the left session, we would be able to rely on non-malleability in that session.

This suggests a natural primitive which we call *unclonable tag-generation*. In a tag-generation protocol, a sender and a receiver interact to agree on a string tag. Unclonability is defined using a game where an adversary MiM participates in three simultaneous sessions: a left session where it acts as the receiver, and two right sessions where it acts as the sender. Unclonability guarantees that at least one of the two right sessions outputs a different tag than the left session, unless all three sessions output a failure symbol $\perp$.

Armed with this primitive, there is a simple construction of unclonable commitments or unclonable proofs. First, agree on tag using an unclonable tag-generation protocol. Then, execute a non-malleable commitment using tag. If both the tag-generation protocol and the non-malleable commitment are non-interactive, then the two messages can be sent in parallel. This yields a non-interactive unclonable commitment. Unclonable proofs can be constructed similarly from non-malleable proofs.

Constructing Unclonable Tag-Generation. A natural idea for unclonable tag-generation is to use the strong unclonability properties of public key quantum money (PKQM). This yields a simple, non-interactive protocol. The sender sends a serial number, banknote pair $(s, |\$_s\rangle)$ to the receiver. The tag is s. The verifier checks that the banknote matches s before outputting s. Since no MiM can create two copies of $|\$_s\rangle$, it can only use s as a tag in a single right session. Unfortunately, public key quantum money is currently only known from strong or poorly-understood assumptions such as indistinguishability obfuscation [AC12, FGH$^+$12, Zha21].

It is immediate to see that in the interactive setting, a *classical* solution is possible. The key observation is that two different receivers may act differently, as long as they send at least one message. First, the receiver sends a uniformly random message r. The sender samples a signature key pair $(\mathsf{sk}, \mathsf{vk})$, then signs r. It sends vk and the signature back to the receiver, who verifies the signature. The tag is the verification key vk. Notice that in the unclonability experiment, the left sender will only sign a single r. However, the two right sessions will use different r's with high probability. Thus, in order to break unclonability, the adversary MiM must forge one of the two signatures. By combining this idea with a post-quantum non-malleable commitment/proof (e.g. [ABG$^+$21, BLS22, LPY23]), it is even possible to construct a post-quantum unclonable commitment/proof. The downside of this approach is that it is interactive, which precludes constructing non-interactive unclonable primitives.

As a result of these ideas, we obtain two constructions of both unclonable commitments and unclonable proofs. One construction is non-interactive, but requires public key quantum money. The other is post-quantum, but requires two rounds of communication.

Non-Interactive Unclonable Commitments without PKQM. In the case of commitments, it is possible to generate the tags in a more flexible way. Since a commitment has both a commitment phase and an opening phase, the tag can be generated in the commitment phase, then confirmed in the opening phase. Unless the adversary can confirm the same tag in both opening phases, it is unable to clone a commitment. This delayed confirmation allows us to use techniques from private-key quantum money, which exists unconditionally [Wie83].

This idea yields our third construction of unclonable commitments, which is non-interactive and can be based only on post-quantum non-interactive non-malleable commitments. As mentioned previously, this is an almost minimal assumption, since unclonability is a strengthening of non-malleability. To generate the tag, the committer samples a random Wiesner state $|x\rangle_\theta$, then commits to (x, θ) using a post-quantum non-interactive non-malleable commitment scheme with tag 0. Call this commitment $\mathsf{com}_{\mathsf{tag}}$. Then, to commit

to a message m, the committer computes a non-interactive non-malleable commitment to m, using $\mathsf{com}_{\mathsf{tag}}$ as the tag. To open the commitment, the committer opens both non-malleable commitments and the verifier checks that $\mathsf{com}_{\mathsf{tag}}$ matches the Wiesner state.

Since the tag $\mathsf{com}_{\mathsf{tag}}$ is statistically binding, it information-theoretically determines a Wiesner state. Thus, in order to use the same tag in all three sessions, the adversarial MiM must clone a Wiesner state *before* its description is revealed. Otherwise, one of the right sessions will contain an invalid commitment. On the other hand, if the adversary uses a different tag in one of the right sessions, then non-malleability guarantees that the value committed inside that session is independent of the value committed in the left session.

However, this intuitive proof fails against a more subtle attack. An adversary may attempt to selectively generate invalid commitments according to the left message m. To do this, it generates an independent Wiesner state $|\widetilde{x}\rangle_{\widetilde{\theta}}$. Then, it mauls the commitment to m so that its right tag commitment $\widetilde{\mathsf{com}}_{\mathsf{tag}}$ matches the Wiesner state if and only if m is a pre-determined message. Crucially, the non-malleability of $\mathsf{com}_{\mathsf{tag}}$ prevents this type of attack. Since $\widetilde{\mathsf{com}}_{\mathsf{tag}}$ must use a different tag (0) from the commitment to m, the value committed inside it must be independent of m. Combining this with the intuitive proof from before shows that in at least one session, the adversary must either make an invalid commitment or commit to a value which is independent of m, and that which case happens must also be independent of m. Therefore this session is independent of m in general.

Unclonable NIZKs Imply PKQM. The reader may observe that the two ideas for unclonable commitments from weaker assumptions both made use of multiple messages to confirm the tag. In our third construction of commitments, we make use of the opening phase to delay the confirmation, giving non-interactive unclonable commitments. However, some primitives, such as proofs, only consist of a single phase. Given this, it is natural to wonder whether non-interactive unclonability for such primitives inherently require public key quantum money. We show that in the case of unclonable non-interactive zero knowledge (NIZKs), the answer is yes; unclonable NIZKs do imply public key quantum money, assuming one-way functions. When combining this with the previous construction of unclonable NIZKs from public key quantum money, we arrive at a loose equivalence between the two.

To construct public key quantum money, we use unclonable NIZKs and commitments. To mint a note, the bank first samples a common reference string (CRS) for the NIZK. Since the bank's only goal is to prevent cloning of banknotes, it may be trusted to do this honestly. Then, it generates a commitment com_0 to 0, which will act as the serial number. Finally, it proves using the unclonable NIZK that com_0 is a commitment to 0. Intuitively, the unclonability of the NIZK guarantees that any adversary which produces two banknotes must be able to find a witness for com_0 being a commitment to 0, just given com_0. However, this would violate the security of the commitment! It is also possible to show the security of the scheme using the related notion of NIZK soundness-unclonability, which guarantees that the adversary cannot convince both right verifiers to accept a statement $\widetilde{x} \notin \mathcal{L}$, even if it receives a simulated proof for a statement x (potentially not in $\mathcal{L}$). We defer the details to the technical section.

2.4 Strong Unclonability

It is also interesting to consider the setting where the right protocols may be *different* from the left protocol. For example, this scenario could be useful when a user wishes to place sealed bids in auctions from multiple auction houses, who might each prefer their own cryptographic libraries.

Negative Results. It is simple to see that strongly unclonable NIZKs are not possible; an adversary can simply ask the verifier to return the NIZK after they are done, then reuse the returned NIZK. We further show that even *interactive* strongly unclonable proofs are not possible.

Intuitively, an adversary can attack a candidate proof by engaging in a single session of a secure multi-party computation (MPC) protocol (for quantum functionalities) with the two verifiers. The MPC protocol emulates a single verifier for the proof. This ideal verifier will always accept the (honest) proof from the left session, so both right verifiers will also accept. We emphasize that the verifiers do not have a direct channel between them, so they must send messages to each other via the adversary. Since the adversary controls the channel between the two verifiers, the MPC protocol must be secure against a dishonest majority in order to guarantee that the right protocols are sound.

One can also imagine a similar attack for commitments. However, we only define unclonable commitments with respect to statistical binding. In order for the right protocols, which are defined using the MPC, to be statistically binding, the MPC must guarantee the correct output even when an *unbounded* adversary controls the majority of the parties.

Positive Results: Commitments. One option for avoiding the impossibility is to use a random oracle (the QROM model). In a dishonest majority setting, an MPC cannot hope to query the QROM, preventing the attack. We show how to construct a strongly unclonable commitment scheme in the QROM model, where unclonability holds if the two right protocols are statistically binding.

To commit to a message m, encrypt m using an unclonable encryption scheme, then commit to the encryption key k using the random oracle. Unclonable encryption can be constructed in the QROM [AKL+22]. As long as k is statistically bound, there is only one possible decryption of the unclonable ciphertext, giving statistical binding. Crucially, random oracles allow commitments which are *both* statistically binding and statistically hiding. The former property allows statistical binding. The latter property ensures k can be removed from the left commitment without affecting the values committed to in the right commitments. Since the right commitments must be statistically binding, they each uniquely determine a value $\widetilde{m}_r$. However, if $\widetilde{m}_1$ and $\widetilde{m}_2$ simultaneously depended on m, then this would violate the security of the underlying unclonable encryption scheme.

2.5 Concurrent Work

In an independent and concurrent work, Jawale and Khurana [JK23] also introduced a notion of unclonable NIZKs and showed that it is roughly equivalent to public-key quantum money. Their notion of *unclonable extractability* is very similar to our notion of *extraction-unclonability*, which supports it as a natural definition of unclonability. They also apply unclonable NIZKs to construct unclonable signatures of knowledge, which we do not consider. On the other hand, we additionally define and construct unclonable commitments, extend our definitions to the interactive case, and investigate the possibility of achieving security against verifiers who do not follow the honest verification procedure (strong unclonability).

2.6 Paper Organization

We define and construct unclonable tag-generation in Section 4. We define unclonable commitments in Section 5.1 and construct them in Section 5.2. We define unclonable proofs in Section 6.1 and construct them in Section 6.2. We prove their equivalence to public key quantum money in Section 6.3. We present our negative results for strong unclonability in Section 7 and the positive results in Section 8.

3 Preliminaries

λ is the security parameter. We denote the projector $\Pi_x^\theta = H^\theta |x\rangle \langle x| H^\theta$ onto the string $x \in \{0,1\}^n$ in the basis $\theta \in \{0,1\}^n$, where $H^\theta = \bigotimes_i H^{\theta_i}$. A non-uniform QPT adversary is a polynomial-time unitary along with a quantum advice state ρ. In general, ρ may be entangled with an external register (e.g. possessed by another adversary).

3.1 Commitments

A quantum bit commitment scheme consists of a commitment phase $\mathsf{Com} = \langle \mathsf{Com}_C, \mathsf{Com}_R \rangle$ and an opening phase $\mathsf{Open} = \langle \mathsf{Open}_C, \mathsf{Open}_R \rangle$.

- Com describes an interaction between a committer C and a receiver R, where C commits to a message $m \in \{0,1\}^n$. R outputs a commitment state and C outputs a decommitment state.
- Open describes an interaction where C reveals m, then attempts to convince R that it committed to m. R outputs m if it is convinced, and $\perp$ otherwise.

It must satisfy correctness, computational or statistical hiding, and computational or statistical binding. For correctness, if C honestly commits to $m \in \{0,1\}^n$, then an honest R will output m in the Open phase. We describe only computational hiding and statistical binding here.

Definition 3.1 (Computational Hiding)**.** *A commitment scheme* $(\mathsf{Com}, \mathsf{Open})$ *is computationally hiding if for all messages $m_0, m_1 \in \{0,1\}^n$, all QPT adversarial receivers R^*, and all QPT distinguishers D,*

$$\left| \begin{array}{r} \Pr[D(\tau_{R^*}) = 1 : (\tau_C, \tau_{R^*}) \leftarrow \mathsf{Com}_{C,R^*}(m_0)] \\ -\Pr[D(\tau_{R^*}) = 1 : (\tau_C, \tau_{R^*}) \leftarrow \mathsf{Com}_{C,R^*}(m_0)] \end{array} \right| \leq \mathrm{negl}(\lambda)$$

Definition 3.2. *A quantum bit commitment scheme* $(\mathsf{Com}, \mathsf{Open})$ *satisfies* statistical binding *if for every adversarial committer C^*, there exists a (possibly inefficient) extractor algorithm* Ext *acting on the receiver's register such that the following properties hold.*

1. ***Correct Value.*** *Let $\mathsf{Valid}_\epsilon(\rho_R, m)$ be the predicate which is 1 if and only if there exists an purification $\rho_{C,R}$ of ρ_R such that $\mathsf{Open}(\rho_{C,R})$ outputs m with probability ϵ. Given a state $\rho_{C,R}$ on registers C and R, let ρ_R be the mixed state resulting from tracing out register C. Then*

$$\Pr\left[\mathsf{Valid}_1(\widetilde{\rho}_R, m) = 1 : \begin{array}{c} (\widetilde{\rho}_R, m) \leftarrow \mathsf{Ext}(\rho_R) \\ \rho_{C^*,R} \leftarrow \langle \mathsf{Com}_{C^*}, \mathsf{Com}_R \rangle \end{array}\right] = \mathrm{negl}(\lambda)$$

   In other words, if the extractor outputs a message m, then there must be an opening to m.

2. ***Unique Value.*** *For all $\epsilon = 1/\operatorname{poly}(\lambda)$,*

$$\Pr\left[\exists m' \neq \perp \textit{ s.t. } \mathsf{Valid}_\epsilon(\widetilde{\rho}_R, m') = 1 : \begin{array}{c} (\widetilde{\rho}_R, m) \leftarrow \mathsf{Ext}(\rho_R) \\ \rho_{C^*,R} \leftarrow \langle \mathsf{Com}_{C^*}, \mathsf{Com}_R \rangle \end{array}\right] = \mathrm{negl}(\lambda)$$

   In other words, the transcript can only *be opened to the extracted message or $\perp$.*

3. ***Indistinguishable Transcript.*** *The following two distributions have negligible trace distance.*

$$\left\{(\tau_{C^*}, m): \begin{array}{c} \rho_{C^*,R} \leftarrow \langle \mathsf{Com}_{C^*}, \mathsf{Com}_R \rangle \\ (\tau_{C^*,R}, m) \leftarrow \langle \mathsf{Open}_{C^*}, \mathsf{Com}_R \rangle (\rho_{C^*,R}) \end{array}\right\}$$

and

$$\left\{(\tau_{C^*}, m): \begin{array}{c} \rho_{C^*,R} \leftarrow \langle \mathsf{Com}_{C^*}, \mathsf{Com}_R \rangle \\ (\widetilde{\rho}_{C^*,R}, \widetilde{m}) \leftarrow (I^{C^*} \otimes \mathsf{Ext}^R)(\rho_{C^*,R}) \\ (\tau_{C^*,R}, m) \leftarrow \langle \mathsf{Open}_{C^*}, \mathsf{Com}_R \rangle (\widetilde{\rho}_{C^*,R}) \end{array}\right\}$$

We note that this definition of statistical binding for quantum bit commitments is slightly different from prior definitions. It is most similar to the extractor-based definition of [AQY22], which has shown to be equivalent to several other definitions such as sum-binding [Yan22, DS23]. Indeed, properties 2 and 3 are together equivalent to [AQY22]'s definition. The only difference is property 1, which requires the extractor to recognize invalid commitments. Looking ahead to non-malleability and unclonability, the ability to recognize invalid commitments will allow us to rule out an adversary which receives a commitment to m_b, then generates either a valid commitment or an invalid one, depending on b.

Observe that any classical bit commitment scheme satisfies this definition. Thus, our constructions can be instantiated from classical bit commitments. Additionally, our (quantum) constructions will satisfy this definition, providing further support for the definition. It is an interesting question to show whether this definition is equivalent to pre-existing notions of statistical binding.

Non-malleable Commitments. Concurrent non-malleability with respect to commitment [PR05b, DDN00] is defined using a man-in-the-middle experiment $G_{\mathsf{MiM},k_L,k_R}(\vec{m})$, where $\vec{m}$ is a vector of k_L bitstrings of length n. In it, there are $k_L + k_R$ concurrent sessions: k_L "left" sessions between honest committers C_i and an adversarial MiM using classical identities tag_i, and k_R "right" sessions between MiM and honest receivers R_i using classical identities $\widetilde{\mathsf{tag}}_i$. In left session i, C_i commits to $\vec{m}_i$, and in the right sessions, MiM attempts to commit to a related sequence of values $\widetilde{m}_1, \ldots, \widetilde{m}_{k_R}$. After both commitment phases have completed, the statistical binding extractor is applied to each R_i's state to obtain $(\rho, \widetilde{m}_1, \ldots, \widetilde{m}_{k_R})$, where ρ is the joint state of all parties in the experiment. The experiment sets $v_i = \widetilde{m}_i$ if $\widetilde{\mathsf{tag}}_i \neq \mathsf{tag}$ and otherwise sets $v_i = \perp$. Finally, it outputs $(\widetilde{\tau}_{C^*}, \widetilde{v}_1, \ldots, \widetilde{v}_{k_R})$, where τ_{MiM} is MiM's final state.

Definition 3.3 (Non-malleability with Respect to Commitment)**.** *A statistically binding commitment scheme is* (k_L, k_R)-concurrently non-malleable with respect to commitment *if for every* $\vec{m}_0, \vec{m}_1 \in \{0,1\}^{nk_L}$*, every QPT man-in-the-middle* MiM*, and every QPT distinguisher* D*,*

$$\left|\begin{array}{r} \Pr[D(\tau_{\mathsf{MiM}}, \widetilde{v}_1, \ldots, \widetilde{v}_k) = 1 : (\tau_{\mathsf{MiM}}, \widetilde{v}_1, \ldots, \widetilde{v}_k) \leftarrow G_{\mathsf{MiM},k_L,k_R}(\vec{m}_0)] \\ -\Pr[D(\tau_{\mathsf{MiM}}, \widetilde{v}_1, \ldots, \widetilde{v}_k) = 1 : (\tau_{\mathsf{MiM}}, \widetilde{v}_1, \ldots, \widetilde{v}_k) \leftarrow G_{\mathsf{MiM},k_L,k_R}(\vec{m}_1)] \end{array}\right| \leq \mathrm{negl}(\lambda)$$

We say it is concurrently non-malleable with respect to commitment if this holds for every $k_L, k_R = \mathrm{poly}(\lambda)$.

The reader may note that this definition differs slightly from the classical case. In the classical case, the value inside the transcript is well defined - it is the unique message which the transcript can be opened to. However, in the quantum setting, the "value" of the transcript is not well-defined, since the transcript may be in superposition over multiple values. The statistical binding extractor is able to measure the value of the transcript, allowing us to settle this nuance.

Several of our constructions will require the use of specific identities/tags. The ability to support arbitrary tags is a natural property for non-malleable commitments. For example, [LPY23] supports arbitrary tags.

3.2 Proofs

Let $\mathcal{L}$ be an NP-language and let $\mathcal{R}$ be the corresponding relation. A proof system for $\mathcal{R}$ consists of an interactive protocol between a prover and a verifier $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$, where the prover takes as input a witness w and both parties take as input a statement $x \in \mathcal{L}$ and at the end of the interaction the verifier returns a bit $\{0,1\}$ denoting acceptance or rejection. We require a proof system to satisfy the following properties.

- (Correctness) For all $(w,x) \in \mathcal{R}$, the verifier in $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ always returns 1.
- (Soundness) For all $x \notin \mathcal{L}$ and all (non-uniform) QPT adversaries $\mathcal{A}$ the probability that the verifier accepts in the interaction $\langle \mathcal{A}, \mathsf{Verify} \rangle$ is negligible in the security parameter. For statistical soundness, this must hold for all (unbounded) quantum adversaries $\mathcal{A}$.[3]
- (Zero-Knowledge) There exists a simulator Sim such that for all (non-uniform) QPT adversaries $\mathcal{A}$ with advice ρ, and all $(w,x) \in \mathcal{R}$ it holds that the following distributions

  $$\mathsf{Sim}(x, \rho, \mathcal{A}) \approx \langle \mathsf{Prove}, \mathcal{A}(\rho) \rangle$$

  are computationally indistinguishable.

If a proof system consists of single message from the prover to the verifier, then we call it *non-interactive* zero-knowledge (NIZK). Sometimes we additionally require that soundness holds also when given access to the simulator (on different statements), in which case we say that the protocol satisfies *simulation-soundness*. A related notion is *simulation-extraction*, which considers an experiment where the adversary receives a proof for a statement $x \in \mathcal{L}$ in a left session and sends a proof for some statement $\widetilde{x}$ in a right session. The left and right sessions are associated with identities tag and $\widetilde{\mathsf{tag}}$, respectively. Simulation-extraction requires the existence of a simulator-extractor which outputs a transcript along with a witness $\widetilde{w}$ and the right verifier's acceptance bit. The first output must satisfy zero-knowledge. Additionally, if $\mathsf{tag} \neq \widetilde{\mathsf{tag}}$ and the right verifier accepted, then $(\widetilde{w}, \widetilde{x}) \in \mathcal{R}$ with probability $1 - \mathrm{negl}(\lambda)$. If these properties hold for any polynomial number of left and right sessions, we say the proof is *concurrently* simulation-sound (respectively, concurrently simulation-extractable).

3.3 Quantum Money

We recall the definition of quantum money [Aar09, Zha19]. For convenience, we only define the version of the protocol where there is a single valid banknote (commonly referred to as a *mini-scheme*), which can be lifted to the more general settings of many banknotes in a straightforward way using standard digital signatures. Quantum money schemes consist of two QPT algorithm $(\mathsf{Gen}_{\mathsf{QM}}, \mathsf{Verify}_{\mathsf{QM}})$ defined as follows.

- $\mathsf{Gen}_{\mathsf{QM}}(1^\lambda)$: On input the security parameter 1^λ, the generation algorithm produces a quantum banknote $|\$\rangle$, along with a serial number s.
- $\mathsf{Verify}_{\mathsf{QM}}(|\$\rangle, s)$: On input a banknote and a matching serial number, the verification algorithm returns a bit $\{0,1\}$, denoting acceptance or rejection, along with a residual state.

For correctness, we require that honestly generated banknotes are always accepted, and furthermore that the verification algorithm does not noticeably disturb the quantum state associated to the banknote. We formalize these requirements below.

[3] Other works sometimes require that proofs are statistically sound and refer to computationally sound proofs as "arguments". Our definitions and results apply to both computational and statistical soundness.

- (Correctness) For all $\lambda \in \mathbb{N}$, there exists a negligible function negl such that

$$\Pr\left[\mathsf{Verify}_{\mathsf{QM}}(|\$\rangle, s)\right] \geq 1 - \mathrm{negl}(\lambda)$$

  where $(|\$\rangle, s) \leftarrow \mathsf{Gen}_{\mathsf{QM}}(1^\lambda)$.

- (Non-Disturbance) For all $\lambda \in \mathbb{N}$, there exists a negligible function negl such that

$$\mathbb{E}\left[|\langle\Psi|\$\rangle|^2\right] \geq 1 - \mathrm{negl}(\lambda)$$

  where $(|\$\rangle, s) \leftarrow \mathsf{Gen}_{\mathsf{QM}}(1^\lambda)$ and $|\Psi\rangle$ is the residual state obtained after running $\mathsf{Verify}_{\mathsf{QM}}(|\$\rangle, s)$.

Security. For security, consider the following game between an adversary $\mathcal{A}$ and a challenger.

- The challenger runs $(|\$\rangle, s) \leftarrow \mathsf{Gen}_{\mathsf{QM}}(1^\lambda)$ and sends $(s, |\$\rangle)$ to $\mathcal{A}$.
- A produces a pair of (possibly entangled) states $|\$_0\rangle$ and $|\$_1\rangle$.
- The challenger accepts if

$$\mathsf{Verify}_{\mathsf{QM}}(|\$_0\rangle, s) = \mathsf{Verify}_{\mathsf{QM}}(|\$_1\rangle, s) = 1.$$

We say that a quantum money mini-scheme is secure if for all QPT $\mathcal{A}$ the probability that the challenger accepts in the above game is negligible in the security parameter.

3.4 Monogamy of Entanglement and Unclonable Encryption

Unclonable Games. Both monogamy of entanglement [TFKW13] and security for unclonable encryption [BL20] can be described in terms of a cloning game [AKL23]. A cloning game $(\mathsf{Setup}, \mathsf{Gen}, \mathsf{Chall}, \mathsf{Verify})$ consists of three phases:

- **Generation:** The challenger samples a key $\mathsf{sk} \leftarrow \mathsf{Setup}(1^\lambda)$, then prepares a state $|\psi\rangle \leftarrow \mathsf{Gen}(\mathsf{sk})$.
- **Splitting:** The challenger sends $|\psi\rangle$ to an adversary A. A outputs a bipartitate state ρ^{R_B,R_C} across two registers R_B, R_C.
- **Challenge:** Adversaries B and C receive the contents of register R_B and R_C, respectively. Additionally, the challenger samples two challenges $r_B, r_C \leftarrow \mathsf{Chall}(\mathsf{sk})$ and sends them to the respective adversary. B and C output states τ_B and τ_C, respectively. The experiment outputs accept if and only if both $\mathsf{Verify}(\mathsf{sk}, r_B, \tau_B)$ and $\mathsf{Verify}(\mathsf{sk}, r_C, \tau_C)$ output accept.

Unclonable Encryption. An unclonable encryption scheme $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ has the same syntax as a standard encryption scheme, which we assume familiarity with. In addition to the standard notion of correctness for an encryption scheme, it must also satisfy indistinguishable-unclonable security. Indistinguishable-unclonable security is defined using a cloning game, which is parameterized by two messages $m_0, m_1 \in \{0,1\}^n$. $\mathsf{Setup}(1^\lambda)$ samples a key $\mathsf{sk} = (k, b)$ consisting of an encryption key k and a bit b. $\mathsf{Gen}(\mathsf{sk}, b)$ encrypts a message m_b. $\mathsf{Chall}(\mathsf{sk})$ outputs the key k to both adversaries. $\mathsf{Verify}(\mathsf{sk}, r, b')$ accepts if an only if $b' = b$. An encryption scheme is unclonable if, for all m_0, m_1, no QPT adversary can win the cloning game with probability $> 1/2 + \mathrm{negl}(\lambda)$.

Several works have previously studied unclonable encryption and its variants [BL20, AK21, AKL$^+$22]. Currently, unclonable encryption is known in the quantum random oracle model [AKL$^+$22].

Monogamy of Entanglement. Monogamy of entanglement can be described by a cloning game where $\mathsf{Setup}(1^\lambda)$ samples uniform $x, \theta \leftarrow \{0,1\}^\lambda$, $\mathsf{Gen}(x,\theta)$ outputs a Wiesner state $|x\rangle_\theta$, $\mathsf{Chall}((x,\theta))$ outputs (θ,θ), and $\mathsf{Verify}((x,\theta),\theta,\tau)$ outputs accept if and only if $\tau = x$. [TFKW13] introduced this game and showed that no adversary can win it with probability better than $\left(1/2 + 1/(2\sqrt{2})\right)^\lambda$.

3.5 Quantum Random Oracle Model

In the random oracle model [BR93], all parties have access to a uniformly sampled random function H. Since quantum adversaries can evaluate hash functions in superposition, we model quantum adversaries to have quantum access to random oracles [BDF+11]. Specifically, we assume that all algorithms have access to the unitary implementing the mapping:

$$|x\rangle |y\rangle \mapsto |x\rangle |y \oplus H(x)\rangle$$

where H is a uniformly sampled random function.

We paraphrase a theorem from [BBBV97] which relates the state of a quantum oracle algorithm A^H to the state of $A^{H'}$, where $H(x) = H'(x)$ for all $x \neq x^*$.

Theorem 3.4 ([BBBV97])**.** *Let $|\psi_i\rangle = \sum_y \alpha_{y,i} |\phi_{y,i}\rangle \otimes |y\rangle^Q$ be the state of A^H at time i, where Q is the query register. Let $|\psi'_i\rangle$ be the state of $A^{H'}$ at time i. Then for all $T \in \mathbb{N}$,*

$$\mathsf{TD}(|\psi_T\rangle, |\psi'_T\rangle) \leq \sqrt{T \sum_{i=1}^{T} |\alpha_{x^*,i}|^2}$$

Observe that $|\alpha_{x^*,i}|^2$ is the probability of getting output x^* when measuring A^H's query register Q in the computational basis at time i. Thus, x^* can be extracted with probability

$$\frac{1}{T} \sum_{i=1}^{T} |\alpha_{x^*,i}|^2 \geq \left(\frac{\mathsf{TD}(|\psi_T\rangle, |\psi'_T\rangle)}{T} \right)^2$$

by measuring the query register at a random timestep $i \leftarrow [1, \ldots, T]$. A similar result is given by the one-way-to-hiding lemma [Unr14].

3.6 Watrous's Rewinding Lemma

Lemma 3.5 ([Wat09])**.** *Let Q be a quantum circuit acting on an ancilla register initialized to 0 and an input register. Let $p(\psi)$ be the probability that $Q|0\rangle|\psi\rangle$ outputs success (i.e. 0). Let $|\phi_0(\psi)\rangle$ be the residual state after measuring Q's success. Let $p_0, q \in (0,1)$ and $\varepsilon \in (0, 1/2)$ be real numbers such that for all input states $|\psi\rangle$,*

1. $|p(\psi) - q| < \epsilon$
2. $p_0(1-p_0) \leq q(1-q)$
3. $p_0 \leq p(\psi)$

Then there exists a general quantum circuit R with size

$$O\left(\frac{\log(1/\varepsilon)\mathsf{Size}(Q)}{p_0(1-p_0)}\right)$$

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

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

4 Unclonable Tag-Generation

In order to construct unclonable primitives, we start by constructing a primitive called unclonable tag-generation protocols. Informally, a tag-generation protocol guarantees that no adversary can receive a generated tag in a left session, then force the same tag in two other sessions of the same protocol. This is a useful property to combine with non-malleable primitives, which only guarantee security if the tags in the left and right sessions are different. Together, they give a very natural construction of unclonable primitives: determine the tag using an unclonable tag-generation protocol, then run the corresponding non-malleable primitive.

4.1 Definition

Tag-Generation Protocols. A tag-generation protocol is a (potentially interactive) protocol between a sender and a receiver. At the end of the protocol, the sender and the receiver output the same string tag.

Tag Cloning Game. The tag cloning game $\mathsf{Cl\text{-}TagGen}(1^\lambda)$ uses a tag-generation protocol TagGen. It involves three session: a "left" session between an honest sender S_L and a QPT adversarial man-in-the-middle MiM, and two right sessions between MiM and an honest right receiver, respectively named R_1 and R_2. The sessions may be interleaved arbitrarily. The game is played as follows:

1. **Tag-Generation:** Each session runs an instance of TagGen.
2. **Output:** Let tag be S_L's output, let $\widetilde{\mathsf{tag}}_1$ be R_1's output, and let $\widetilde{\mathsf{tag}}_2$ be R_2's output. The adversary wins (output 1) if $\mathsf{tag} = \widetilde{\mathsf{tag}}_1 = \widetilde{\mathsf{tag}}_2 \neq \bot$. Otherwise the game outputs 0.

Definition 4.1 (Unclonability for Tag-Generation)**.** *A tag-generation scheme Π is* unclonable *if for all QPT* MiM*,*

$$\Pr\Big[\mathsf{Cl\text{-}TagGen}(1^\lambda) = 1\Big] \leq \mathrm{negl}$$

4.2 Constructions

A Classical Interactive Protocol. We observe that even classically, two different receivers may act differently. Thus, we may use interaction to realize a classical unclonable tag-generation protocol.

Lemma 4.2 (Interactive Classical Unclonable Tag-Generation)**.** *Assuming post-quantum one-way functions, there exists a two-round, post-quantum tag-generation protocol which is unclonable.*

1. The receiver samples a uniform bitstring $r \leftarrow \{0,1\}^\lambda$, then sends it to the sender.
2. The sender samples a digital signature key pair $(\mathsf{sk}, \mathsf{vk})$. It computes $\sigma \leftarrow \mathsf{Sign}(\mathsf{sk}, r)$, then sends (vk, σ) to the receiver.
3. **Output:** The sender outputs vk. If $\mathsf{Verify}(\mathsf{vk}, \sigma, r)$ accepts, the receiver outputs vk. Otherwise it outputs $\perp$.

Figure 2: Post-Quantum Unclonable Tag-Generation

Proof. The construction is given in Figure 2. Note that post-quantum digital signatures can be constructed from post-quantum one-way functions. Say that MiM violated unclonability, and consider an execution where $\mathsf{tag} = \widetilde{\mathsf{tag}}_1 = \widetilde{\mathsf{tag}}_2 \neq \perp$. By assumption, such executions occur with noticeable probability. Since no tag is equal to $\perp$, it must be the case that $\mathsf{Verify}(\widetilde{\mathsf{tag}}_1, \widetilde{\sigma}_1, \widetilde{r}_1)$ and $\mathsf{Verify}(\widetilde{\mathsf{tag}}_2, \widetilde{\sigma}_2, \widetilde{r}_2)$ both accepted. Furthermore, $\widetilde{r}_1 \neq \widetilde{r}_2$ with probability $1 - 2^{-\lambda}$ over the randomness of R_1 and R_2. Therefore either $r \neq \widetilde{r}_1$ or $r \neq \widetilde{r}_2$. Without loss of generality, say it is $\widetilde{r}_i$. Thus, given MiM, we could break the unforgeability of digital signatures by querying the unforgeability challenger on r, using the returned signature σ to finish the unclonability game, and outputting $\widetilde{\sigma}_i$. □

Unfortunately, using an interactive tag-generation protocol to build an unclonable protocol results in an interactive protocol. In order to construct non-interactive unclonable primitives, we will need a non-interactive tag-generation protocol.

Unclonable Tag-Generation From Public Key Quantum Money. Public key quantum money scheme has a very strong unclonability property. This gives a very simple and natural unclonable tag-generation protocol.

1. The sender generates a banknote $|\$\rangle$ with serial number s. It sends both to the receiver.
2. **Output:** The sender outputs s. If $\mathsf{Verify}(|\$\rangle, s)$ accepts, the receiver outputs s. Otherwise it outputs $\perp$.

Figure 3: Non-Interactive Unclonable Tag-Generation from Public Key Quantum Money

Lemma 4.3 (Unclonable Tag-Generation from PKQM)**.** *Assuming public key quantum money, there exists a non-interactive unclonable tag-generation protocol.*

Proof. The construction is given in Figure 3. Say that some MiM violated unclonability. Then with noticeable probability, $s = \widetilde{s}_1 = \widetilde{s}_2$ and both $\left|\widetilde{\$}_1\right\rangle$ and $\left|\widetilde{\$}_2\right\rangle$ are valid banknotes for s. This is a direct contradiction of the security of public key quantum money. □

5 Unclonable Commitments

5.1 Definitions

Informally, an unclonable commitment guarantees that no man-in-the-middle can receive a commitment to m, then commit to related messages in two other sessions of the same commitment protocol. This is very similar to the guarantee provided by non-malleable commitments. The key difference is that an unclonable commitment does not require that the committers in the three sessions use distinct identities, or even that they send different messages. As a direct consequence, *unclonability is meaningful even if the adversary attempts to directly forward the left session.*

Commitment Cloning Game. The commitment cloning game $\mathsf{Cl\text{-}Com}(1^\lambda, m_0, m_1)$ uses a statistically binding commitment protocol $(\mathsf{Com}, \mathsf{Open})$. It involves three session: a "left" session between an honest committer C_L and an adversarial man-in-the-middle MiM, and two right sessions between MiM and an honest right receiver, respectively named R_1 and R_2. The sessions may be interleaved arbitrarily. Additionally, it uses two QPT distinguishers D_1 and D_2. The game is played as follows:

1. **Commitment:** In the left session, C_L samples a random bit b, then commits to m_b using Com. Simultaneously in the right sessions, MiM interacts with R_1 and R_2 as the committer in two executions of Com. At the end of the execution, MiM splits its internal state into two registers MiM_1 and MiM_2.
2. **Output:** Let Ext be the statistical binding extractor for $(\mathsf{Com}, \mathsf{Open})$ (see Definition 3.2).Let $\mathcal{R}_1$ and $\mathcal{R}_2$ be the internal registers of R_1 and R_2, respectively. Compute $(\mathcal{R}_1, \widetilde{\mathcal{M}}_1) \leftarrow \mathsf{Ext}(\mathcal{R}_1)$ and $(\mathcal{R}_2, \widetilde{\mathcal{M}}_2) \leftarrow \mathsf{Ext}(\mathcal{R}_2)$, where registers $\widetilde{\mathcal{M}}_1$ and $\widetilde{\mathcal{M}}_2$ contains the extracted messages for right session one and two, respectively. Compute $D_1(\mathsf{MiM}_1, \widetilde{\mathcal{M}}_1)$ and $D_2(\mathsf{MiM}_2, \widetilde{\mathcal{M}}_2)$, then measure their respective output bits b_1 and b_2. The adversary wins (output 1) if $b_1 = b_2 = b$, and otherwise the game outputs 0.

Definition 5.1 (Unclonability for Commitments)**.** *A commitment scheme* $(\mathsf{Com}, \mathsf{Open})$ *is* unclonable *if for all (non-uniform) QPT adversaries* MiM*, all (non-uniform) QPT distinguisher pairs* (D_1, D_2) *and all message pairs* $(m_0, m_1) \in \{0,1\}^\ell$*,*

$$\Pr\Big[\mathsf{Cl\text{-}Com}(1^\lambda, m_0, m_1) = 1\Big] \leq \frac{1}{2} + \mathrm{negl}(\lambda)$$

It is also interesting to consider a stronger notion where the right sessions may use different (statistically binding) commitment schemes than the left session. The corresponding game $\mathsf{Cl\text{-}Com}_{\mathsf{strong}}(1^\lambda, m_0, m_1)$ takes the same form as $G(1^\lambda, m_0, m_1)$, except the right sessions use commitment schemes $(\widetilde{\mathsf{Com}}_1, \widetilde{\mathsf{Open}}_1)$ and $(\widetilde{\mathsf{Com}}_2, \widetilde{\mathsf{Open}}_2)$. These right commitments must satisfy statistical binding (Definition 3.2).

Definition 5.2 (Strong Unclonability)**.** *A commitment scheme* $(\mathsf{Com}_L, \mathsf{Open}_L)$ *is* strongly unclonable *if for all right commitment schemes* $(\widetilde{\mathsf{Com}}_1, \widetilde{\mathsf{Open}}_1)$ *and* $(\widetilde{\mathsf{Com}}_2, \widetilde{\mathsf{Open}}_2)$*, every non-uniform QPT adversary* MiM*, every (non-uniform) QPT distinguisher pair* (D_1, D_2) *and all message pairs* $(m_0, m_1) \in \{0,1\}^n$*,*

$$\Pr\Big[\mathsf{Cl\text{-}Com}_{\mathsf{strong}}(1^\lambda, m_0, m_1) = 1\Big] \leq \frac{1}{2} + \mathrm{negl}(\lambda)$$

Since strong unclonability is a property about the composition of *different* protocols, it is much more difficult to achieve. We examine this notion further in Section 8.

Many-Many Unclonability. A natural extension is to consider an adversary which receives commitments in k left sessions and attempts to produce $k + r$ commitments to related values in the right sessions. Similar settings have been previously studied in the context of program-copy protection [Aar09, LLQZ22]. In the case of commitments, we additionally aim to generalize the notion of concurrent non-malleable commitments [DDN00, PR05a]. In concurrent non-malleable commitments, the *joint* values of all right commitments whose tags differ from the left commitments must be independent of the *joint* values of all left commitments. Intuitively, many-many unclonability guarantees that if there are k left sessions, then at most k right sessions are correlated with them; the others must be independent. The many-many commitment cloning game $\mathsf{Cl\text{-}Com}_{k,r}(1^\lambda, \vec{m}_0, \vec{m}_1)$ for k left sessions and $k + r$ right sessions is played as follows:

1. **Commitment:** In the left sessions, the challenger samples a random bit b, then commits to $\vec{m}_{b,i}$ for each $i \in [k]$. Simultaneously in the right sessions, MiM interacts as the committer with the $k + r$ honest receivers. At the end of this phase, MiM chooses a partition $P = (P_1, \ldots, P_{k+1})$ of $[k + r]$ and splits its state into $k + 1$ registers $\mathsf{MiM}_1, \ldots, \mathsf{MiM}_{k+1}$.
2. **Output:** For every $i \in [k + 1]$, compute $(\mathcal{R}_i, \widetilde{\mathcal{M}}_i) \leftarrow \mathsf{Ext}(\mathcal{R}_i)$, then compute $D_i(\mathsf{MiM}_i, \widetilde{\mathcal{M}}_i)$ and measure its output bit b_i. The adversary wins (output 1) if $b_i = b$ for every $i \in [k + 1]$. Otherwise the experiment outputs 0.

Definition 5.3 (Many-Many Unclonability for Commitments)**.** *A commitment scheme* $(\mathsf{Com}, \mathsf{Open})$ *is* many-many unclonable *if for all* $k = \mathrm{poly}(\lambda)$*,* $r = \mathrm{poly}(\lambda)$*, all (non-uniform) QPT adversaries* MiM*, all (non-uniform) QPT distinguishers* $D_1, \ldots, D_{k+1}$*, and all message vectors* $(\vec{m}_0, \vec{m}_1) \in \{0, 1\}^{nk}$*,*

$$\Pr\Big[\mathsf{Cl\text{-}Com}_{k,r}(1^\lambda, \vec{m}_0, \vec{m}_1) = 1\Big] \leq \frac{1}{2} + \mathrm{negl}(\lambda)$$

5.2 Constructions

A simple construction for unclonable commitments is to use an unclonable tag-generation protocol to generate a tag for a non-malleable commitment. This guarantees that one of the two right sessions uses a different tag than the left session. In this case, we can rely on the security of the non-malleable commitment.

Theorem 5.4. *Assuming* n_1*-round non-malleable commitments and* n_2*-round unclonable tag-generation protocols, there exists an unclonable commitment with* $n_1 + n_2$ *or* $n_1 + n_2$ *or* $n_1 + n_2 - 1$ *rounds.*

Proof. The construction is given in Figure 4. Hiding is immediate from the underlying non-malleable commitment. Binding and hiding follow directly from NMCom.

We show unclonability by reducing to the non-malleability of $(\mathsf{NMCom}, \mathsf{NMOpen})$ (Definition 3.3). The non-malleability adversary $\mathsf{MiM}_{\mathsf{nm}}$ internally runs the cloning adversary MiM_U. It internally runs TagGen in all three sessions to determine the tags tag, $\widetilde{\mathsf{tag}}_1$, and $\widetilde{\mathsf{tag}}_2$. By the unclonability of TagGen, $\widetilde{\mathsf{tag}}_r \neq \mathsf{tag}$ for some r. $\mathsf{MiM}_{\mathsf{nm}}$ chooses tag as the tag for the external left session and $\widetilde{\mathsf{tag}}_r$ as the tag for the external right session. It then plays the non-malleability experiment by forwarding messages between the external sessions and their respective internal sessions, while internally emulating an honest interaction in the unchosen internal right session. Finally, the non-malleability distinguisher D_{nm} simply runs the cloning distinguisher $D_{U,r}$ for session r. Observe that since $\widetilde{\mathsf{tag}}_r \neq \mathsf{tag}$, D_{nm} receives the same extracted value as $D_{U,r}$ with overwhelming probability. Thus, if the cloning distinguishers simultaneously succeed with probability $1/2 + \epsilon$, then D_{nm} succeeds with probability $\geq 1/2 + \epsilon - \mathrm{negl}(\lambda)$.

□

Tools. The construction uses the following tools.

- A non-malleable commitment $(\mathsf{NMCom}, \mathsf{NMOpen})$ with respect to commitment.
- An unclonable tag-generation protocol TagGen.

$\underline{\mathsf{UCom}(1^\lambda, m)}$**:** The committer and receiver do the following:

1. Compute a tag $\mathsf{tag} \leftarrow \mathsf{TagGen}$. If $\mathsf{tag} = \bot$, abort.
2. Run the non-malleable commitment $\langle \mathsf{NMCom}_C(m), \mathsf{NMCom}_R \rangle(\mathsf{tag})$.

   If the same party sends the last message of TagGen and the first message of $\langle \mathsf{NMCom}_C(m), \mathsf{NMCom}_R \rangle(\mathsf{tag})$, they may send these messages at the same time.

$\underline{\mathsf{UOpen}(1^\lambda)}$**:** Run the opening phase of the non-malleable commitment $\langle \mathsf{NMOpen}_C, \mathsf{NMOpen}_R \rangle(\mathsf{tag})$.

Figure 4: An Unclonable Commitment

If the non-malleable commitment is *concurrent* non-malleable, then the same construction is *many-many* unclonable.

Theorem 5.5. *Assuming n_1-round concurrent non-malleable commitments and n_2-round unclonable tag-generation protocols, there exists a many-many unclonable commitment with $n_1 + n_2$ or $n_1 + n_2 - 1$ rounds.*

Proof. Consider an execution of the unclonable commitment game with k left sessions and $k + r$ right sessions. Due to the unclonability of TagGen, for each left session there is at most one right session with the same identity $\mathsf{tag} \neq \bot$, except with negligible probability. This can be reduced to the $(1 \to 2)$ unclonability of the tag-generation by randomly selecting one of the $k = \mathrm{poly}(\lambda)$ left sessions and two of the $k + r = \mathrm{poly}(\lambda)$ right sessions. Therefore at least one of the $k + 1$ sets of right sessions does not include any sessions whose tags match a left session. Many-many unclonability can be reduced to the concurrent non-malleability of $(\mathsf{NMCom}, \mathsf{NMOpen})$ similarly to the proof of Theorem 5.4; identify these sessions at the end of the commitment phase, then run the corresponding cloning distinguisher. □

When combining the construction in Figure 4 with the post-quantum tag-generation protocol and post-quantum non-malleable commitments, we get a *post-quantum* construction of unclonable commitments. Post-quantum non-malleable commitments were first introduced by [ABG+21]. Later work improved the assumptions to only require post-quantum one-way functions [BLS22], and even achieved a constant-round version [LPY23].

Corollary 5.6 (Post-Quantum Interactive Unclonable Commitments)**.** *Assuming post-quantum one-way functions, there exist constant-round post-quantum unclonable commitments.*

Proof. This is immediate from Theorem 5.4, Lemma 4.2 (two-round post-quantum unclonable tag-generation from one-way functions), and [LPY23]. □

Corollary 5.7 (Unclonable Commitments from PKQM)**.** *Assuming non-interactive non-malleable commitments and public key quantum money, there exist non-interactive unclonable commitments.*

Proof. This is immediate from Theorem 5.4 and Lemma 4.3 (non-interactive tag-generation from public key quantum money). □

We note that non-interactive non-malleable commitments with respect to commitment[4] can be constructed in the CRS model from any non-malleable public key encryption (PKE) scheme with perfect correctness. Non-malleable PKE can be constructed from any PKE with security against chosen-plaintext attacks [CDMW18]. Although [CDMW18] proved only classical security, we observe that their construction is also post-quantum secure. At a high level, their construction uses a CRS which consists of many public encryption keys. Each tag is associated with a different set of these keys. To encrypt a message m, it is first encoded as a matrix M, then each entry in the matrix is encrypted using a separate public key from the CRS, according to the tag. If the left and right sessions use different tags tag and $\widetilde{\mathsf{tag}}$, then it is possible to decrypt a number of entries from M without learning anything about the messages encrypted in the left session, by using keys which appear in $\widetilde{\mathsf{tag}}$ but not in tag. [CDMW18] uses a clever encoding scheme related to Reed-Solomon codes to ensure that this is enough to break the right commitment (if one controls the CRS) without breaking the left commitment, so the right committed message must be independent of the left committed message. We note that this extractor is straightline and also succeeds against a quantum adversary.

We now return to the construction sketch of non-interactive non-malleable commitments. The CRS consists of a public key. To commit to a message, encrypt it under the public key in the CRS, and to open, reveal the randomness. Perfect correctness guarantees binding, since there is a unique message that was encrypted. Although non-malleable PKE is typically defined *without* a tag, a tag t can be generically added by committing to $t\|m$ instead of m [PR05b].

5.2.1 Non-Interactive Construction from Weaker Assumptions

Unfortunately, public key quantum money is currently only known from strong assumptions such as post-quantum indistinguishability obfuscation [AC12, FGH+12, Zha21]. In order to build an unclonable commitment from weaker assumptions, we will make use of the fact that commitments have both a commitment phase and an opening phase. This structure allows for the tag to be generated in the commitment phase, then "confirmed" in the opening phase. The delayed verification allows for the usage of techniques from private-key quantum money in order to guarantee that the tag cannot be cloned.

Theorem 5.8. *If there exist post-quantum non-interactive commitments which are $(1,4)$-concurrent non-malleable with respect to commitment, then there exist non-interactive unclonable commitments.*

Proof. The construction is given in Figure 5. Hiding follows directly from $(\mathsf{NMCom}, \mathsf{NMOpen})$. The statistical binding extractor invokes the statistical binding extractor for $(\mathsf{NMCom}, \mathsf{NMOpen})$ on both $\mathsf{nmcom}_{\mathsf{tag}}$ and $\mathsf{nmcom}_{\mathsf{msg}}$. The extracted messages are parsed as (x, θ) and m, respectively. It measures $|\psi\rangle$ with respect to $\{\Pi_x^\theta, I - \Pi_x^\theta\}$. Finally, it outputs the leftover state, along with $m' = m$ if the measurement result was Π_x^θ, and $m' = \bot$ otherwise. If $(x, \theta) = \bot$, the extracted message is $m' = \bot$. It is straightforward to verify the three statistical binding properties for this extractor, by combining the statistical binding of

[4]Non-interactive non-malleable commitments were previously studied in [DIO98]. However, they achieve non-malleability with respect to opening, which is weaker than non-malleability with respect to commitment.

Tools. The construction uses a post-quantum non-interactive commitment $(\mathsf{NMCom}, \mathsf{NMOpen})$ which is concurrent non-malleable with respect to commitment.

$\underline{\mathsf{UCom}_C(1^\lambda, m)}$**:** The committer does the following.

1. Sample $x, \theta \leftarrow \{0,1\}^\lambda$.
2. Compute $(\mathsf{nmcom}_{\mathsf{tag}}, d_{\mathsf{tag}}) \leftarrow \mathsf{NMCom}_C(\mathsf{tag} = 0, (x, \theta))$.
3. Compute $(\mathsf{nmcom}_{\mathsf{msg}}, d_{\mathsf{msg}}) \leftarrow \mathsf{NMCom}_C(\mathsf{tag} = \mathsf{nmcom}_{\mathsf{tag}}, m)$.
4. Send $(\mathsf{nmcom}_{\mathsf{tag}}, \mathsf{nmcom}_{\mathsf{msg}}, |x\rangle_\theta)$ to the receiver.

$\underline{\mathsf{UCom}_R(1^\lambda, (\mathsf{nmcom}_{\mathsf{tag}}, \mathsf{nmcom}_{\mathsf{msg}}, |\psi\rangle))}$**:** Reject the commitment if $\mathsf{nmcom}_{\mathsf{tag}} = 0$.

$\underline{\mathsf{UOpen}_C(1^\lambda)}$**:** The committer sends the openings $(d_{\mathsf{tag}}, (x, \theta))$ and (d_{msg}, m) to the receiver.

$\underline{\mathsf{UOpen}_R(1^\lambda, (\mathsf{nmcom}_{\mathsf{tag}}, \mathsf{nmcom}_{\mathsf{msg}}, |\psi\rangle), ((d_{\mathsf{tag}}, (x, \theta)), (d_{\mathsf{msg}}, m))}$**:** The receiver does the following:

1. Verify the openings to $\mathsf{nmcom}_{\mathsf{tag}}$ and $\mathsf{nmcom}_{\mathsf{msg}}$. If either reject, output $\perp$. Otherwise, continue.
2. Measure $H^\theta |\psi\rangle$ in the basis θ. If the result is x, output m. Otherwise, output $\perp$.

Figure 5: A Non-Interactive Unclonable Commitment from Post-Quantum Non-Malleable Commitments

$(\mathsf{NMCom}, \mathsf{NMOpen})$ with the observation that any honest receiver will either perform the same measurement or output $\perp$.

We reduce unclonability to the concurrent non-malleability of $(\mathsf{NMCom}, \mathsf{NMOpen})$, with four right sessions (Definition 3.3). Say that MiM_U, $D_{U,1}$, $D_{U,2}$ win the cloning game with probability $1/2 + \varepsilon$, for noticeable ε. We construct a non-malleability adversary $\mathsf{MiM}_{\mathsf{nm}}$ and distinguisher D_{nm} which win the concurrent non-malleability game with probability $1/2 + \varepsilon - \mathrm{negl}(\lambda)$. $\mathsf{MiM}_{\mathsf{nm}}$ interacts with MiM_U in an internal unclonability experiment, as follows.

1. Sample $x, \theta \leftarrow \{0,1\}^\lambda$. Generate $\mathsf{nmcom}_{\mathsf{tag}} \leftarrow \mathsf{NMCom}(\mathsf{tag} = 0, (x, \theta))$, subject to $\mathsf{nmcom}_{\mathsf{tag}} \neq 0$. Pick $\mathsf{nmcom}_{\mathsf{tag}}$ as the tag tag_L for the left session of the non-malleability experiment.
2. Receive a commitment $\mathsf{nmcom}_{\mathsf{msg}}$, which has $\mathsf{tag} = \mathsf{nmcom}_{\mathsf{tag}}$.
3. Send $(\mathsf{nmcom}_{\mathsf{tag}}, \mathsf{nmcom}_{\mathsf{msg}}, |x\rangle_\theta)$ to MiM_U, internally. Receive $\left(\widetilde{\mathsf{nmcom}}_{\mathsf{tag},1}, \widetilde{\mathsf{nmcom}}_{\mathsf{msg},1}, \widetilde{\rho}_1\right)$ and $\left(\widetilde{\mathsf{nmcom}}_{\mathsf{tag},2}, \widetilde{\mathsf{nmcom}}_{\mathsf{msg},2}, \widetilde{\rho}_2\right)$ from MiM_U in right sessions 1 and 2 of the internal commitment cloning game, respectively.
4. Send each of $\widetilde{\mathsf{nmcom}}_{\mathsf{tag},1}$, $\widetilde{\mathsf{nmcom}}_{\mathsf{msg},1}$, $\widetilde{\mathsf{nmcom}}_{\mathsf{tag},1}$, and $\widetilde{\mathsf{nmcom}}_{\mathsf{msg},1}$ in one of the four right sessions of the external non-malleability experiment. Keep $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ as local state.

The corresponding concurrent non-malleability distinguisher D_{nm} does the following.

1. D_{nm} is intialized with $\mathsf{MiM}_{\mathsf{nm}}$'s leftover local state, which includes $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$. Additionally, it receives four extracted values $\widetilde{v}_{\mathsf{tag},1}$, $\widetilde{v}_{\mathsf{msg},1}$, $\widetilde{v}_{\mathsf{tag},2}$, and $\widetilde{v}_{\mathsf{msg},2}$ from the non-malleability challenger.

2. D_{nm} attempts to extract the value of the unclonable commitment in right session 1 of the internal commitment cloning game, using $\widetilde{v}_{\mathsf{tag},1}$ and $\widetilde{v}_{\mathsf{msg},1}$. It checks whether the tag $\widetilde{\mathsf{tag}}_1 = \widetilde{\mathsf{nmcom}}_{\mathsf{tag},1}$ is valid by parsing $\widetilde{v}_{\mathsf{tag},1} = (x_1, \theta_1)$ and measuring whether $\widetilde{\rho}_1$ is $|x\rangle_\theta$. The tag is valid if $\widetilde{v}_{\mathsf{tag},1} \neq \bot$ and the measurement returns success. If the tag is valid, then the extracted value is $\widetilde{v}_1 = v_{\mathsf{msg},1}$. Otherwise, the extracted value is $\widetilde{v}_1 = \bot$. We say that D_{nm} correctly extracted $\widetilde{v}_1$ if $\widetilde{\mathsf{tag}}_1 \neq \mathsf{tag}_L$ or $\widetilde{\mathsf{tag}}_1$ is *not* a valid tag.

3. D_{nm} does the same to extract $\widetilde{v}_2$ for the internal right session two.

4. If D_{nm} correctly extracted $\widetilde{v}_1$, then it outputs $D_{U,1}(\widetilde{v}_1)$. Otherwise, it outputs $D_{U,2}(\widetilde{v}_2)$.

Observe that whenever $\widetilde{\mathsf{tag}}_1 \neq \mathsf{tag}_L$, the external non-malleability challenger sets $\widetilde{v}_{\mathsf{msg},1}$ to be the value extracted from $\widetilde{\mathsf{nmcom}}_{\mathsf{msg},1}$. Therefore in this case, $\widetilde{v}_1$ is exactly the same as the output of the statistical binding extractor for UCom on $\left(\widetilde{\mathsf{nmcom}}_{\mathsf{tag},1}, \widetilde{\mathsf{nmcom}}_{\mathsf{msg},1}, \widetilde{\rho}_1\right)$. Similarly, D_{nm} correctly extracts this message if $\widetilde{\mathsf{tag}}_1$ is not a valid tag, since the message is $\bot$. A similar statement applies for right session 2 of the internal cloning game.

Observe that whenever D_{nm} correctly extracts $\widetilde{v}_1$, $D_{U,1}(\widetilde{v}_1)$ is identically distributed to $D_{U,1}$'s output in the commitment cloning experiment. A similar statement holds for $\widetilde{v}_2$ and $D_{U,2}$. Thus, D_{nm} *loses* the non-malleability game only if it fails to correctly extract both $\widetilde{v}_1$ and $\widetilde{v}_2$, or if at least one of $D_{U,1}(\widetilde{v}_1)$ and $D_{U,2}(\widetilde{v}_2)$ incorrectly guess the left session's message bit. Claim 5.9 shows that the former happens with $\mathrm{negl}(\lambda)$ probability. By assumption, the latter occurs with probability $\leq 1/2 - \varepsilon$. The union bound implies that $\mathsf{MiM}_{\mathsf{nm}}$ and D_{nm} win the non-malleability game with probability $\geq 1 - (1/2 - \varepsilon + \mathrm{negl}(\lambda)) = 1/2 + \varepsilon - \mathrm{negl}(\lambda)$.

Claim 5.9. *D_{nm} correctly extracts at least one of $\widetilde{v}_1$ and $\widetilde{v}_2$, except with $\mathrm{negl}(\lambda)$ probability.*

Proof. D_{nm} fails to extract both messages only if $\mathsf{tag}_L = \widetilde{\mathsf{tag}}_1 = \widetilde{\mathsf{tag}}_2$ and both right tags are valid. Say the first condition holds. Since $(\mathsf{NMCom}, \mathsf{NMOpen})$ is statistically binding, the values of these commitments are either $\bot$ (in which case the tag is invalid), or (x, θ). Thus, in this case D_{nm} fails to extract both messages only if it measures $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$ with respect to $\{\Pi_x^\theta, I - \Pi_x^\theta\}$ and receives success in both cases. Let $\widetilde{\rho}$ be the joint state of $\widetilde{\rho}_1$ and $\widetilde{\rho}_2$. It is sufficient to show that

$$\mathrm{Tr}\left[\left(\Pi_x^\theta \otimes \Pi_x^\theta\right) \widetilde{\rho} \left(\Pi_x^\theta \otimes \Pi_x^\theta\right)\right] = \mathrm{negl}(\lambda)$$

Observe that $\widetilde{\rho}$ does not change between the end of the commitment phase and when the measurement is applied. Thus, it is sufficient to argue about $\widetilde{\rho}$ in the experiment which ends at the end of the commitment phase. Consider a hybrid experiment where $\mathsf{nmcom}_{\mathsf{tag}}$ is a commitment to 0, instead of a commitment to (x, θ). Let $\widetilde{\rho}_H$ be the joint state resulting in this hybrid. By the hiding of $(\mathsf{NMCom}, \mathsf{NMOpen})$,

$$\left|\mathrm{Tr}\left[\left(\Pi_x^\theta \otimes \Pi_x^\theta\right) \widetilde{\rho} \left(\Pi_x^\theta \otimes \Pi_x^\theta\right)\right] - \mathrm{Tr}\left[\left(\Pi_x^\theta \otimes \Pi_x^\theta\right) \widetilde{\rho}_H \left(\Pi_x^\theta \otimes \Pi_x^\theta\right)\right]\right| = \mathrm{negl}(\lambda)$$

Otherwise, an adversary could distinguish between $\mathsf{Com}((x, \theta))$ and $\mathsf{Com}(0)$ by measuring the joint state with respect to $\{\left(\Pi_x^\theta \otimes \Pi_x^\theta\right), I - \left(\Pi_x^\theta \otimes \Pi_x^\theta\right)\}$.

Since com and nmcom are independent of (x, θ), the fact that

$$\mathrm{Tr}\left[\left(\Pi_x^\theta \otimes \Pi_x^\theta\right) \widetilde{\rho}_H \left(\Pi_x^\theta \otimes \Pi_x^\theta\right)\right] = \mathrm{negl}(\lambda)$$

immediately reduces to monogamy of entanglement [TFKW13]. □

□

The construction in Figure 5 is also *many-many* unclonable (Definition 5.3).

Theorem 5.10. *If there exist post-quantum non-interactive commitments which are concurrent non-malleable with respect to commitment, then there exist non-interactive many-many unclonable commitments.*

Proof Sketch. The proof is very similar to the proof of Theorem 5.8, so we omit the full details. At a high level, we define a $\mathsf{MiM}_{\mathsf{nm}}$ which outputs two non-malleable commitments for each right session of the cloning game. The distinguisher D_{nm} attempts to extract the values of the unclonable commitments using the values revealed in the non-malleability game. If it succeeds in extracting all values in at least one distinguishing set, it runs the corresponding cloning distinguisher on the extracted values. D_{nm} loses the distinguishing game *only* if it fails to extract one value from every distinguishing set or if at least one of the cloning distinguishers would guess incorrectly. By assumption, the latter occurs with probability $\leq 1/2 - \epsilon$.

The crux of the argument is to show that D_{nm} successfully extracts every value from some distinguishing set with overwhelming probability. If this is the case, then D_{nm} wins the non-malleability game with probability $\geq 1/2 + \epsilon - \mathrm{negl}(\lambda)$, which violates the security of $(\mathsf{NMCom}, \mathsf{NMOpen})$. Similarly to Claim 5.9, it is possible to show that for every left commitment in the cloning game, at most one right commitment in the cloning game validly uses the same tag, except with negligible probability. Since there are k left commitments and $k+1$ distinguishing sets, at least one distinguishing set only contains commitments which do *not* validly use any tags from a left session. The values revealed by the non-malleability game suffice to extract every value from this set. □

6 Unclonable Proofs

6.1 Definition

Intuitively, an unclonable proof should guarantee that an adversary cannot use a simulated left proof to produce two new proofs without knowing either witness. This is very similar to the related notion of non-malleable zero knowledge. However, a key difference is unclonability does not require the right proofs to be different than the left proof for its guarantee to hold. Thus, *unclonability is meaningful even if the adversary attempts to directly forward the left proof.*

Proof Cloning Experiment. The proof cloning experiment $\mathsf{Cl\text{-}Arg}(1^\lambda, x, w, \rho)$ uses a proof $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ for a language $\mathcal{L}$ and is parameterized by a statement-witness pair $(x, w) \in R_{\mathcal{L}}$ and a quantum advice string ρ. It involves three sessions: a "left" session between an honest left prover $P_L(x, w)$ and an adversarial man-in-the-middle $\mathsf{MiM}(\rho)$, and two right sessions between MiM and an honest right verifier, respectively named V_1 and V_2. The sessions may be interleaved arbitrarily. In the left session, P_L attempts to convince MiM that $x \in \mathcal{L}$, using $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$. In the right sessions, MiM chooses statements $\widetilde{x_1}$ and $\widetilde{x_2}$ adaptively, then attempts to convince V_1 and V_2 that $\widetilde{x}_1 \in \mathcal{L}$ and $\widetilde{x}_2 \in \mathcal{L}$, respectively, using the same proof system. $\mathsf{Cl\text{-}Arg}(1^\lambda, x, w, \rho)$ outputs a tuple $(\tau_{\mathsf{MiM}}, (\widetilde{x}_1, \widetilde{a}_1), (\widetilde{x}_2, \widetilde{a}_2))$ consisting of the view of MiM, the statements $\widetilde{x}_1$, $\widetilde{x}_2$ argued in the right sessions, and the acceptance bits $\widetilde{a}_1$, $\widetilde{a}_2$ of the two right verifiers (where 1 is accept).

Definition 6.1 (Simulation-Extraction-Unclonability)**.** *A proof system* $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ *is* SE-unclonable *if there exists a QPT simulator-extractor* SE*, which outputs a tuple* $((\tau_{\mathsf{MiM}}, (\widetilde{x}_1, \widetilde{a}_1), (\widetilde{x}_2, \widetilde{a}_2)), \widetilde{w_1}, \widetilde{w_2})$*, such that for every (non-uniform) QPT adversary* MiM *with quantum advice* ρ *and every* $(x, w) \in R_{\mathcal{L}}$*,*

1. ***Zero Knowledge.*** $\mathsf{CI\text{-}Arg}(1^\lambda, x, w, \rho)$ *is computationally indistinguishable from the first output* $(\tau_{\mathsf{MiM}}, (\widetilde{x}_1, \widetilde{a}_1), (\widetilde{x}_2, \widetilde{a}_2))$ *of* $\mathsf{SE}(1^\lambda, x, \rho)$*.*
2. ***Witness Extraction.*** *If* $\mathsf{SE}(1^\lambda, x, \rho)$ *outputs acceptance bits* $\widetilde{a}_1 = \widetilde{a}_2 = 1$*, then at least one of the following holds, except with negligible probability:* $(\widetilde{x}_1, \widetilde{w}_1) \in \mathsf{Rel}_{\mathcal{L}}$ *or* $(\widetilde{x}_2, \widetilde{w}_2) \in \mathsf{Rel}_{\mathcal{L}}$*.*

It is also interesting to consider the case where the right sessions may use different proof systems $\langle \widetilde{\mathsf{Prove}_1}, \widetilde{\mathsf{Verify}_1} \rangle$ and $\langle \widetilde{\mathsf{Prove}_2}, \widetilde{\mathsf{Verify}_2} \rangle$ than the left session. Note that these proof systems may be for different languages $\widetilde{\mathcal{L}}_1$ and $\widetilde{\mathcal{L}}_2$ than the left proof system. We denote the corresponding experiment as $\mathsf{CI\text{-}Arg}_{\mathsf{strong}}(1^\lambda, x, w, \rho)$.

Definition 6.2 (Strong SE-Unclonability)**.** *A proof system* $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ *is* strongly SE-unclonable *if there exists a QPT simulator-extractor* SE *which outputs a tuple* $((\tau_{\mathsf{MiM}}, (\widetilde{x}_1, \widetilde{a}_1), (\widetilde{x}_2, \widetilde{a}_2)), \widetilde{w_1}, \widetilde{w_2})$*, such that for every (non-uniform) QPT adversary* MiM *with quantum advice* ρ *and every* $(x, w) \in R_{\mathcal{L}}$*,*

1. ***Zero Knowledge.*** $\mathsf{CI\text{-}Arg}_{\mathsf{strong}}(1^\lambda, x, w, \rho)$ *is computationally indistinguishable from the first output of* $\mathsf{SE}(1^\lambda, x, \rho)$*.*
2. ***Witness Extraction.*** *If* $\mathsf{SE}(1^\lambda, x, \rho)$ *outputs acceptance bits* $\widetilde{a}_1 = \widetilde{a}_2 = 1$*, then at least one of the following holds, except with negligible probability:* $(\widetilde{w}_1, \widetilde{x}_1) \in \mathsf{Rel}_{\widetilde{\mathcal{L}}_1}$ *or* $(\widetilde{w}_2, \widetilde{x}_2) \in \mathsf{Rel}_{\widetilde{\mathcal{L}}_2}$*.*

Since strong SE-unclonability is a property about the composition of *different* protocols, it is much more difficult to achieve. We explore this notion further in Section 7.

Soundness-Unclonability. An alternative notion only requires that at least one of the two right sessions retains its soundness, even if the left session is simulated for a potentially false statement. Note that in this case, the adversary can violate the soundness of one right session simply by forwarding the left session.

Definition 6.3 (Soundness Unclonability)**.** *A proof* $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ *is* soundness-unclonable *if there exists a QPT simulator* $\mathcal{S}$ *such that for every QPT* MiM *with quantum advice* ρ*,*

1. ***Zero Knowledge.*** *For every* $(x, w) \in \mathsf{Rel}_{\mathcal{L}}$*,* $\mathcal{S}(1^\lambda, x, \rho)$ *is computationally indistinguishable from* $\mathsf{CI\text{-}Arg}(1^\lambda, x, w, \rho)$*.*
2. ***Simulation-Soundness.*** *For a simulated execution* ν*, let* $\mathsf{SV}(\nu)$ *be the number of right sessions* i *where* R_i *outputs accept, but* $\widetilde{x}_i \notin \mathcal{L}$*, i.e. a soundness violation occurs. For every* $x \in \{0,1\}^n$*,*

$$\Pr\Big[\mathsf{SV}(\nu) > 1 : \nu \leftarrow \mathcal{S}(1^\lambda, x, \rho)\Big] = \mathrm{negl}(\lambda)$$

We may additionally consider the strong variant, which permits any proof in the right sessions. The relation between SE-unclonability and soundness-unclonability seems similar to the relation between simulation-extraction [PR05b] and simulation-soundness [Sah99] in non-malleable zero knowledge, which [JP14] shows are incomparable.

Many-Many Unclonability. A natural extension is to consider an adversary which sees k proofs in the left sessions and attempts to produce $k+r$ proofs in the right sessions, for polynomial k and r. Such settings have previously been studied in the context of program copy-protection [Aar09, LLQZ22]. Intuitively, $k \to k+r$ unclonability should guarantee that the adversary must know at least r witnesses (or that at least r sessions are sound). We denote the corresponding game as $\mathsf{Cl\text{-}Arg}_{k,r}$.

Definition 6.4 (Many-Many SE-Unclonability)**.** *A proof system* $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ *is* many-many SE-unclonable *if for every* $k = \mathrm{poly}(\lambda)$*,* $r = \mathrm{poly}(\lambda)$*, there exists a QPT simulator-extractor* $\mathsf{SE}_{k,r}$ *which outputs a tuple* $((\rho_{\mathsf{MiM}}, (\widetilde{x}_i, \widetilde{a}_i)_{i\in[k+r]}), (\widetilde{w}_i)_{i\in[k+r]})$*, such that for every (non-uniform) QPT adversary* MiM *with quantum advice* ρ *and for every* $(x, w) \in R_{\mathcal{L}}$*,*

1. ***Zero Knowledge.*** $\mathsf{Cl\text{-}Arg}_{k,r}(1^\lambda, x, w, \rho)$ *is computationally indistinguishable from the first output of* $\mathsf{SE}_{k,r}(1^\lambda, x, \rho)$*.*
2. ***Witness Extraction.*** *For a simulated execution* ν*, let* $\mathsf{Acc}(\nu) = \sum_{i=1}^{k+r} \widetilde{a}_i$ *be the number of right sessions which output accept. Let* $W(\nu, (\widetilde{w}_i)_{i\in[k+r]})$ *be the number of right sessions* i *such that* $(\widetilde{x}_i, \widetilde{w}_i) \in \mathsf{Rel}_{\mathcal{L}}$*. Then*

$$\Pr\Big[\mathsf{Acc}(\nu) > W(\nu, (\widetilde{w}_i)_{i\in[k+r]}) + k : (\nu, (\widetilde{w}_i)_{i\in[k+r]}) \leftarrow \mathsf{SE}_{k,r}(1^\lambda, x, \rho)\Big] = \mathrm{negl}(\lambda)$$

Definition 6.5 (Many-Many Soundness Unclonability)**.** *A proof* $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$ *is* many-many soundness-unclonable *if for every* $k = \mathrm{poly}(\lambda)$ *and* $r = \mathrm{poly}(\lambda)$*, there exists a QPT simulator* $\mathcal{S}$ *such that for every QPT* MiM *with quantum advice* ρ*,*

1. ***Zero Knowledge.*** *For every* $(x, w) \in \mathsf{Rel}_{\mathcal{L}}$*,* $\mathcal{S}(1^\lambda, x, \rho)$ *is computationally indistinguishable from* $\mathsf{Cl\text{-}Arg}_{k,r}(1^\lambda, x, w, \rho)$*.*
2. ***Simulation-Soundness.*** *For a simulated execution* ν*, let* $\mathsf{SV}(\nu)$ *be the number of right sessions* i *where* R_i *outputs accept, but* $\widetilde{x}_i \notin \mathcal{L}$*, i.e. a soundness violation occurs. For every* $x \in \{0,1\}^n$*,*

$$\Pr\Big[\mathsf{SU}(\nu) > k : \nu \leftarrow \mathcal{S}(1^\lambda, x, \rho)\Big] = \mathrm{negl}(\lambda)$$

6.2 Constructions

Theorem 6.6. *Assuming* n_1*-round unclonable tag-generation protocols and* n_2*-round simulation-sound proofs, there exist soundness-unclonable proofs with* $n_1 + n_2$ *or* $n_1 + n_2 - 1$ *rounds.*

Proof. The construction is given in Figure 6. Soundness is immediate from the underlying proof. The simulator constructs an adversary for the zero knowledge property of the underlying simulation-sound proof by internally emulating the two right verifiers. Consider an execution of the simulator where V_1 and V_2 respectively accept. Say that TagGen output tag in the left session and $\widetilde{\mathsf{tag}}_1$, $\widetilde{\mathsf{tag}}_2$ in the right sessions in this execution. By the security of TagGen, $\mathsf{tag} \neq \widetilde{\mathsf{tag}}_b$ for some $b \in \{1, 2\}$. Otherwise, one of the right tags is $\perp$, and so the corresponding right receiver would have rejected the proof. By the simulation-soundness of $\langle \mathsf{Prove}, \mathsf{Verify} \rangle$, right session b is sound. Therefore $\widetilde{x}_b \in \mathcal{L}$, except with negligible probability, since R_b accepted. □

Theorem 6.7. *Assuming* n_1*-round unclonable tag-generation protocols and* n_2*-round simulation-extractable proofs, there exist SE-unclonable proofs with* $n_1 + n_2$ *or* $n_1 + n_2 - 1$ *rounds.*

Tools. The construction uses the following tools.

- A simulation-extractable (respectively, simulation-sound) proof $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$.
- An unclonable tag-generation protocol TagGen.

$\underline{\langle \mathsf{UProve}, \mathsf{UVerify} \rangle(1^\lambda, x)}$**:** The committer and receiver do the following:

1. Run the tag-generation protocol TagGen. Let tag be the output of the protocol.
2. Run the proof $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle(\mathsf{tag}, x)$ using tag as the non-malleable tag.

Round Reduction Optimization. If the same party sends the last message of TagGen and the first message of $\langle \mathsf{Prove}, \mathsf{Verify} \rangle(\mathsf{tag}, x)$, they may send the messages at the same time.

Figure 6: An SE-Unclonable (respectively, Simulation-Sound) Proof

Proof. The construction is given in Figure 6. Soundness is immediate from the underlying proof. To show SE-unclonability, we construct the simulator-extractor SE_U. We start by constructing a halfway simulator-extractor $\mathsf{SE}_{U,1/2}$ which will succeed in simulating the adversary's view, but will only extract a valid witness with probability $1/2 \pm \mathrm{negl}(\lambda)$. Given an unclonability adversary MiM_U^*, we will define an adversary $\mathsf{MiM}_{\mathsf{nm}}$ for the non-malleability game with $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$. We will then use $\mathsf{MiM}_{\mathsf{nm}}$ to define $\mathsf{SE}_{U,1/2}$.

Non-Malleability Adversary. $\mathsf{MiM}_{\mathsf{nm}}$ uniformly samples $b \leftarrow \{1, 2\}$. Set $\bar{b}$ to be the unchosen value. It internally runs the unclonability experiment using MiM_U. It runs the sessions of the unclonability experiment as follows:

- **Left Session.** Before the left TagGen completes, internally run an interaction between an honest left prover and MiM_U. When the left TagGen completes with output tag, choose tag as the tag for the left session of the external non-malleability experiment. Afterwards, when it receives a left message m_{nm} in the non-malleability experiment, it forwards it to MiM_U as a left message in the unclonability experiment. Similarly, it forwards messages from MiM_U in the left session of the unclonability experiment to the left session of the external non-malleability experiment.
- **Right Session** b**.** Before $\widetilde{\mathsf{TagGen}}_b$ completes, internally run an interaction between an honest R_b and MiM_U. When the left $\widetilde{\mathsf{TagGen}}_b$ completes with output $\widetilde{\mathsf{tag}}_b$, choose $\widetilde{\mathsf{tag}}_b$ as the tag for the right session of the external non-malleability experiment. Afterwards, when it receives a right message $\widetilde{m}_{\mathsf{nm}}$ from the non-malleability experiment, it forwards it to MiM_U as a message from R_b. Similarly, upon receiving a message from MiM_U to R_b, it forwards it to the right session of the external non-malleability experiment.
- **Right Session** $\bar{b}$**.** Internally run an interaction between an honest $R_{\bar{b}}$ and MiM_U.

Half Simulator-Extractor. Let $\mathsf{SE}_{\mathsf{nm}}$ be the simulator-extractor for the underlying simulation-extractable proof $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$. $\mathsf{SE}_{U,1/2}$ is defined as follows.

1. Run $\mathsf{SE}_{\mathsf{nm}}$ on $\mathsf{MiM}_{\mathsf{nm}}$ to obtain $\mathsf{MiM}_{\mathsf{nm}}$'s view and a witness $\widetilde{w}_b$. $\mathsf{MiM}_{\mathsf{nm}}$'s view consists of MiM_U's view $\rho_{\mathsf{MiM},U}$, b, and $R_{\overline{b}}$'s view. Let $a_{\overline{b}}$ be $R_{\overline{b}}$'s acceptance bit.

2. If $\mathsf{tag} = \widetilde{\mathsf{tag}}_{\overline{b}} \neq \widetilde{\mathsf{tag}}_b$, set the extraction success bit $s = 1$ (representing successful extraction). If $\mathsf{tag} = \widetilde{\mathsf{tag}}_b \neq \widetilde{\mathsf{tag}}_{\overline{b}}$, set $s = 0$. Otherwise, sample s uniformly at random from $\{0, 1\}$.

3. Output $\rho_{\mathsf{MiM},U}$, the witness and acceptance bit $(\widetilde{w}_b, a_b)$ corresponding to right session b, the witness and acceptance bit $(\widetilde{w}_{\overline{b}} = \bot, a_{\overline{b}})$ corresponding to right session b, and the extraction success bit s.

Observe that $\rho_{\mathsf{MiM},U}$ is computationally indistinguishable from MiM_U's view in the proof cloning experiment. If this were not the case, then the simulation-extractability of $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ could be broken by implementing $\mathsf{MiM}_{\mathsf{nm}}$ using MiM_U.

Claim 6.8.

$$\Pr[s = 1 \wedge a_1 = a_2 = 1 \wedge (\widetilde{w}_b, \widetilde{x}_b) \notin \mathcal{R}] = 1 - \mathrm{negl}(\lambda)$$

Proof. If $s = 1$, then there are three cases. In the first case, $\mathsf{tag} = \widetilde{\mathsf{tag}}_{\overline{b}} \neq \widetilde{\mathsf{tag}}_b$. Since $\widetilde{\mathsf{tag}}_b \neq \mathsf{tag}$, the simulation-extractability of $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ guarantees that $(\widetilde{w}_b, \widetilde{x}_b) \in \mathcal{R}$ whenever $a_b = 1$, except with negligible probability. In the second case, $\mathsf{tag} \neq \widetilde{\mathsf{tag}}_{\overline{b}}$ and $\mathsf{tag} \neq \widetilde{\mathsf{tag}}_b$. The proof for case two is the same as for case one. In the third case, $\widetilde{\mathsf{tag}}_1 = \widetilde{\mathsf{tag}}_2 = \mathsf{tag}$. By the unclonability of TagGen, all three tags are $\bot$, except with negligible probability. Therefore both right verifiers reject in this case, i.e. $a_1 = a_2 = 0$. □

Claim 6.9. *Define the projector Π_s onto the space where $\mathsf{SE}_{U,1/2}$ outputs $s = 1$. Fix any polynomial-size unitary U_{MiM}. Let $p(|\psi\rangle)$ be the probability that a measurement of $\{\Pi_s, I - \Pi_s\}$ on $\mathsf{SE}_{U,1/2}(U_{\mathsf{MiM}}, |\psi\rangle)$ outputs success (result Π_s). For every quantum advice string $|\psi\rangle$ and every $\varepsilon = 1/\operatorname{poly}(\lambda)$,*

$$|p(|\psi\rangle) - 1/2| \leq \varepsilon$$

Proof. We can divide the success behavior into two cases. If *exactly* one of the two right tags is equal to the left tag, then $s = 1$ if and only if $\mathsf{tag} = \widetilde{\mathsf{tag}}_b$. Otherwise, s is sampled uniformly at random. The success probability in the latter case is clearly exactly $1/2$. It remains to be shown that the success probability in the former case is close to $1/2$. Consider a hybrid experiment which runs the real proof unclonability game and outputs MiM_U's view along with a uniformly random $b \leftarrow \{1, 2\}$. The probability that $\widetilde{\mathsf{tag}}_b \neq \mathsf{tag}$ is negligibly far between the hybrid experiment and the simulated experiment. Otherwise, the zero knowledge property of $\mathsf{SE}_{\mathsf{nm}}$ could be violated by running $\mathsf{MiM}_{\mathsf{nm}}$ and checking whether $\widetilde{\mathsf{tag}}_b \neq \mathsf{tag}$, where b is the internal value sampled by $\mathsf{MiM}_{\mathsf{nm}}$. Since we have conditioned on exactly one of the two right tags being equal to the left tag, the probability that $\widetilde{\mathsf{tag}}_b \neq \mathsf{tag}$ in the hybrid experiment is exactly $1/2$. Thus, in the simulated experiment, this probability is negligibly far from $1/2$. □

Full Simulator-Extractor. The full simulator-extractor SE_U uses Watrous's rewinding lemma with small perturbations [Wat09] to run $\mathsf{SE}_{U,1/2}$ until $s = 1$. It outputs everything that $\mathsf{SE}_{U,1/2}$ output, except for s. By Claim 6.9, $\mathsf{SE}_{U,1/2}$ satisfies the requirements for the rewinding lemma (Lemma 3.5). Furthermore, Claim 6.8 shows that SE_U satisfies witness extraction when $s = 1$, which happens with overwhelming probability.

□

As it turns out, the construction in Figure 6 is also *many-many* unclonable if it is instantiated with a concurrent non-malleable proof.

Theorem 6.10. *Assuming n_1-round unclonable tag-generation protocols and n_2-round concurrent simulation-sound proofs, there exist many-many soundness-unclonable proofs with $n_1 + n_2$ or $n_1 + n_2 - 1$ rounds.*

Proof. In a tag-generation experiment where a man-in-the-middle acts as the receiver in k left sessions and as the sender in $k + r$ right sessions, for each left session there is at most one right session with the same output $\mathsf{tag} \neq \bot$, except with negligible probability. In particular, at least r right sessions have tags which either do not belong to any left session or are $\bot$, except with negligible probability. This can be reduced to the $(1 \to 2)$ unclonability of the tag-generation by randomly selecting one of the $k = \mathrm{poly}(\lambda)$ left sessions and two of the $k + r = \mathrm{poly}(\lambda)$ right sessions.

If $k+r'$ right sessions accept, then these sessions all have tags which are not $\bot$. By the previous property, r' of these tags also do not belong to any left session. Therefore the concurrent simulation-soundness of $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ implies that the statements in these sessions belong to the NP language. □

Theorem 6.11. *Assuming n_1-round unclonable tag-generation protocols and n_2-round concurrent simulation-extractable proofs, there exist many-many SE-unclonable proofs with $n_1 + n_2$ or $n_1 + n_2 - 1$ rounds.*

Proof. Define an adversary $\mathsf{MiM}_{\mathsf{nm}}$ for the underlying non-malleable proof $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ which internally interacts with the cloning adversary MiM_U. $\mathsf{MiM}_{\mathsf{nm}}$ internally runs the tag-generation procedure in each session to determine the tags, then forwards the messages of $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ between MiM_U and the external non-malleability game. The simulator-extractor for the unclonable proof runs the simulator-extractor for the underlying non-malleable proof $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ on $\mathsf{MiM}_{\mathsf{nm}}$. Recall from the proof of Theorem 6.10 that at least r right sessions have tags which either do not belong to any left session or are $\bot$, except with negligible probability. If $k + r'$ right sessions accept, then at least r' of these tags are not $\bot$ and do not belong to any left session. The concurrent simulator-extractor for $\langle \mathsf{NMProve}, \mathsf{NMVerify} \rangle$ extracts witnesses for these sessions, except with negligible probability. □

Corollary 6.12. *Assuming post-quantum one-way functions, there exist post-quantum interactive soundness-unclonable proofs and post-quantum interactive simulation-sound proofs.*

Proof. This is immediate from Theorem 6.7, Theorem 6.6, Lemma 4.2, and the fact that post-quantum one-way functions imply simulation-sound proofs [GLM23]. Post-quantum simulation-extractable proofs are implied by post-quantum bounded-concurrent secure two party computation, which [GLM23] also constructs from post-quantum one-way functions.

□

Corollary 6.13. *Assuming simulation-extractable (respectively, simulation-sound) NIZKs and public key quantum money, there exist SE-unclonable (respectively, simulation-sound) NIZKs.*

Proof. This is immediate from Theorem 6.7, Theorem 6.6, and Lemma 4.3. Note that the round-reduction optimization can be applied to make the construction non-interactive. □

Simulation-sound NIZKs can be constructed in the common reference string (CRS) model from any NIZK using one-way functions [Sah99]. Furthermore, a simulation-sound NIZK can be compiled to be simulation-extractable using public key encryption, again in the CRS model. We observe that this construction also holds against quantum adversaries. Sahai's compiler uses many NIZKs under independent CRS's. By using strong-one-time signatures as the tag, the compiler forces any adversary to use a different set of

CRS's in the right session than in the left session, unless it copies the tag. Thus, at least one part of the right NIZK will use an honestly generated CRS, which is sound. We note that this proof also succeeds against a quantum adversary.

Furthermore, a simulation-sound NIZK can be compiled to be simulation-extractable using public key encryption, again in the CRS model. To do so, publish a public encryption key pk in the CRS, then have the prover encrypt their witness under pk and prove using the simulation-sound NIZK that it encrypted a valid witness for x. The simulator-extractor chooses the CRS, encrypts 0, simulates the left NIZK, and finally decrypts the ciphertext from the right NIZK. Simulation-soundness guarantees that the right ciphertext decrypts to a valid witness. We note that this simulator-extractor is straightline and also succeeds against a quantum adversary.

6.3 Relation to Public Key Quantum Money

As we saw in the previous section, it is possible to construct unclonable zero-knowledge proofs using public-key quantum money and a non-malleable zero-knowledge proof. If the non-malleable proof is non-interactive, then so is the resulting unclonable proof. Since NIZKs are inherently non-interactive, it is natural to wonder whether an unclonable NIZK can be used to construct public-key quantum money. We show in this section that this is indeed the case.

Theorem 6.14. *If one-way functions and simulation-extractable (respectively, simulation-sound) NIZKs exist, then the existence of SE-unclonable NIZKs (respectively, soundness-unclonable NIZKs) is equivalent to the existence of public-key quantum money.*

Proof. We prove the forward implication in Lemma 6.15 and the backward implication in Corollary 6.13 □

Lemma 6.15. *If post-quantum one-way functions and either SE-unclonable NIZKs or soundness-unclonable NIZKs exist, then public key quantum money also exists.*

Proof. We construct a quantum money mini-scheme, which can be transformed into a fully-fledged public key quantum money scheme using signatures. To generate a banknote, first honestly generate a common reference string crs for the NIZK.[5] Next, sample a post-quantum, non-interactive, statistically binding commitment $\mathsf{com} \leftarrow \mathsf{Com}(0)$.[6] Finally, sample an unclonable NIZK $|\pi\rangle$ for the language $\mathcal{L}_0$ of commitments to 0 and instance com. The serial number is com and the quantum money state is $|\pi\rangle$. $\mathsf{Verify}_{\mathsf{QM}}(\mathsf{com}, |\pi\rangle)$ outputs the result of the NIZK verification procedure $\mathsf{Verify}_{\mathsf{NIZK}}(\mathsf{com}, |\pi\rangle)$.

Consider the following hybrid experiments, where a challenger generates a quantum money state and an adversary attempts to clone it. The probability of the adversary successfully cloning in each of these hybrids is computationally close to the others.

- H_0: This is the original experiment. The experiment outputs whether the adversary successfully clones, i.e. whether $\mathsf{Verify}_{\mathsf{QM}}$ accepts both states.
- H_1: The same as the original experiment, except $|\pi\rangle$ is simulated. Indistinguishability from H_0 reduces to the zero knowledge property of the NIZK.

[5] In general, the bank can be trusted, since its only goal is to prevent the duplication of banknotes. This differs from quantum *lightning* [Zha19], where an adversary may mint its own notes and wants to produce two with the same serial number.

[6] These can be constructed in the CRS model from post-quantum one-way functions, e.g. using Naor's commitment scheme [Nao90].

- H_2: The same as H_1, except $\mathsf{com} \leftarrow \mathsf{Com}(1)$ is a commitment to 1 instead of to 0. Indistinguishability from H_1 reduces to the hiding of Com.

Say that in H_2, the adversary produces two states $|\widetilde{\pi}_1\rangle$ and $|\widetilde{\pi}_2\rangle$ such that $\mathsf{Verify}_{\mathsf{QM}}(\mathsf{com}, |\widetilde{\pi}_1\rangle)$ and $\mathsf{Verify}_{\mathsf{QM}}(\mathsf{com}, |\widetilde{\pi}_2\rangle)$ both output accept. Since the quantum money verification procedure just runs the NIZK verification procedure, $\mathsf{Verify}_{\mathsf{NIZK}}(\mathsf{com}, |\widetilde{\pi}_r\rangle)$ also accepts for both $r = 1, 2$. However, since Com is statistically binding, $\mathsf{com} \notin \mathcal{L}_0$. If the NIZK was SE-unclonable, this is a contradiction, since the simulator-extractor cannot extract a valid witness for either right session. Similarly, this is also a contradiction if the NIZK was instead soundness-unclonable, since MiM breaks soundness in both right sessions. Therefore the adversary cannot clone the quantum money state in H_2, except with $\mathrm{negl}(\lambda)$ probability. Since the cloning success probability is computationally close in H_0 and H_2, no computationally bounded adversary can clone banknotes, except with $\mathrm{negl}(\lambda)$ probability. □

As previously discussed at the end of Section 6.2, simulation-sound NIZKs can be constructed from any one-way function and any NIZK. Additionally, simulation-extractable NIZKs can be constructed from any public-key encryption scheme and any NIZK. Thus we obtain the following corollary of Theorem 6.14.

Corollary 6.16. *If one-way functions and NIZKs exist, then the existence of soundness-unclonable NIZKs is equivalent to the existence of public-key quantum money. Furthermore, if public-key encryption and NIZKs exist, then the existence of SE-unclonable NIZKs is equivalent to the existence of public-key quantum money.*

7 Strong Unclonability: Negative Results

We begin with a very simple attack ruling out strongly unclonable NIZKs.

Theorem 7.1. *There do not exist strongly SE-unclonable or soundness-unclonable NIZKs.*

Proof. As an explicit attack, MiM first forwards the left NIZK to R_1. R_1 verifies it, then returns it to MiM. By correctness of the NIZK and the Gentle Measurement Lemma [Win99], the returned state is statistically close to the original NIZK. Finally, MiM sends the returned NIZK to R_2 and R_2 verifies it. □

Interactive Zero Knowledge. Next, we rule out even *interactive* strongly unclonable proofs. At a high level, the man-in-the-middle and two right verifiers will agree to run an MPC which allows them to act as a single verifier for the left session. It is important to note that V_1 and V_2 do *not* have an authenticated channel. Instead, they must pass messages to each other through the man-in-the-middle. Since the man-in-the-middle may tamper with these messages, the MPC must be secure against a dishonest majority in order for the resulting proof to be sound.

Theorem 7.2. *Assuming stateful secure multiparty computation for quantum functionalities with dishonest majority, there do not exist strongly SE-unclonable proofs in the plain model, except for languages in BQP. Furthermore, there do not exist strongly soundness-unclonable proofs in the plain model.*[7]

[7] We note that the attack given in the proof uses computationally sound right protocols. Definition 6.2 permits any right proofs, whether computationally or statistically sound. In principle, we could design a weaker definition which only considers statistically sound right protocols, and the attack would not apply.

Remark. Commitments imply MPC with dishonest majority [BCKM21]. Statefulness can be generically and unconditionally added by the ideal functionality authenticating and encrypting messages to itself. In more detail, each time the ideal functionality is queried, the parties run a new MPC execution. In addition to their inputs for the ideal functionality, each party also inputs a classical message authentication code (MAC) key and an encryption key. A designated party inputs the authenticated previous state of the ideal functionality ($\perp$ if this is the first interaction). The ideal functionality checks the authentication of this state using each of the MAC keys, one-by-one. If every check passes, it decrypts the state using each encryption key and proceeds. After computing the query response, the ideal functionality also generates a new set of MAC and encryption keys (one per party) and authenticates/encrypts its current state. In addition to outputting the query response, it outputs the authenticated state to a designated party and outputs one MAC key and one encryption key to each party. We leave open whether strongly SE-unclonable NIZKs imply MPC.

Proof. We describe an explicit attack. Define a stateful MPC protocol between three parties P_1, P_2, and P_3, where P_3 provides input. The ideal functionality acts as the verifier in the left proof system. In other words, it interprets the input as a message from the prover in the left proof, then computes and outputs the verifier's next message according to the left proof system. At the end of the protocol, it outputs the ideal verifier's output to all parties. This is possible since the ideal verifier outputs a classical bit.

In right session 1, V_1 and MiM interact in an execution of this MPC protocol, where V_1 controls P_1 and MiM controls both P_2 and P_3. In right session 2, V_2 and MiM interact in an execution of this MPC protocol, where V_2 controls P_2 and MiM controls both P_1 and P_3. In both sessions, MiM, acting as P_3, is supposed to input messages from the left proof system into the MPC.

Claim 7.3. *Both right proof systems are computationally sound.*

Proof. This is immediate from the security of the MPC against a dishonest majority and the soundness of the left proof system. □

To carry out the attack, MiM forwards messages between the sessions. More explicitly, consider round i of both sessions. MiM receives messages $m_i^{1\to 2}$ and $m_i^{1\to 3}$ from V_1, as well as messages $m_i^{2\to 1}$ and $m_i^{2\to 3}$ from V_2. It computes messages $m_i^{3\to 1}$ and $m_i^{3\to 2}$ honestly according to the MPC protocol. Then, it sends $m_i^{2\to 1}$ and $m_i^{3\to 1}$ to V_1 and sends $m_i^{1\to 2}$ and $m_i^{3\to 2}$ to V_2.

Observe that due to the message forwarding, MiM, V_1, and V_2 are effectively participating in a single execution of the MPC protocol. Thus, whenever the ideal verifier (which is emulated by the MPC) would accept a statement $x \in \mathcal{L}$, both V_1 and V_2 will accept the same statement $x \in \mathcal{L}$.

Claim 7.4. *The left proof cannot be soundness-unclonable.*

Proof. Consider an execution where the left proof is simulated for a *false* statement. Then the right protocols both accept the same false statement, breaking soundness. □

Claim 7.5. *If the left proof is SE-unclonable for a language $\mathcal{L}$, then $\mathcal{L} \in \mathsf{BQP}$.*

Proof. To decide a statement x, run the simulator-extractor on x and MiM to obtain a witness w. Check whether w is a witness for $x \in \mathcal{L}$. If $x \in \mathcal{L}$, then both V_1 and V_2 will accept. Then, by definition of SE-unclonability, the simulator-extractor must output a valid witness for at least one of the two right sessions. Note that the simulator-extractor is still well-defined for $x \notin \mathcal{L}$, but by definition, it cannot output a valid witness. Since MiM is defined independently of x, the simulator-extractor thus decides $\mathcal{L}$. □

□

Overcoming the Impossibility for Interactive Protocols. There are several potential approaches for overcoming the impossiblity result for interactive protocols. Each of them revolves around disrupting either the correctness of the attack or the security of the MPC protocol.

- **QROM.** A dishonest-majority MPC cannot correctly implement random oracle calls, so strongly unclonable commitments/proofs may be possible in the QROM.
- **Unclonability with Respect to Statistical Soundness.** We can similarly increase the requirements of the attack by insisting that the right proofs be statistically sound. This crucially relies on the observation that MiM can tamper with the messages between V_1 and V_2, so the MPC must be secure against a dishonest majority in order for the right protocols to be sound.

We give a construction of strongly unclonable commitments in the QROM in Section 8, and leave the other directions open.

8 Strong Unclonability: Positive Results

8.1 Commitments: Strong Unclonability with Respect to Statistical Binding

Our construction is in the QROM and uses any unclonable-indistinguishable encryption scheme $(\mathsf{UGen}, \mathsf{UEnc}, \mathsf{UDec})$ with information-theoretic unclonability as a black box. Such schemes are known in the QROM, and constructing them in the plain model is an open question [AKL+22].

$\underline{\mathsf{UCom}_C(1^\lambda, m)}$**:** The sender does the following:

1. Sample an unclonable encryption key $k \leftarrow \mathsf{UGen}(1^\lambda)$.
2. Encrypt $|\psi\rangle \leftarrow \mathsf{UEnc}(k, m)$.
3. Sample $r \leftarrow \{0,1\}^\lambda$, then query the random oracle to get $h \leftarrow H(r, k)$.
4. Send $(|\psi\rangle, h)$ to the receiver.

$\underline{\mathsf{UOpen}_C(1^\lambda)}$**:** The sender sends (r, k) to the receiver.

$\underline{\mathsf{UOpen}_R(1^\lambda, (|\psi\rangle, h))}$**:** The receiver does the following:

1. Query the random oracle to get $h' \leftarrow H(r, k)$. If $h \neq h'$, output $\perp$. Otherwise continue.
2. Decrypt and output $m \leftarrow \mathsf{UDec}(k, |\psi\rangle)$.

Figure 7: A Strongly Unclonable Commitment with Respect to Statistical Binding

Theorem 8.1. *There exists a non-interactive commitment scheme in the QROM which is unclonable with respect to any (right) statistically binding commitment scheme. This holds unconditionally against any adversary which makes polynomially-many queries to the random oracle.*

Proof. The construction is given in Figure 7. Hiding follows from a hybrid argument where first $H(r,k)$ is replaced with a independently sampled uniform value h', then $|\psi\rangle$ is generated as an encryption of a fixed string. To see statistical binding, consider the extractor which queries the random oracle until it finds a value (r',k') such that $H(r',k') = H(r,k)$, then decrypts $|\psi\rangle$ using k. If H is sufficiently expanding, then with overwhelming probability there does not exist $(r',k') \neq (r,k)$ satisfying this. Thus, any honest receiver will output $\perp$ if not given $(r,k) = (r',k')$ during the opening phase, and otherwise it will attempt to decrypt $|\psi\rangle$ using k. This is exactly what the extractor does, so the joint mixed state over the transcript and message in the real opening is statistically close to the extracted version.

To see unclonability with respect to statistical binding, consider the following hybrids.

- $\mathcal{H}_0$: The unclonable commitment game with input m_0. Recall that game outputs a single bit which indicates whether the adversary successfully committed to a related message in both right sessions.
- $\mathcal{H}_1$: In the commitment phase, the random oracle is replaced with one where $H(r,k)$ is removed. In particular, for a uniformly random h', MiM and the right receivers are given access to

$$H'(x) = \begin{cases} h' & \text{if } x = (r,k) \\ H(x) & \text{otherwise} \end{cases}$$

  In the extraction/output phase, the extractors still get access to H.
- $\mathcal{H}_2$: In the commitment phase, MiM receives a uniformly random h' instead of $H(r,k)$ in the left session. Additionally, MiM and the right receivers are given access to H during the commitment phase. After the commitment phase, the extractors are given access to H'.
- $\mathcal{H}_3$: The left committer commits to m_1 instead of m_0.
- $\mathcal{H}_4$: Undo the changes in hybrids $\mathcal{H}_2$ and $\mathcal{H}_1$. This is the unclonable commitment game with input m_1.

Claim 8.2. *$\mathcal{H}_0$ and $\mathcal{H}_1$ are statistically close.*

Proof. It is sufficient to show that the state of the MiM and right receivers at the end of the commitment phase is statistically close in trace distance in both hybrids. After the commitment phase, the extractor pairs in each hybrid have unbounded query access to the same oracle H, which can be realized by simply giving them each the full description of H. In this case, both extractors are unitary operations, so the distance between their input states cannot increase.

Theorem 3.4 guarantees that if the states at the end of the commitment phase in $\mathcal{H}_0$ and $\mathcal{H}_1$ had trace distance ϵ, then (r,k) could be recovered with probability $(\epsilon/T)^2$. Since r is uniformly sampled from $\{0,1\}^\lambda$, we know that $(\epsilon/T)^2 \leq 2^{-\lambda}$. Furthermore, $T = \mathrm{poly}(\lambda)$ during the query phase, so $\epsilon = \mathrm{negl}(\lambda)$. □

Claim 8.3. $\mathcal{H}_1 = \mathcal{H}_2$

Proof. This is immediate from the definition of a random oracle. □

Claim 8.4. *If the encryption scheme is information-theoretically indistinguishable-unclonable, then $\mathcal{H}_2$ and $\mathcal{H}_3$ are statistically close.*

Proof. Say this were not the case for some MiM. Then there exists an unbounded distinguisher pair (D_1, D_2) which can simultaneously distinguish the outputs of their respective extractors between $\mathcal{H}_2$ and $\mathcal{H}_3$. We can use MiM, the right receivers, the extractors, and (D_1, D_2) to violate the information-theoretic unclonability of the encryption scheme.

The splitter receives $|\psi\rangle \leftarrow \mathsf{UEnc}(k, m_b)$, then picks a uniformly random h' and sends $(|\psi\rangle, h')$ to the adversarial receiver in the commitment phase. It outputs the states of the two right receivers at the end of the commitment phase, as well as h' and a uniformly random r.

The guessers receive k from the unclonable encryption challenger. They run their respective (inefficient) extractor on the state received from the splitter, using k, h', r, and access to H to implement H'. At the end, they run their respective distinguisher on the extracted value, then output the result. Since the only difference between the two hybrids is the choice of b, the simultaneous distinguishing advantage of the guessers is identical to simultaneous distinguishing advantage between hybrids $\mathcal{H}_2$ and $\mathcal{H}_3$. By assumption, this is noticeable. This contradicts the information-theoretic unclonability of the encryption scheme. □

Claim 8.5. *$\mathcal{H}_3$ and $\mathcal{H}_4$ are statistically close.*

Proof. The proof is exactly the same as the transitions between hybrids $\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$ in reverse. □

□

Acknowledgements

G.M. is supported by the European Research Council through an ERC Starting Grant (Grant agreement No. 101077455, ObfusQation).

J.R. is supported in part by the NSF award 1916939, DARPA SIEVE program, a gift from Ripple, a DoE NETL award, a JP Morgan Faculty Fellowship, a PNC center for financial services innovation award, and a Cylab seed funding award.

References

[Aar09] Scott Aaronson. Quantum copy-protection and quantum money. In *Proceedings of the 24th Annual IEEE Conference on Computational Complexity, CCC 2009, Paris, France, 15-18 July 2009*, pages 229–242. IEEE Computer Society, 2009.

[ABG+21] Amit Agarwal, James Bartusek, Vipul Goyal, Dakshita Khurana, and Giulio Malavolta. Post-quantum multi-party computation. In Anne Canteaut and François-Xavier Standaert, editors, *EUROCRYPT 2021, Part I*, volume 12696 of *LNCS*, pages 435–464. Springer, Heidelberg, October 2021.

[AC12] Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In Howard J. Karloff and Toniann Pitassi, editors, *44th ACM STOC*, pages 41–60. ACM Press, May 2012.

[AK21] Prabhanjan Ananth and Fatih Kaleoglu. Unclonable encryption, revisited. In Kobbi Nissim and Brent Waters, editors, *Theory of Cryptography - 19th International Conference, TCC 2021,*

Raleigh, NC, USA, November 8-11, 2021, Proceedings, Part I, volume 13042 of *Lecture Notes in Computer Science*, pages 299–329. Springer, 2021.

[AKL+22] Prabhanjan Ananth, Fatih Kaleoglu, Xingjian Li, Qipeng Liu, and Mark Zhandry. On the feasibility of unclonable encryption, and more. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 212–241. Springer, Heidelberg, August 2022.

[AKL23] Prabhanjan Ananth, Fatih Kaleoglu, and Qipeng Liu. Cloning games: A general framework for unclonable primitives. In Helena Handschuh and Anna Lysyanskaya, editors, *Advances in Cryptology - CRYPTO 2023 - 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part V*, volume 14085 of *Lecture Notes in Computer Science*, pages 66–98. Springer, 2023.

[AKN+23] Shweta Agrawal, Fuyuki Kitagawa, Ryo Nishimaki, Shota Yamada, and Takashi Yamakawa. Public key encryption with secure key leasing. In Carmit Hazay and Martijn Stam, editors, *Advances in Cryptology - EUROCRYPT 2023 - 42nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Lyon, France, April 23-27, 2023, Proceedings, Part I*, volume 14004 of *Lecture Notes in Computer Science*, pages 581–610. Springer, 2023.

[APV23] Prabhanjan Ananth, Alexander Poremba, and Vinod Vaikuntanathan. Revocable cryptography from learning with errors. Cryptology ePrint Archive, Report 2023/325, 2023. `https://eprint.iacr.org/2023/325`.

[AQY22] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. Cryptography from pseudorandom quantum states. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part I*, volume 13507 of *LNCS*, pages 208–236. Springer, Heidelberg, August 2022.

[BBBV97] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh V. Vazirani. Strengths and weaknesses of quantum computing. *SIAM J. Comput.*, 26(5):1510–1523, 1997.

[BC23] Anne Broadbent and Eric Culf. Uncloneable cryptographic primitives with interaction. *CoRR*, abs/2303.00048, 2023.

[BCKM21] James Bartusek, Andrea Coladangelo, Dakshita Khurana, and Fermi Ma. One-way functions imply secure computation in a quantum world. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part I*, volume 12825 of *LNCS*, pages 467–496, Virtual Event, August 2021. Springer, Heidelberg.

[BDF+11] Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. In Dong Hoon Lee and Xiaoyun Wang, editors, *ASIACRYPT 2011*, volume 7073 of *LNCS*, pages 41–69. Springer, Heidelberg, December 2011.

[BGG+23] James Bartusek, Sanjam Garg, Vipul Goyal, Dakshita Khurana, Giulio Malavolta, Justin Raizes, and Bhaskar Roberts. Obfuscation and outsourced computation with certified deletion. Cryptology ePrint Archive, Report 2023/265, 2023. `https://eprint.iacr.org/2023/265`.

[BL20] Anne Broadbent and Sébastien Lord. Uncloneable quantum encryption via oracles. In Steven T. Flammia, editor, *15th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2020, June 9-12, 2020, Riga, Latvia*, volume 158 of *LIPIcs*, pages 4:1–4:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2020.

[BLS22] Nir Bitansky, Huijia Lin, and Omri Shmueli. Non-malleable commitments against quantum attacks. In Orr Dunkelman and Stefan Dziembowski, editors, *Advances in Cryptology - EUROCRYPT 2022 - 41st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Trondheim, Norway, May 30 - June 3, 2022, Proceedings, Part III*, volume 13277 of *Lecture Notes in Computer Science*, pages 519–550. Springer, 2022.

[BR93] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In Dorothy E. Denning, Raymond Pyle, Ravi Ganesan, Ravi S. Sandhu, and Victoria Ashby, editors, *ACM CCS 93*, pages 62–73. ACM Press, November 1993.

[BS17] Shalev Ben-David and Or Sattath. Quantum tokens for digital signatures. Cryptology ePrint Archive, Report 2017/094, 2017. `https://eprint.iacr.org/2017/094`.

[BSW16] Mihir Bellare, Igors Stepanovs, and Brent Waters. New negative results on differing-inputs obfuscation. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT 2016, Part II*, volume 9666 of *LNCS*, pages 792–821. Springer, Heidelberg, May 2016.

[CDMW18] Seung Geol Choi, Dana Dachman-Soled, Tal Malkin, and Hoeteck Wee. A black-box construction of non-malleable encryption from semantically secure encryption. *J. Cryptol.*, 31(1):172–201, 2018.

[CLLZ21] Andrea Coladangelo, Jiahui Liu, Qipeng Liu, and Mark Zhandry. Hidden cosets and applications to unclonable cryptography. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part I*, volume 12825 of *LNCS*, pages 556–584, Virtual Event, August 2021. Springer, Heidelberg.

[DDN00] Danny Dolev, Cynthia Dwork, and Moni Naor. Nonmalleable cryptography. *SIAM J. Comput.*, 30(2):391–437, 2000.

[DIO98] Giovanni Di Crescenzo, Yuval Ishai, and Rafail Ostrovsky. Non-interactive and non-malleable commitment. In *30th ACM STOC*, pages 141–150. ACM Press, May 1998.

[DS23] Marcel Dall'Agnol and Nicholas Spooner. On the necessity of collapsing for post-quantum and quantum commitments. In Omar Fawzi and Michael Walter, editors, *18th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2023, July 24-28, 2023, Aveiro, Portugal*, volume 266 of *LIPIcs*, pages 2:1–2:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023.

[FGH+12] Edward Farhi, David Gosset, Avinatan Hassidim, Andrew Lutomirski, and Peter W. Shor. Quantum money from knots. In Shafi Goldwasser, editor, *ITCS 2012*, pages 276–289. ACM, January 2012.

[GLM23] Vipul Goyal, Xiao Liang, and Giulio Malavolta. On concurrent multi-party quantum computation. In Helena Handschuh and Anna Lysyanskaya, editors, *Advances in Cryptology - CRYPTO*

2023 - 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part V, volume 14085 of *Lecture Notes in Computer Science*, pages 129–161. Springer, 2023.

[Goy11] Vipul Goyal. Constant round non-malleable protocols using one way functions. In Lance Fortnow and Salil P. Vadhan, editors, *43rd ACM STOC*, pages 695–704. ACM Press, June 2011.

[JK23] Ruta Jawale and Dakshita Khurana. Unclonable non-interactive zero-knowledge. *IACR Cryptol. ePrint Arch.*, page 1532, 2023.

[JP14] Abhishek Jain and Omkant Pandey. Non-malleable zero knowledge: Black-box constructions and definitional relationships. In Michel Abdalla and Roberto De Prisco, editors, *SCN 14*, volume 8642 of *LNCS*, pages 435–454. Springer, Heidelberg, September 2014.

[KN22] Fuyuki Kitagawa and Ryo Nishimaki. Functional encryption with secure key leasing. In Shweta Agrawal and Dongdai Lin, editors, *Advances in Cryptology - ASIACRYPT 2022 - 28th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, December 5-9, 2022, Proceedings, Part IV*, volume 13794 of *Lecture Notes in Computer Science*, pages 569–598. Springer, 2022.

[LLQZ22] Jiahui Liu, Qipeng Liu, Luowen Qian, and Mark Zhandry. Collusion resistant copy-protection for watermarkable functionalities. In Eike Kiltz and Vinod Vaikuntanathan, editors, *Theory of Cryptography - 20th International Conference, TCC 2022, Chicago, IL, USA, November 7-10, 2022, Proceedings, Part I*, volume 13747 of *Lecture Notes in Computer Science*, pages 294–323. Springer, 2022.

[LPY23] Xiao Liang, Omkant Pandey, and Takashi Yamakawa. A new approach to post-quantum non-malleability. In *64th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2023, Santa Cruz, CA, USA, November 6 - November 9, 2023*. IEEE, 2023.

[Nao90] Moni Naor. Bit commitment using pseudo-randomness. In Gilles Brassard, editor, *CRYPTO'89*, volume 435 of *LNCS*, pages 128–136. Springer, Heidelberg, August 1990.

[PR05a] Rafael Pass and Alon Rosen. Concurrent non-malleable commitments. In *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2005), 23-25 October 2005, Pittsburgh, PA, USA, Proceedings*, pages 563–572. IEEE Computer Society, 2005.

[PR05b] Rafael Pass and Alon Rosen. New and improved constructions of non-malleable cryptographic protocols. In Harold N. Gabow and Ronald Fagin, editors, *37th ACM STOC*, pages 533–542. ACM Press, May 2005.

[Sah99] Amit Sahai. Non-malleable non-interactive zero knowledge and adaptive chosen-ciphertext security. In *40th FOCS*, pages 543–553. IEEE Computer Society Press, October 1999.

[TFKW13] Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New Journal of Physics*, 15(10):103002, oct 2013.

[Unr14] Dominique Unruh. Revocable quantum timed-release encryption. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 129–146. Springer, Heidelberg, May 2014.

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

[Wie83] Stephen Wiesner. Conjugate coding. *SIGACT News*, 15(1):78–88, jan 1983.

[Win99] Andreas J. Winter. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory*, 45(7):2481–2485, 1999.

[WZ82] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

[Yan22] Jun Yan. General properties of quantum bit commitments (extended abstract). In Shweta Agrawal and Dongdai Lin, editors, *Advances in Cryptology - ASIACRYPT 2022 - 28th International Conference on the Theory and Application of Cryptology and Information Security, Taipei, Taiwan, December 5-9, 2022, Proceedings, Part IV*, volume 13794 of *Lecture Notes in Computer Science*, pages 628–657. Springer, 2022.

[Zha19] Mark Zhandry. Quantum lightning never strikes the same state twice. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part III*, volume 11478 of *LNCS*, pages 408–438. Springer, Heidelberg, May 2019.

[Zha21] Mark Zhandry. White box traitor tracing. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part IV*, volume 12828 of *LNCS*, pages 303–333, Virtual Event, August 2021. Springer, Heidelberg.