Post-quantum Resettably-Sound Zero Knowledge*

Nir Bitansky† Michael Kellner‡ Omri Shmueli§

March 2021

Abstract

We study post-quantum zero-knowledge (classical) protocols that are sound against *quantum resetting attacks*. Our model is inspired by the classical model of resetting provers (Barak-Goldreich-Goldwasser-Lindell, FOCS '01), providing a malicious efficient prover with oracle access to the verifier's *next-message-function*, fixed to some initial random tape; thereby allowing it to effectively reset (or equivalently, rewind) the verifier. In our model, the prover has *quantum access* to the verifier's function, and in particular can query it in superposition.

The motivation behind quantum resettable soundness is twofold: First, ensuring a strong security guarantee in scenarios where quantum resetting may be possible (e.g., smart cards, or virtual machines). Second, drawing intuition from the classical setting, we hope to improve our understanding of basic questions regarding post-quantum zero knowledge.

We prove the following results:

- **Black-Box Barriers.** Quantum resetting exactly captures the power of black-box zero knowledge quantum simulators. Accordingly, resettable soundness cannot be achieved in conjunction with black-box zero knowledge, except for languages in **BQP**. Leveraging this, we prove that constant-round public-coin, or three message, protocols cannot be black-box post-quantum zero-knowledge. For this, we show how to transform such protocols into quantumly resettably sound ones. The transformations are similar to classical ones, but their analysis is significantly more challenging due to the essential difference between classical and quantum resetting.
- **A Resettably-Sound Non-Black-Box Zero-Knowledge Protocol.** Under the (quantum) Learning with Errors assumption and quantum fully-homomorphic encryption, we construct a post-quantum resettably-sound zero knowledge protocol for **NP**. We rely on non-black-box simulation techniques, thus overcoming the black-box barrier for such protocols.
- **From Resettable Soundness to The Impossibility of Quantum Obfuscation.** Assuming one-way functions, we prove that any quantumly-resettably-sound zero-knowledge protocol for **NP** implies the impossibility of quantum obfuscation. Combined with the above result, this gives an alternative proof to several recent results on quantum unobfuscatability.

*This work was supported by European Union Horizon 2020 Research and Innovation Program via ERC Project REACT (Grant 756482), by ISF grants 18/484 and 19/2137, by Len Blavatnik and the Blavatnik Family Foundation, and the Blavatnik Interdisciplinary Cyber Research Center at Tel Aviv University.

†Tel-Aviv University, E-mail: nirbitan@tau.ac.il. Member of the Check Point Institute of Information Security.

‡Tel-Aviv University, E-mail: kellner@mail.tau.ac.il.

§Tel-Aviv University, E-mail: omrishmueli@mail.tau.ac.il.

Contents

1 Introduction

Zero-knowledge protocols, introduced by Goldwasser, Micali, and Rackoff [GMR89], are a cornerstone of cryptography. They allow proving the validity of any statement in **NP** without revealing anything but its validity [GMW91]. After over three and a half decades of research, zero knowledge protocols are well understood in terms of their expressiveness and round complexity, and various enhancements of zero knowledge have been considered.

In this work, we consider zero knowledge protocols with *post-quantum security*, namely, protocols that can be executed by classical parties, but where both soundness and zero knowledge are guaranteed against efficient quantum adversaries. Here, starting from the seminal work of Watrous [Wat09], our understanding of post-quantum zero knowledge has been gradually improving, and yet it is still far behind our understanding of classical zero knowledge. Indeed, beyond the obvious need for post-quantum computational assumptions, the design and analysis of post-quantum zero knowledge protocols is challenged by quantum phenomena such as the no-cloning theorem [WZ82] and state disturbance [FP96], which often deem classical techniques insufficient.

Resettable Soundness. We focus on the notion of *resettable soundness*, introduced by Barak, Goldreich, Goldwasser, and Lindell [BGGL01] and by Micali and Reyzin [MR01a]. In the classical setting, resettably-sound protocols remain sound even against a prover that has the ability to reset the honest verifier to its initial state and random tape, and repeat the interaction in any way it chooses (equivalent to the ability to rewind the verifier to any previous message). The threat of reset attacks arises in various settings, when fresh randomness cannot be generated on the fly and parties are subject to physical resets. Examples include verifiers that run on smart cards or virtual machines, or when a verifier interacts with arbitrarily many provers, but wishes to use the same randomness and keep a small, constant-size inner state that does not depend on the number of interactions. Accordingly security against resetting attacks has received much attention [CGGM00, KP01, MR01b, DGS09, GS09a][COSV12, OV12, COPV13, COP+14, BP15, CPS16].

Beyond the protection it provides in the above settings, resettable soundness has played an important role in understanding a foundational question regarding (classical) zero knowledge protocols — the gap between black box zero knowledge and non black box zero knowledge. In the first, the zero knowledge simulator can only access the verifier as a black box, whereas in the second, it can make explicit use of the verifier's code. Indeed, resettably-sound protocols cannot have a black-box zero knowledge simulator [BGGL01]; roughly speaking, this is because a resetting prover effectively has the same rewinding power as a zero knowledge simulator, and can accordingly use any black box simulation strategy in order to cheat. In fact, several other black-box zero knowledge impossibilities can be derived by a reduction to the impossibility of resettably sound black-box zero knowledge [GK96b, BGGL01, PTW11].

This Work: Quantum Resettable Soundness. We investigate resettable soundness in the quantum setting. That is, we consider classical protocols that are sound against *quantum resetting attacks* and (plain) zero knowledge against quantum malicious verifiers. Our goal is twofold: First, constructing such protocols to deal with resetting scenarios in a quantum world. Second, in light of the role that resettable soundness plays in the classical setting, we may expect that in the quantum setting too, understanding resettable soundness would shed light on basic questions regarding post-quantum zero knowledge.

1.1 Contributions

We first model resetting attack in a quantum world and define the corresponding notion of resettable soundness. We consider a strong definition that provides the resetting prover *quantum access* to the honest verifier's *next message function*, for some fixed verifier randomness. In particular, the resetting prover may not only rewind the verifier, but also do it in superposition. This model aims to capture the worst possible behavior of an efficient quantum attacker in a setting where resetting is possible. Furthermore, the model aims to capture the capabilities of a black box zero knowledge simulator in the quantum setting. Throughout, we restrict attention to efficient resetting provers and accordingly to arguments [BCC88] (offering computational soundness) rather than proofs (offering statistical soundness).

We next describe our results regarding the construction and implications of the above notion of resettable soundness (further discussion of the model and definition can be found in the technical overview below).

Quantum Black Box Barriers. As intended our definition provides a quantum resetting prover with the power of a quantum black-box zero knowledge simulator. This yields a black box barrier analogous to the one in the classical setting.

Observation 1.1 (Informal)**.** *Post-quantum resettably-sound* **black-box** *zero knowledge is impossible, except for languages in* **BQP**.

Building on this fact, we then prove that the Goldreich-Krawczyk black box zero knowledge barriers from the classical setting [GK96b] translate to the quantum setting. More generally, we show that under minimal assumptions, any *three-message* or *constant-round public-coin* zero-knowledge protocol can be converted into a quantum resettably-sound argument, while preserving black-box zero knowledge.

Theorem 1.2 (Informal)**.** *Assuming post-quantum one-way functions, post-quantum zero knowledge protocols that are* **three message or constant-round public-coin**, *with a negligible soundness error, can be made resettably sound. Such protocols cannot be black-box zero knowledge, except for languages in* **BQP**.

The transformation behind the theorem is in fact the same as the corresponding classical transformation [BGGL01]. The analysis, however, is different and more challenging due to the essential difference between classical resetting and quantum resetting, which is *superposition resetting attacks* (see technical overview).

The resulting black-box barrier holds for general zero knowledge protocols, in particular, for arguments. In the case of proofs, there is evidence that three-message or constant-round public-coin zero knowledge (for non-trivial languages) is impossible altogether (even non-black-box) [BLV06, KRR17, FGJ18]. In the case of black-box zero knowledge, this was proven (unconditionally) Jain, Kolla, Midrijanis, and Reichardt [JKMR09]. Finally, we note that like in the classical setting, the resulting barriers, in fact, hold also in a semi-black-box model where the simulator is allowed to depend on the circuit size of the simulated verifier. In the fully black-box model, the barriers can be proven without relying on one way functions.

A Resettably-Sound Protocol via Quantum Non-Black-Box Techniques. Aiming to constructing post-quantum resettably-sound zero knowledge, we are faced with the above mentioned black-box impossibility. In the classical setting, the corresponding black box impossibility of resettably-sound can be circumvented relying on *non-black-box simulation.* Indeed, the pioneering work of Barak shows how to construct constant-round public-coin zero knowledge arguments from collision-resistant hashing [Bar01], to which one can apply the [BGGL01] transformation to obtain resettable soundness. In the quantum setting, however, constant-round public-coin zero knowledge arguments for now remain out of reach.

Nevertheless, under standard assumptions (Quantum Learning with Errors [Reg05] and Quantum Fully-Homomorphic Encryption [Bra18a, Mah18a]) we construct a post-quantum resettably-sound zero knowledge protocol relying on (quantum) non-black-box simulation.

Theorem 1.3 (Informal)**.** *Assuming the hardness of* QLWE *and the existence of* QFHE *there exists a post-quantum resettably sound zero-knowledge argument for* **NP**.

Our construction starts from the recent construction of post-quantum constant-round (non-black-box) zero-knowledge [BS20] and modifies it. While non-black-box techniques do not seem inherent for constant round zero knowledge with plain soundness (see [CCY20] in related work), in our setting they become essential. While the non-black-box technique we use is similar to that of [BS20], resettable soundness, requires a new proof, which encounters several technical challenges emerging from quantum resetting.

From Resettable Soundness to Quantumly Unobfuscatable Functions. In the classical setting, resettably-sound zero knowledge is known to be intimately related to the impossibility of virtual black box obfuscation [BGI+12]. In particular, assuming one-way functions any resettably-sound zero knowledge protocol for **NP** implies *a family of unobfuscatable functions* [BP15]. We show that this result translates also to the quantum setting; specifically there exists classical function families that cannot be obfuscated as quantum states according to the quantum virtual black box notion of Alagic and Fefferman [AF16].

Theorem 1.4 (Informal)**.** *If there exists a post-quantum resettably-sound zero-knowledge argument for* **NP** *and post-quantum one-way functions, then quantum virtual black-box obfuscation is impossible.*

Such an impossibility was recently shown by Ananth and La Placa [AP20] and by Alagic, Brakerski, Dulek, and Schaffner [ABDS20]. The combination of Theorems 1.3,1.4 yields an alternative, albeit more complicated, proof of this result (under similar assumptions). We note that differently from the classical setting where the impossibility of black box obfuscation is unconditional, in the quantum setting it relies on QLWE and strongly relies on quantum homomorphic encryption. Following the above theorem, advancement in the construction of quantumly resettably sound protocols, and in particular the construction of constant-round public-coin or three-message protocols, is likely to also advance our understanding of quantum unobfuscatability.

2 Technical Overview

In this section, we provide a technical overview of the paper.

2.1 Defining Post-quantum Resettable Soundness

In the classical setting [BGGL01], a resetting attack by a malicious prover rP is modeled by providing the prover oracle access to the *next-message function* of honest verifier $\mathsf{V}(x, \cdot\,; r)$ for the common input x and randomness r that is sampled uniformly and fixed once and for all. The prover then has the ability to query a partial transcript ts, including prover messages up to some round i, and obtain back the verifier message in round $i+1$. In a successful attack, after polynomially many queries, the prover manages to output a full transcript ts for some false statement x, which yet convinces the verifier $\mathsf{V}(x, \mathsf{ts}; r)$.

Aiming to generalize this to the quantum setting, there are two conceivable definitions. The first considers quantum provers, which are only given *classical access* to $\mathsf{V}(x, \cdot\,; r)$. The second, which we consider in this work, provides the prover with *quantum access* to $\mathsf{V}(x, \cdot\,; r)$; namely, access to the unitary map $|\mathsf{ts}\rangle|y\rangle \mapsto |\mathsf{ts}\rangle|y \oplus \mathsf{V}(x, \mathsf{ts}; r)\rangle$; in particular, it may now query $\mathsf{V}(x, \cdot\,; r)$ in superposition. While the first may still provide meaningful security in settings where classical access can be enforced, the second resists stronger resetting scenarios in which the attacker can perform quantum resetting. Furthermore, our definition captures the abilities of a black-box zero-knowledge simulator, and will thus be useful for proving black-box barriers on post-quantum zero knowledge.

Proving that resettably-sound protocols cannot be *black box* zero knowledge, except for languages in **BQP**, now follows a standard argument similar to the classical one [BGGL01]. Roughly, speaking this is because a quantum resetting prover has the ability to run a quantum black-box simulator for the verifier $\mathsf{V}(x, \cdot\,; r)$, in order to produce a cheating transcript. Indeed, by zero knowledge and completeness, for any true statement x, the simulator almost always generates an accepting transcript, and unless it can decide the underlying language (meaning that it is in **BQP**), it must also be able to do so for some false statements.

Variants. A natural strengthening of the above definition allows the prover to also choose the statements x that it provides the oracle with; namely get access to $\mathsf{V}(\cdot\,, \cdot\,; r)$. In the body, we prove that this stronger notion can be obtained from the simpler notion assuming subexponentially-secure (post-quantum) pseudorandom functions. We note that all the implications of resettable soundness shown in this work, already follow from the simpler notion of resettable soundness.

Also, as already noted we restrict attention to efficient resetting provers, namely arguments. We note that classically, resettably-sound zero knowledge proofs, namely against unbounded provers, are only possible for trivial languages [BGGL01], and this carries over to the quantum setting. Again, all implications shown in this work already follow from resettably-sound zero knowledge arguments.

2.2 3-Message and Constant-Round-Public-Coin Protocols Can be Made Resettably Sound

We now explain how 3-message protocols and constant-round public-coin protocols are made resettably sound. The transformation does not change the honest prover, and thus preserves black box zero knowledge, and any other privacy guarantee, such as witness indistinguishability (which we will use later on). This in

turn yields quantum black-box zero-knowledge barriers on 3-message or constant-round public-coin protocols (with a negligible soundness error).

3-Message Protocols. The transformation for three-message protocols is essentially identical to the classical one [BGGL01]. Given the original verifier V for the protocol, we consider a new verifier $\tilde{\mathsf{V}}$ whose randomness consists of a random seed k for a pseudorandom function secure under quantum access [Zha12]. Given a statement x and first prover message α, the verifier $\tilde{\mathsf{V}}$ derives randomness r by applying the PRF and derives the second message β, by applying the original verifier with corresponding randomness:

$$r = \mathsf{PRF}_k(\alpha), \qquad \beta = \mathsf{V}(x, \alpha; r) \ .$$

As expected $\tilde{\mathsf{V}}(x, \alpha, \beta, \gamma; k)$ accepts if the original verifier $\mathsf{V}(x, \alpha, \beta, \gamma; r)$ accepts.

In the classical setting, resettable soundness is proven by a relatively simple reduction to the soundness of the original protocol. In the quantum setting, however, proving security is significantly more challenging. Before we address these challenges let us start by recalling the classical reduction to develop basic intuition. We are given a resetting prover rP, which without loss of generality, never makes the same query twice, and always queries the oracle $\tilde{\mathsf{V}}$ on the cheating transcript it eventually outputs. Roughly speaking, the reduction, which aims to cheat V in a single interaction, will aim to embed this interaction in a random position in an execution of the resetting $\mathsf{rP}^{\tilde{\mathsf{V}}(x,\cdot\,;k)}$ and forward that execution to the external verifier V. All other executions are internally simulated by the reduction. By pseudorandomness, the view of the simulated rP is indistinguishable from its view in a resetting attack and will include some cheating execution. With noticeable probability (inverse proportional to the number of queries that rP makes), the reduction hits the cheating execution and wins.

In the quantum setting, however, it is not a-priori clear how such a reduction would work. In particular, any query made by rP to $\tilde{\mathsf{V}}$ may now include a super-position of super-polynomially many transcripts. Furthermore, merely observing the prover queries disrupts its state and could affect the probability it produces a cheating transcript. Embedding an execution at a random position is also tricky. When we forward some message α to the external verifier, and obtain back a message β, we have to answer consistently with β all oracle queries to α. However, whereas in the classical case, we could assume that no α is queried more than once (because queries can be stored), now it may be that α takes part in all superposition queries that the prover makes.

Similar difficulties arise when trying to prove the soundness of the Fiat-Shamir transformation [FS86] in the quantum random oracle model [BDF+10], and were, in fact, successfully circumvented in recent works [LZ19, DFMS19, DFM20]. Indeed, both in the Fiat-Shamir setting and in our setting, we can still hope to obtain an analog of the classical reduction. Specifically, by measuring a random query made by rP, forwarding the result α to the external verifier, and consistently answering with β any *future query* α by *reprogramming* the classical function $\tilde{\mathsf{V}}$.

The intuition is that for the prover to succeed in outputting a convincing transcript (α, β, γ), the message α has to appear in one of his super-position queries with noticeable weight; otherwise, it gains almost no information on the corresponding verifier message β, and will fail to break soundness. Furthermore, when measuring such a query we are likely to obtain α, without disturbing the prover's state too much (in the extreme case that α occurs with probability one, the state is not disturbed at all). If the reduction hits the first such query (where α is significant), then it suffices that it is consistent with α in future queries and does not have to worry about past queries.

This intuition is elegantly captured and made rigorous by Don, Fehr, Majenz, and Schaffner [DFMS19, DFM20]. They prove reprogramming and simulation lemmas that establish the validity of (a slight variant of) the described reduction in the case of Fiat Shamir, where the message β is chosen uniformly at random. In our setting, β is an arbitrary message derived by the verifier. Nevertheless, relying on their reprogramming lemma, we can prove an appropriate simulation lemma for our setting.

A Useful Generalization: Many-Round Almost Resettable Protocols. We also show a generalization of the three-message transformation that allows to take any *single-prefix resettably-sound protocol* and make it (fully) resettably sound. Single-prefix resettably sound protocols are almost resettably sound. They

allow the resetting prover to use a *single classical first message* and accordingly obtain a single response to this message from the verifier. Only starting from the prover's next message it is allowed to quantumly reset; namely all interactions (even if in super-position) start with the same classical prover message and verifier response. A three message protocol is indeed the simplest example of a single-prefix resettably-sound protocol, since the verifier has a single message, and if this message is not reset, then there is no resetting whatsoever, and resettable soundness is synonymous to plain soundness.

This generalization turns out to be useful, and is used later on in our construction of a resettably sound (non-black-box) zero knowledge protocol for **NP**. To obtain this generalization, we first extend the reprogramming lemma from [DFM20] to the case of reprogramming an entire oracle, specified by some prefix. This allows us to extend the previously described reduction, which given a fully resetting prover can turn it into a single prefix resetting prover. The difference is that now rather than obtaining from the external verifier a response β to the measured α, it obtains oracle access to an oracle $\tilde{\mathsf{V}}(x, \alpha, \cdot\,; r)$ specified by the prefix α (and implicitly a response β). This oracle effectively allows to perform resetting attacks, but only starting from the next prover message.

Constant-Round Public-Coin Protocols. Another example where classical resettable soundness can be achieved is that of constant round public-coin protocols. Also here we obtain an analogous transformation in the quantum setting, now based on *multi-value reprogramming lemmas* from [DFM20], used there to deal with multi-message Fiat Shamir.

Beyond 3-Message or Constant-Round Public-Coin? We note that we should not hope to transform arbitrary protocols into resettably-sound ones; indeed, multi-message post-quantum zero knowledge protocols for **NP** do exist, and are even public coin [Wat09]. But what does it take for a protocol to be (transformable to) resettably sound? Here one bottleneck is the (in)ability of the reduction to simulate internally the interactions that are not forwarded to the external verifier. More specifically, the question is whether the reduction could simulate *continuations* that start consistently with the external verifier and than diverge. In general private-coin protocols, this may not be possible as the private coins of the external verifier are not known to the reduction. In contrast, in three-message protocols this is not a problem, as there is nothing to continue (the verifier has a single message). Similarly, also in public coin protocols, simulating continuations is easy — the reduction samples the random messages on its own.

This is, however, not the only bottleneck. A second bottleneck is that the reduction has to *hit the cheating execution* with noticeable probability, and since the reduction has to guess on the fly which messages to forward to the external verifier, this probability may decrease exponentially in the number of rounds. Hence, even for public coin protocols, the transformation only works for a constant number of rounds. In fact, this is tight — the round complexity of Watrous' zero knowledge public-coin proofs [Wat09] can be reduced to any super constant function $\omega(1)$. (For instance, by starting from Blum's Hamiltonicity protocol [Blu86] that has constant soundness, repeating it in parallel logarithmically many times, and then sequentially $\omega(1)$ times.)

2.3 Constructing a Resettably Sound Non-Black-Box Zero-Knowledge Protocol

We now outline the main ideas and techniques behind our construction of a resettably-sound non-black-box zero-knowledge protocol for **NP**. Our starting point is the post-quantum zero knowledge protocol of Bitansky and Shmueli [BS20]. We next describe the main challenges in turning this protocol into a quantumly resettably sound protocol.

A Bird's Eye View of the BS Protocol. At a high level (and oversimplifying), the BS protocol consists of two phases. First, the verifier provides a quantum extractable commitment to a challenge message. Then the parties execute a standard zero knowledge sigma protocol to prove the statement x, where the verifier opens the commitment from the first phase. The extractor for the first-phase commitment is non-black-box, using the code of a sender (the verifier in this case), it can extract the underlying message while faithfully simulating the quantum state of the sender. This gives rise to a corresponding non-black-box simulation strategy, which first extracts the verifier challenge and can then cheat in the sigma protocol.

Already at this level, one can see that the protocol is not resettably sound, even classically, let alone quantumly. A resetting prover can first run the verifier until the opening phase, obtain the challenge, then reset the verifier, and like the simulator use the obtained challenge to cheat in the sigma protocol. Indeed, the reason that the actual simulator in the BS protocol does not follow this black-box strategy is that it does not work for malicious quantum verifiers, whereas a resetting prover only has to cheat a classical verifier.

Following the above observation, we change the above high level blueprint. We rely on the Feige-Lapidot-Shamir [FLS99] *trapdoor paradigm.* In the first-phase, the BS extractable commitment is used to set up a trapdoor statement t. In the second phase, the prover provides a witness-indistinguishable proof that either x is a true statement or t is a true statement. To guarantee soundness, the trapdoor statement is set up so that it is indistinguishable from a false statement, and thus relying on the soundness of the second-phase proof, a convincing proof must mean that x is a true statement. In contrast, a simulator given the code of the verifier should be able to efficiently extract a witness for the trapdoor statement t, and can then use it in the second phase proof indistinguishably from the prover (who uses the witness for x).

Given that we are interested in quantum resettable soundness, we have to guarantee that the indistinguishability of the trapdoor statement t from a false statement, holds even against quantum resetting attacks. Furthermore, we have to guarantee that the second-phase proof is resettably sound. For the latter, we can use standard constant-round public-coin witness-indistinguishable proofs; indeed, we have already shown that such proofs can be made quantumly-resettably sound, while preserving witness indistinguishability. The more involved part is establishing indistinguishability of the trapdoor statement from a false one under resetting.

A Resettably-Secure Trapdoor Phase. We now dive deeper into the construction of a resettably-secure trapdoor phase. In terms of extractability (of a trapdoor witness), we first present a trapdoor phase that is only extractable against a restricted class of verifiers that are *non-aborting and explainable*. The notion of non-aborting explainable verifiers considers verifiers whose messages can always be *explained* as a behavior of the honest (classical) verifier with respect to *some* randomness (finding this explanation may be inefficient); in particular, they never abort. This simpler setting will already capture the main challenges we need to deal with. We will later discuss how this restriction is removed.

Similarly to the BS extractable commitment, we rely on three basic tools:

- *Quantum fully-homomorphic encryption* (QFHE) — an encryption scheme that allows to homomorphically apply any polynomial-size quantum circuit C to an encryption of x to obtain a new encryption of $C(x)$, proportional in size to the result $|C(x)|$ (the size requirement is known as *compactness*).

- *Compute-and-compare program obfuscation* (CCO). A compute-and-compare program $\mathbf{CC}[f, v, z]$ is given by a function f (represented as a classical circuit) and a target string v in its range; it accepts every input x such that $f(x) = v$, and rejects all other inputs. A corresponding obfuscator compiles any such program into a program $\widetilde{\mathbf{CC}}$ with the same functionality. In terms of security, provided that the target v has high entropy conditioned on f, the obfuscated program is computationally indistinguishable from a simulated dummy program that is independent of f, v, z, and rejects *all* inputs.

- *Secure function evaluation* (SFE) that can be thought of as homomorphic encryption with an additional *circuit privacy* guarantee, which says that the result of homomorphic evaluation of a circuit, reveals nothing about the evaluated circuit to the decryptor, except of course from the result of evaluation.

We now describe a (still simplified) trapdoor phase, which is essentially the same as the BS extractable commitment, except for how the randomness of the verifier is handled. In the trapdoor phase the verifier has two randomized steps; we denote the randomness used in these rounds by r_1 and r_2, respectively.

1. The prover P samples a secret key sk for SFE, and sends a commitment cmt to sk.

2. The verifier V uses randomness r_1 to sample:

    - two random strings u and v,

- a secret key sk' for an FHE scheme,
- an FHE encryption $\mathsf{ct}'_u = \mathsf{QFHE.Enc}_{\mathsf{sk}'}(u)$ of u,
- an obfuscation $\widetilde{\mathbf{CC}}$ of $\mathbf{CC}[f, v, \mathsf{sk}']$, where $f = \mathsf{QFHE.Dec}_{\mathsf{sk}'}$ is the FHE decryption circuit.

It then sends $(\mathsf{ct}'_u, \widetilde{\mathbf{CC}})$ to the prover P.

3. The prover P:
   - sends $\mathsf{ct}_{u'}$, a string u' encrypted using SFE (the honest prover sets u' arbitrarily).
   - proves using a resettably-sound witness-indistinguishable argument that $\mathsf{ct}_{u'}$ is a valid SFE encryption corresponding to the secret key sk underlying the commitment cmt.
4. The verifier V:
   - uses he SFE homomorphic evaluation to compute the function $C_{u\to v}$ that given input u, returns v (and otherwise $\perp$).
   - To derive the randomness for this evaluation, V interprets its randomness r_2 as a seed for a pseudorandom function and applies it to the prover messages $(\mathsf{cmt}, \mathsf{ct}_{u'})$.
   - V then returns the resulting ciphertext to P.
5. The trapdoor statement t is set to be:

"There exists a ciphertext ct^* *that the program* $\widetilde{\mathbf{CC}}$ *does not reject."*

Basic Intuition. We start by building basic intuition on how the above protocol achieves the goal of a trapdoor phase. For starters we will ignore the resetting attacks, and recall the intuition from BS. Then we will address the main challenges in proving resettable security, and how they are met. (A reader familiar with BS may want to skip directly to the resettable security paragraph.)

Let us start by explaining how a non-black-box simulator can use the circuit of an explainable verifier in order to obtain a witness proving the trapdoor statement. The simulator acts honestly in the first step, and then obtains the CC obfuscation $\widetilde{\mathbf{CC}}$ and FHE encryption ct'_u of the string u. The main point is that now the simulator can *homomorphically continue the protocol under the FHE encryption.* That is, it will evaluate the (quantum) verifier under the encryption, where is has the secret u *in the clear* and can use it in the SFE protocol to obtain back the secret target value v (the hiding of SFE encryption is used to argue that such an execution is indistinguishable from a real one where a dummy encryption is sent). Going back out of the encryption, the simulator now actually holds an encryption ct^* of v, and in particular $\widetilde{\mathbf{CC}}$ does not reject ct^*, but rather outputs the FHE secret key sk'. Thus, the ciphertext ct^* obtained by the simulator is a valid trapdoor witness. The reason we require $\widetilde{\mathbf{CC}}$ to output sk', rather than an arbitrary accept value, is for the simulator to be able to decrypt the internal verifier quantum state and faithfully continue the simulation.

We now turn to explain why to a malicious (but for now, non-resetting) prover, who does not obtain the code of the verifier, the trapdoor statement is indistinguishable from a false statement. Specifically, we would like to argue that we can replace the obfuscation $\widetilde{\mathbf{CC}}$ with a simulated one that rejects all inputs. To see this, we first argue that the prover cannot send an SFE encryption $\mathsf{ct}_{u'}$ such that $u' = u$, except with negligible probability. Indeed, given only the first sender message $(\mathsf{ct}'_u, \widetilde{\mathbf{CC}})$, the receiver obtains no information about u. Hence, we can invoke the CCO security and replace the obfuscation $\widetilde{\mathbf{CC}}$ with a simulated one, which is independent of the secret FHE key sk. This, in turn, allows us to invoke the security of encryption to argue that the first message $(\mathsf{ct}'_u, \widetilde{\mathbf{CC}})$ hides u. While this means that the prover does not obtain u in the clear, we still need to argue that it cannot send an encryption of u. This is done using a non-uniform reduction and is exactly the purpose of the prover commitment cmt to the SFE secret key sk, which allows us to provide the reduction with sk as non-uniform advice. Having established that no SFE encryption of u is sent we can

invoke the circuit privacy guarantee to completely remove the value v from the prover's view and now we also replace $\widetilde{\mathbf{CC}}$ with a simulated one that rejects all inputs.

Resettable Security. The above argument establishing indistinguishability of the trapdoor statement from a false statement, does not consider resettable attackers. We now discuss the difficulties arising from resetting attacks and how they are dealt with.

Recall that a resetting quantum attacker may perform super-position queries. Accordingly, now when arguing that it cannot produce an SFE encryption of u, we would like to argue that SFE encryptions of u have negligible weight in any query made by rP; in other words, projecting the queries on the space of non-u queries has little affect on the experiment. Indeed, we can prove this if the resetting prover is guaranteed to always use the same SFE encryption key, in which case we can non-uniformly hardwire this key into our reduction like before. The problem is that a resetting prover may start many executions, each with a different SFE key; in fact it can run exponentially many such executions in super-position. This is where we use our reduction to *single-prefix resetting provers* (discussed in the previous section). The reduction allows us to obtain new prover that in all executions sends the same commitment cmt and uses the same secret key; any resetting attempt is done from the next message and onward.

Having established that the prover queries do not include encryptions of the secret u (or rather have a small projection on this space), we would like to invoke as before the circuit privacy guarantee. However, this should be done with care. The problem is the prover still has the ability to send many ciphertexts and receive evaluations on each one of them. This is the reason we invoke a pseudorandom function to derive randomness in this step, which ensures that each evaluation uses (pseudo)independent randomness. Proving security, however, is not straightforward. In the classical setting, this is not an issue — the overall number of queries is polynomial and thus we can use a standard hybrid argument, invoking circuit privacy polynomially many times. In the quantum setting, however, where queries include a super-position over exponentially many ciphertexts, this is unclear. In fact, there is a basic problem here, which we find interesting on its own. Assume that for two efficient samplers $S_0(x)$ is computationally indistinguishable from $S_1(x)$ for any input x; are the two oracles $F_i(x) := S_i(x; R(x))$ indistinguishable (quantumly), when R is a random function? Zhandry [Zha12] shows that this is the case if $S_i(x) = S_i(y)$ for any x, y, but the general case is unclear.

Fortunately, in our case, we can take a straightforward approach to solve it, by guaranteeing that circuit privacy is statistical, and ensuring that the statistical error is smaller than the total number of ciphertexts in the support, and thus a naive hybrid argument still works. Doing so again requires care, as the size of SFE ciphertexts and the statistical security guaranteed may be related. We show how to deal with this by forcing the prover to also commit to the randomness used in SFE encryptions so that the number of hybrids only depends (exponentially) on the fixed length of the encrypted plaintext.

General Verifiers. In the described trapdoor protocol, we have made two simplifying assumptions regarding the verifier — that it is explainable and that it is non-aborting. We deal with the first restriction using a common approach based on witness indistinguishable proofs by the verifier [BKP19, BS20]. This time however, we need to rely on *resettable* statistical witness indistinguishability. Statistically-witness-indistinguishable ZAPs are known under super-polynomial hardness of QLWE [GJJM20, BFJ+20] and are resettable as they only include one round. We also give a solution using only polynomial hardness of QLWE, based on Unruh's notion of collapse binding statistically-hiding hash functions, which leads to statistical witness-indistinguishable protocols [Unr16b, Unr16a], while these protocols are not resettably-witness-indistinguishable as is, we show how to make them resettably secure.

As for dealing with verifier aborts, we rely on a general approach from [BS20], which roughly asserts that it is sufficient to be able to construct two separate zero knowledge simulators, one for verifiers that do not abort and one for verifiers that do, and which do not affect the probability of aborting (more than negligibly). They show that two such simulators can always be combined to one full-fledged simulator using Watrous' rewinding lemma [Wat09].

2.4 From Resettable Soundness to Quantum Unobfuscatability

Finally, we outline the construction of quantumly unobfuscatable functions from resettably-sound zero-knowledge protocols for **NP** and one-way functions. Informally, an unobfuscatable function family is a family of classical functions $\{f_k\}$ indexed by a secret k. Given quantum oracle access to a random f_k in the family, no efficient quantum learner should be able to learn some secret function $s(k)$ of the key. In contrast, given any quantum state ρ and quantum circuit C such that for some k and and all inputs x, $C(\rho, x)$ computes the classical value $f_k(x)$, one could efficiently extract from C and ρ the corresponding secret $s(k)$.

Our construction closely follows the construction of classically unobfuscatable functions from classical resettably sound zero knowledge protocols [BP15], while making some adaptations to the analysis stemming from the difference between the classical and quantum settings. Roughly speaking, our family of functions $\{f_{r,\varphi,s}\}$ is indexed by randomness r and statement φ for the (honest) verifier given by our resettably-sound protocol, and some secret s. The statement φ is taken from some **NP** language $\mathcal{L}$ where random statements $\varphi \in \mathcal{L}$ are indistinguishable from statement not in $\mathcal{L}$ (for instance pseudorandom strings vs random strings for a sufficiently stretching pseudorandom generator). The function generally computes the verifier next message function $\mathsf{V}(\varphi, \cdot; r)$ with two exceptions. For some fixed public input `statement`, the function will output the statement φ. Also, given any accepting transcript ts, the function outputs its secret s.

To argue unlearnability, we show that any efficient quantum learner L that given oracle access to a random $f_{r,\varphi,s}$ finds s can be transformed into a prover that violates quantum resettable soundness. For this, we first show that any learner that manages find s with noticeable probability, can be translated into a learner that that given access to $\mathsf{V}(\varphi, \cdot; r)$ finds an accepting transcript ts, still with noticeable probability. For this we rely on a *quantum one-way to hiding lemma* by Ambainis, Hamburg, and Unruh [AHU19]. We then rely on the fact that φ is indistinguishable from a false statement to deduce that the prover will also succeed for no statements and thus break resettable soundness.

Finally, we show that we can use the non-black-box zero knowledge simulator to extract an accepting transcript with overwhelming probability. Given a quantum circuit C and state ρ implementing the function $f_{r,\varphi,s}$, say perfectly (although almost perfectly would still do). We can realize a quantum circuit along with quantum auxiliary input ρ that implement the verifier $\mathsf{V}(\varphi, \cdot; r)$. Here perfect correctness guarantees that when the constructed verifier computes its next messages, the state ρ is not disturbed, and thus we can repeatedly compute next messages. We can now run our non-black-box simulator (which also works relative to quantum auxiliary input), and by zero knowledge and completeness obtain an accepting transcript.

2.5 More Related Work

We now mention additional related work, and elaborate on some of the related works mentioned earlier.

Classical Resettable Security. The notion of resetting attacks was first considered by Canetti, Goldreich, Goldwasser, and Micali [CGGM00]. They defined and constructed protocols that are zero knowledge against resetting attacks. Resettable soundness was then introduced and achieved by Barak, Goldreich, Goldwasser, and Lindell [BGGL01]. Deng, Sahai, and Goyal showed how to construct a simultaneously resettable zero knowledge protocol [DGS09], this result was later followed by Goyal [Goy13] who gave a public coin protocol, by Chung, Ostrovsky, Pass and Visconti [COP+14] who gave a protocol based on one-way functions, and by Chongchitmate, Ostrovsky, and Visconti [COV17] who gave a constant round protocol, based on various standard assumptions. Goyal and Sahai [GS09b] and Goyal and Maji [GM11] defined and constructed varioues forms of resettable secure computation. Bitansky and Paneth [BP12, BP13, BP15] constructed resettably-sound protocols with various improved features based on unobfuscatability. Chung, Pass, and Seth [CPS13] constructed resettably-sound zero knowledge based on one-way functions. Finally, Chung, Ostrovsky, Pass, and Venkitasubramaniam [COP+14] presented a 4-round resettably sound zero-knowledge based on one-way functions.

Post-Quantum Zero-Knowledge for NP. The study of post-quantum zero-knowledge (QZK) protocols was initiated by Van De Graaf [VDGC97], who first observed that traditional zero-knowledge simulation

techniques, based on rewinding, fail against quantum verifiers. Subsequent work has further explored different flavors of zero knowledge and their limitations [Wat02], and also demonstrated that relaxed notions such as zero-knowledge with a trusted common reference string can be achieved [Kob03, DFS04]. Watrous [Wat09] was the first to show that the barriers of quantum information theory can be crossed, demonstrating a post-quantum zero-knowledge protocol for **NP** (in a polynomial number of rounds). A constant round non-black-box zero knowledge protocol was constructed by Bitansky and Shmueli [BS20] based on QLWE and quantum fully homomorphic encryption. Following up, Agarwal, Bartusek, Goyal, Khurana, and Malavolta [ABG+20] extended the BS construction to obtain parallel zero knowledge based on spooky encryptions for relations computable by quantum circuits.

Very recently Chia, Chung and Yamakawa [CCY20] showed that the Goldreich-Kahan protocol [GK96a] satisfies a relaxed notion called (post-quantum) ε-zero-knowledge; the protocol is based on collapse binding hash functions in the case of proofs, and on one-way functions in the case of arguments.

Barriers for 3-Message and Constant-Round Public-Coin Proofs. Classically, 3-message and constant-round public-coin zero knowledge arguments are subject to black-box barriers [GK96b], but can in fact be classically achieved using non-black-box simulation (under appropriate computational assumptions) [Bar01, BKP18]. In the case of proofs, there is evidence that they are unlikely to exists altogether (including non-black-box zero knowledge). Specifically, constant-round public-coin proofs do not exist assuming appropriate Fiat-Shamir hash functions [FS86, DNRS03, BLV06]. Kalai and the Rothblums [KRR17] gave such an instantiation of a Fiat Shamir hash assuming subepxoenential indistinguishability obfuscation, and strong forms of point obfuscation. Jain, Fleischhacker, and Goyal [FGJ18] extended their impossibility to also rule out three-message proofs. The mentioned implications also hold in the quantum setting, assuming post-quantum analogs of the corresponding assumptions. Jain, Kolla, Midrijanis and Reichardt [JKMR09] showed that for black-box zero knowledge, proofs can be ruled out unconditionally.

Simulating Quantum Oracles. Quantum oracles have been a fundamental aspect of quantum computation from the start. Querying the oracle in superposition created the need to develop new proof techniques. Specifically when proving security of quantum protocols in the Quantum Random Oracle Model ([BDF+10]). The main issue is the lack of ability to record the queries asked by the adversaries and to easily reprogram the answers. Nevertheless, many results were achieved even without these abilities [Zha12, Unr14, Zha15, ES15, Unr15, TU16, ABB+17, KLS18]. Following Zhandry's work [Zha18] on recording random oracles, many other results were proven such as the Fiat-Shamir transform [LZ19, DFMS19, DFM20], the Micali CS Proofs [CMS19], 4-round Luby-Rackoff construction [HI19] and more.

Quantum Obfuscation. Quantum obfuscation was first proposed by [AF16]. It's impossibility is not implied by the impossibility proved in [BGI+12]. In recent work, [ABDS20] showed the impossibility of such schemes based on the hardness of QLWE. A related stronger notion called Secure Software Leasing was dealt in [AP20] and [KNY20], showing the impossibility of such generic scheme (based on QLWE and the existence of QFHE), but the possibility of such schemes for restricted classes of functions (pseudo-random functions and evasive functions) under sub-exponential QLWE.

3 Preliminaries

3.1 Notations and Quantum Formalism

The following notations will be used throughout the paper,

- Let PPT stand for probabilistic polynomial-time algorithm and QPT stand for quantum polynomial-time algorithm.
- Let $[n]$ for $n \in \mathbb{N}$ stand for the set $\{0, \ldots, n-1\}$.
- We use sans-serif block letters (A) to denote quantum circuits and algorithms, upper-case letters (A) to denote quantum registers and small-case letters or Greek letters (a, α) to denote classical strings.

- We use the ket Dirac notation $|\psi\rangle$ to denote a pure quantum state. Mixed states will be denoted by Greek letters (for example ρ). For a pure state corresponding to the classical string x in the computational basis, we write $|x\rangle$.
- For a system $R = R_1 \otimes R_2 \cdots \otimes R_k$ on k qubits, $\mathcal{M}(R)$ denotes measure all qubits in R, for a set $S \subseteq [k]$, $\mathcal{M}_S(R)$ denotes measures the qubits in S. We abuse this notation, and sometime measure a function result on the input, and not an actual subset of qubits (for example for a function f, $\mathcal{M}_f(R)$ measures the value of f computed on register R via unitary to an empty register and then measuring and discarding this register).
- For a mixed state ρ we write $\|\rho\|_{\mathrm{tr}}$ to denote its trace norm, $\|\rho\|_{\mathrm{tr}} = \mathrm{Tr}\left(\sqrt{\rho^\dagger \rho}\right)$. For mixed states ρ, σ we denote by $\mathrm{TD}(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_{\mathrm{tr}}$ the trace distance.
- We use Id to denote the identity operator.
- For a distribution D, we write $x \leftarrow D$ to denote a sampling of x according to D. To signify a uniformly random sample from a fixed set S we write $x \leftarrow S$.

3.1.1 The Adversarial Model and Quantum Indistinguishability

Throughout, efficient adversaries are modeled as quantum circuits with non-uniform quantum advice (i.e. quantum auxiliary input). Formally, a polynomial-size adversary $\mathsf{A} = \{\mathsf{A}_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$, consists of a polynomial-size non-uniform sequence of quantum circuits $\{\mathsf{A}_\lambda\}_{\lambda\in\mathbb{N}}$, and a sequence of polynomial-size mixed quantum states $\{\rho_\lambda\}_{\lambda\in\mathbb{N}}$. For simplicity, we shall write only A where λ is clear from context.

For an interactive quantum adversary in a classical protocol, it can be assumed without loss of generality that its output message register is always measured in the computational basis at the end of computation. This assumption is indeed without the loss of generality, because whenever a quantum state is sent through a classical channel then qubits decohere and are effectively measured in the computational basis.

Indistinguishability in the Quantum Setting.

- Let $f : \mathbb{N} \to [0, 1]$ be a function.
  - f is negligible if for every constant $c \in \mathbb{N}$ there exists $N \in \mathbb{N}$ such that for all $n > N$, $f(n) < n^{-c}$.
  - f is noticeable if there exists $c \in \mathbb{N}, N \in \mathbb{N}$ such that for every $n \geq N$, $f(n) \geq n^{-c}$.
  - f is overwhelming if it is of the form $1 - \mu(n)$, for a negligible function μ.
- We use $\mathrm{poly}(n)$ to denote an unspecified polynomial in n, $p(n)$ and $\mathrm{negl}(n)$ to denote a negligible function in n $\mu(n)$.
- We may consider random variables over bit strings or over quantum states. This will be clear from the context.
- For two random variables X and Y supported on quantum states, quantum distinguisher circuit D with, quantum auxiliary input ρ, and $\mu \in [0, 1]$, we write $X \approx_{\mathsf{D},\rho,\mu} Y$ if

$$|\Pr[\mathsf{D}(X; \rho) = 1] - \Pr[\mathsf{D}(Y; \rho) = 1]| \leq \mu \ .$$

- Two ensembles of random variables $\mathcal{X} = \{X_i\}_{\lambda\in\mathbb{N}, i\in I_\lambda}$, $\mathcal{Y} = \{Y_i\}_{\lambda\in\mathbb{N}, i\in I_\lambda}$ over the same set of indices $I = \cup_{\lambda\in\mathbb{N}} I_\lambda$ are said to be *computationally indistinguishable*, denoted by $\mathcal{X} \approx_c \mathcal{Y}$, if for every polynomial-size quantum distinguisher $\mathsf{D} = \{\mathsf{D}_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ there exists a negligible function $\mu(\cdot)$ such that for all $\lambda \in \mathbb{N}, i \in I_\lambda$,

$$X_i \approx_{\mathsf{D}_\lambda,\rho_\lambda,\mu(\lambda)} Y_i \ .$$

- The trace distance between two distributions X, Y supported over quantum states, denoted $\mathrm{TD}(X,Y)$, is a generalization of statistical distance to the quantum setting and represents the maximal distinguishing advantage between two distributions supported over quantum states, by unbounded quantum algorithms. We thus say that ensembles $\mathcal{X} = \{X_i\}_{\lambda\in\mathbb{N}, i\in I_\lambda}$, $\mathcal{Y} = \{Y_i\}_{\lambda\in\mathbb{N}, i\in I_\lambda}$, supported over quantum states, are statistically indistinguishable (and write $\mathcal{X} \approx_s \mathcal{Y}$), if there exists a negligible function $\mu(\cdot)$ such that for all $\lambda \in \mathbb{N}, i \in I_\lambda$,

$$\mathrm{TD}(X_i, Y_i) \leq \mu(\lambda) \quad .$$

3.1.2 Quantum Oracles

In this section we define the notions related to oracle aided algorithms used throughout this paper. Primarily, we use the superscript A^{B} notation to denote that algorithm A has oracle access to algorithm B. In the classical scenario, it implies that A can query $\mathsf{B}(x)$ for arbitrary x, and obtain the result. We consider the standard notion of quantum oracle access to a classical function, which implies the ability to query in superposition. Formally, an oracle call to a classical function f is given by an application of the unitary map $|x\rangle|b\rangle \mapsto |x\rangle|b \oplus f(x)\rangle$.

We also consider the notion of access to a quantum oracle. By oracle access to a quantum circuit B, we mean access to the unitary that purifies the circuit B (such a unitary always exists by standard quantum purification techniques).

3.2 Interactive Protocols

We explicitly define the notions we use, regarding interactive protocols, as definitions vary across different papers in this domain, and we wish to avoid any confusion regarding the results we achieve.

3.2.1 Interactive Protocol

Definition 3.1 (Classical Proof and Argument Systems for NP)**.** *Let $\langle \mathsf{P}, \mathsf{V}\rangle$ be a protocol with an honest* PPT *prover P and an honest* PPT *verifier V for a language $\mathcal{L} \in \mathbf{NP}$, satisfying:*

1. **Completeness***: There exists a negligible function $\mu(\cdot)$ such that for any $\lambda \in \mathbb{N}, x \in \mathcal{L} \cap \{0,1\}^\lambda, w \in \mathcal{R}_\mathcal{L}(x)$,*

$$\Pr\left[\langle \mathsf{P}(w), \mathsf{V}\rangle(x) = 1\right] = 1 - \mu(\lambda) \quad .$$

2. **Soundness***: The protocol satisfies one of the following.*

   - **Computational Soundness***: For any quantum polynomial-size prover $\tilde{\mathsf{P}} = \left\{\tilde{\mathsf{P}}_\lambda, |\psi_\lambda\rangle\right\}_{\lambda\in\mathbb{N}}$, there exists a negligible function $\mu(\cdot)$ such that for any security parameter $\lambda \in \mathbb{N}$ and any $x \in \{0,1\}^\lambda \setminus \mathcal{L}$,*

$$\Pr\left[\left\langle \tilde{\mathsf{P}}_\lambda(|\psi_\lambda\rangle), \mathsf{V}\right\rangle(x) = 1\right] \leq \mu(\lambda) \quad .$$

     A protocol with computational soundness is called an argument.

   - **Statistical Soundness***: There exists a negligible function $\mu(\cdot)$, such that for any (unbounded) prover $\tilde{\mathsf{P}}$, any security parameter $\lambda \in \mathbb{N}$, and any $x \in \{0,1\}^\lambda \setminus \mathcal{L}$,*

$$\Pr\left[\left\langle \tilde{\mathsf{P}}, \mathsf{V}\right\rangle(x) = 1\right] \leq \mu(\lambda) \quad .$$

     A protocol with statistical soundness is called a proof.

Remark. *For an argument (i.e. a protocol with computational soundness), we can assume without the loss of generality that the amount of randomness used by the verifier is the size of the security parameter λ. The reason, is that if the randomness needed is more than that, we can use a PRG on randomness of size the*

security parameter and obtain a pseudorandom string in the appropriate length. To prove soundness, we first use the security of the PRG and move from using a pseudorandom string for the randomness of the verifier to using a truly random string. After we moved to using a truly random string for the verifier's randomness, the soundness security proof is identical to the original protocol's soundness proof.

Remark. *A protocol is called public-coin if all of the verifier's messages are uniformly random strings.*

3.2.2 Zero-Knowledge Protocols

For common input x, we denote by $\mathsf{OUT}_{\mathsf{V}}\langle \mathsf{P},\mathsf{V}\rangle$ the output of V in the protocol. For honest verifiers, this output will be a single bit indicating acceptance or rejection of the proof. Malicious quantum verifiers may have arbitrary quantum output (which is formally captured by the verifier outputting its inner quantum state).

Definition 3.2 (Post-Quantum Zero-Knowledge Classical Protocol)**.** *Let $\langle \mathsf{P},\mathsf{V}\rangle$ be a classical protocol (argument or proof) for a language $\mathcal{L}\in\mathbf{NP}$ as in Definition 3.1. The protocol is quantum zero-knowledge if it satisfies:*

***Post-Quantum Zero-Knowledge**: There exists a post-quantum polynomial-time simulator Sim, such that for any quantum polynomial-size verifier $\mathsf{V}^*=\{\mathsf{V}^*_\lambda,|\psi_\lambda\rangle\}_{\lambda\in\mathbb{N}}$,*

$$\left\{\mathsf{OUT}_{\mathsf{V}^*}\left(\langle P\left(w\right),\mathsf{V}^*_\lambda\left(|\psi_\lambda\rangle\right)\rangle\right)\right\}_{x,w,\lambda}\approx_c\left\{\mathsf{Sim}\left(x,\mathsf{V}^*_\lambda,|\psi_\lambda\rangle\right)\right\}_{x,w,\lambda}\quad,$$

where $\lambda\in\mathbb{N}$, $x\in\mathcal{L}\cap\{0,1\}^\lambda$, $w\in\mathcal{R}_{\mathcal{L}}\left(x\right)$.

Remark. *If Sim only uses V^* in a black-box way, meaning,*

$$\mathsf{Sim}\left(x,\mathsf{V}^*_\lambda,|\psi_\lambda\rangle\right):=\mathsf{Sim}^{\mathsf{V}^*_\lambda}\left(x,|\psi_\lambda\rangle\right)\quad,$$

then we say that the protocol is post-quantum black-box zero-knowledge classical protocol.

3.2.3 Witness Indistinguishability

We rely on constant-round public-coin protocols for **NP** that are witness-indistinguishable; that is, proofs that use different witnesses for the same statement are computationally indistinguishable. The proofs we use are also delayed-input, which means that all but the last message of the protocol can be computed independently of the instance x and witness w. More precisely, we use two flavors or WI protocols: one is 3-message proof systems (with statistical soundness) and quantum computational WI, and the second is 4-message argument systems (with quantum computational soundness) and statistical WI.

Definition 3.3 (Computational WI Protocol for **NP**)**.** *A classical protocol $\langle \mathsf{P},\mathsf{V}\rangle$ for a language $\mathcal{L}\in\mathbf{NP}$ (as in Definition 3.1) is computationally witness-indistinguishable if it satisfies:*

***Computational Witness Indistinguishability:** For every quantum polynomial-size verifier $\mathsf{V}^*=\{\mathsf{V}^*_\lambda,\rho_\lambda\}_\lambda$,*

$$\left\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle\mathsf{P}\left(w_0\right),\mathsf{V}^*_\lambda\left(\rho_\lambda\right)\rangle\left(x\right)\right\}_{\lambda,x,w_0,w_1}\approx_c\left\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle\mathsf{P}\left(w_1\right),\mathsf{V}^*_\lambda\left(\rho_\lambda\right)\rangle\left(x\right)\right\}_{\lambda,x,w_0,w_1}\quad,$$

where $\lambda\in\mathbb{N},x\in\mathcal{L}\cap\{0,1\}^\lambda$, and $w_0,w_1\in\mathcal{R}_{\mathcal{L}}\left(x\right)$ are witnesses for x.

For the statistical WI arguments we use sub-exponential statistical security.

Definition 3.4 (Sub-exponential Statistical WI Protocol for **NP**)**.** *A classical protocol $\langle \mathsf{P},\mathsf{V}\rangle$ for a language $\mathcal{L}\in\mathbf{NP}$ (as in Definition 3.1) is statistically, sub-exponentially witness-indistinguishable if it satisfies:*

***Statistical Witness Indistinguishability:** There exists a constant $\varepsilon\in(0,1)$ such that the following two ensembles have statistical distance bounded by $O(2^{-\lambda^\varepsilon})$, for every verifier $\mathsf{V}^*=\{\mathsf{V}^*_\lambda\}_{\lambda\in\mathbb{N}}$:*

$$\left\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle\mathsf{P}\left(w_0\right),\mathsf{V}^*_\lambda\left(\rho_\lambda\right)\rangle\left(x\right)\right\}_{\lambda,x,w_0,w_1}\quad,\quad\left\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle\mathsf{P}\left(w_1\right),\mathsf{V}^*_\lambda\left(\rho_\lambda\right)\rangle\left(x\right)\right\}_{\lambda,x,w_0,w_1}\quad,$$

where $\lambda\in\mathbb{N},x\in\mathcal{L}\cap\{0,1\}^\lambda$, and $w_0,w_1\in\mathcal{R}_{\mathcal{L}}\left(x\right)$ are witnesses for x.

Definition 3.5 (Delayed-Input Protocol for **NP**)**.** *A classical protocol* $\langle \mathsf{P}, \mathsf{V} \rangle$ *for a language* $\mathcal{L} \in \mathbf{NP}$ *(as in Definition 3.1) is delayed-input if it satisfies:*

Delayed-Input: *All messages of the protocol can be computed independently of the instance* $x \in \mathcal{L}$ *and witness* $w \in \mathcal{R}_{\mathcal{L}}(x)$*, but the last message in the protocol, where the prover computes it as a function of* x, w *and possibly previous and additional randomness.*

Instantiations. 3-message, public-coin classical proof systems with computational WI follow from classical zero-knowledge proof systems such as the parallel repetition of the Hamiltonicity protocol [Blu86], which is in turn based on non-interactive perfectly-binding commitments. For the proof system to be WI against quantum attacks, we need the non-interactive commitments to be computationally hiding against quantum adversaries, which can be instantiated for example from QLWE. 4-message, public-coin classical argument systems with sub-exponential statistical WI follow from the same protocol, only that the commitment of the prover is instantiated using collapse-binding commitments [Unr16b, Unr16a] which in turn are based on QLWE. For the protocols to have the delayed-input property, the Hamiltonicity protocol of [FLS99] is used.

3.3 Additional Tools

3.3.1 Pseudo-Random Generator

Definition 3.6 (Pseudo-Random Generator)**.** *A Pseudo-Random Generator (PRG) with stretch* $\ell(\cdot)$ *is a (deterministic) efficient algorithm* PRG *that maps an* n *bit string into a* $n + \ell(n)$ *bit string such that,*

$$\left\{\mathsf{PRG}(x) \,\middle|\, x \leftarrow \{0,1\}^{\lambda}\right\}_{\lambda} \approx_c \left\{y \,\middle|\, y \leftarrow \{0,1\}^{\lambda+\ell(\lambda)}\right\}_{\lambda} .$$

3.3.2 Pseudo-Random Functions

Definition 3.7 (Pseudo-Random Function Familiy)**.** *A pseudo-random function family is a function* $\mathsf{PRF} : \{0,1\}^{\lambda} \times \mathcal{X}(\lambda) \to \mathcal{Y}(\lambda)$ *where* $\{0,1\}^{\lambda}$ *is the key-space, and* $\mathcal{X}$ *and* $\mathcal{Y}$ *are the domain and range, with the following security guarantee,*

- **Indistinguishability from True Random Function***: Any efficient quantum adversary* A*, with oracle access to a randomly sampled function cannot distinguish it from oracle access to a truly random function. Meaning that for every non-uniform adversary* $\mathsf{A} = \{\mathsf{A}_{\lambda}, \rho_{\lambda}\}_{\lambda}$ *there exists a negligible function* $\mu(\cdot)$ *such that for every* $\lambda \in \mathbb{N}$

$$\left| \Pr_{k \leftarrow \{0,1\}^{\lambda}} \left[\mathsf{A}_{\lambda}^{\mathsf{PRF}_k}(\rho_{\lambda}) = 1\right] - \Pr_{R \leftarrow \mathcal{Y}^{\mathcal{X}}} \left[\mathsf{A}_{\lambda}^{R}(\rho_{\lambda}) = 1\right] \right| \leq \mu(\lambda) .$$

Instantiations [Zha12] shows that such pseudo-random functions can be constructed by standard constructions such as [GGM86] assuming quantum secure one-way functions.

3.3.3 Compute-and-Compare Obfuscation

We define compute-and-compare (CC) circuits and obfuscators for CC circuits.

Definition 3.8 (Compute-and-Compare Circuit)**.** *Let* $f : \{0,1\}^n \to \{0,1\}^{\lambda}$ *be a circuit, and let* $u \in \{0,1\}^{\lambda}, z \in \{0,1\}^*$ *be strings. Then* $\mathbf{CC}[f,u,z](x)$ *is a circuit that returns* z *if* $f(x) = u$*, and* $\perp$ *otherwise.* $\mathbf{CC}[f,u,z]$ *has a canonical description from which* f*,* u*, and* z *can be read.*

We now define compute-and-compare (CC) obfuscators (with perfect correctness). In what follows Obf is a PPT algorithm that takes as input a CC circuit $\mathbf{CC}[f,u,z]$ and outputs a new circuit $\widetilde{\mathbf{CC}}$.

Definition 3.9 (CC obfuscator)**.** *A* PPT *algorithm* Obf *is a compute-and-compare obfuscator if it satisfies:*

1. **Perfect Correctness:** *For any circuit* $f : \{0,1\}^n \to \{0,1\}^\lambda$, $u \in \{0,1\}^\lambda$ *and* $z \in \{0,1\}^*$,

$$\Pr\left[\forall x \in \{0,1\}^n : \widetilde{\mathbf{CC}}(x) = \mathbf{CC}[f,u,z](x) \;\middle|\; \widetilde{\mathbf{CC}} \leftarrow \mathsf{Obf}(\mathbf{CC}[f,u,z])\right] = 1 \ .$$

2. **Simulation:** *There exists a* PPT *simulator* **Sim** *such that for every two polynomials* $\ell_1(\cdot)$, $\ell_2(\cdot)$,

$$\{\widetilde{\mathbf{CC}} \mid u \leftarrow \{0,1\}^\lambda, \widetilde{\mathbf{CC}} \leftarrow \mathsf{Obf}(\mathbf{CC}[f,u,z])\}_{\lambda,f,z} \approx_c \{\mathbf{Sim}(1^{\ell_1(\lambda)}, 1^{\ell_2(\lambda)}, 1^\lambda)\}_{\lambda,f,z} \ ,$$

where $\lambda \in \mathbb{N}$, $f : \{0,1\}^n \to \{0,1\}^\lambda$ *is a* $\ell_1(\lambda)$*-size circuit,* $z \in \{0,1\}^{\ell_2(\lambda)}$. *With overwhelming probability over the simulator's randomness, it outputs a circuit that outputs* $\perp$ *on all inputs.*

Instantiations. Compute-and-compare obfuscators with almost-perfect correctness are constructed in [GKW17, WZ17] based on QLWE. CC obfuscators with perfect correctness are constructed [GKVW20] by Goyal, Koppula, Vusirikala and Waters, also based on QLWE.

3.3.4 Non-Interactive Commitments

We define non-interactive commitment schemes.

Definition 3.10 (Non-Interactive Commitment)**.** *A non-interactive commitment scheme is given by a* PPT *algorithm* $\mathsf{Com}(\cdot)$ *with the following syntax:*

- $\mathsf{cmt} \leftarrow \mathsf{Com}\left(1^\lambda, x\right)$: *A randomized algorithm that takes as input a security parameter* 1^λ *and input* $x \in \{0,1\}^*$*, and outputs a commitment* cmt.

The commitment algorithm satisfies:

1. **Perfect Binding:** *For any* $\lambda_0, \lambda_1 \in \mathbb{N}$, $x_0, x_1, r_0, r_1\{0,1\}^*$, $\mathsf{Com}\left(1^{\lambda_0}, x_0; r_0\right) = \mathsf{Com}\left(1^{\lambda_1}, x_1; r_1\right)$ *implies* $x_0 = x_1$.

2. **Computational Hiding:** *For any polynomial* $\ell(\cdot)$,

$$\{\mathsf{Com}\left(1^\lambda, x_0\right)\}_{\lambda,x_0,x_1} \approx_c \{\mathsf{Com}\left(1^\lambda, x_1\right)\}_{\lambda,x_0,x_1} \ ,$$

where $\lambda \in \mathbb{N}$, $x_0, x_1 \in \{0,1\}^{\ell(\lambda)}$.

Instantiations. The above non-interactive commitments are known based on various standard assumptions, including QLWE [GHKW17, LS19].

3.3.5 Quantum Fully Homomorphic Encryption

We rely on quantum fully homomorphic encryption, specifically, a scheme where a classical input can be encrypted classically and a quantum input quantumly. The formal definition follows.

Definition 3.11 (Quantum Fully-Homomorphic Encryption)**.** *A quantum fully homomorphic encryption scheme is given by six algorithms* (QFHE.Gen, QFHE.Enc, QFHE.QEnc, QFHE.Dec, QFHE.QDec, QFHE.Eval) *with the following syntax:*

- $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{QFHE.Gen}(1^\lambda)$: *A* PPT *algorithm that given a security parameter* 1^λ*, samples a classical public key* pk *and a classical secret key* sk.

- $\mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_{\mathsf{pk}}(x)$: *A* PPT *algorithm that takes as input a classical string* $x \in \{0,1\}^*$ *and outputs a classical ciphertext* ct.

- $|\phi\rangle \leftarrow \mathsf{QFHE.QEnc}_{\mathsf{pk}}(|\psi\rangle)$: *A* QPT *algorithm that takes as input a quantum state* $|\psi\rangle$ *and outputs a quantum ciphertext* $|\phi\rangle$.

- $x \leftarrow \mathsf{QFHE.Dec}_{\mathsf{sk}}(\mathsf{ct})$: *A* PPT *algorithm that takes as input a classical ciphertext* ct *and outputs a string* x*.*
- $|\psi\rangle \leftarrow \mathsf{QFHE.QDec}_{\mathsf{sk}}(|\phi\rangle)$: *A* QPT *algorithm that takes as input a quantum ciphertext* $|\phi\rangle$ *and outputs a quantum state* $|\psi\rangle$*.*
- $|\hat{\phi}\rangle \leftarrow \mathsf{QFHE.Eval}_{\mathsf{pk}}(C, \mathsf{ct}, |\phi\rangle)$: *A* QPT *algorithm that takes as input a general quantum circuit* C*, a classical ciphertext* ct *and a quantum ciphertext* $|\phi\rangle$ *and outputs an evaluated quantum ciphertext* $|\hat{\phi}\rangle$

The scheme satisfies the following.

- **Quantum Semantic Security:** *For every polynomial* $\ell(\cdot)$*,*

$$\left\{ (\mathsf{ct}, |\phi\rangle) \;\middle|\; \begin{array}{l} (\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{QFHE.Gen}\left(1^\lambda\right), \\ \mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_{\mathsf{pk}}(x_0), \\ |\phi\rangle \leftarrow \mathsf{QFHE.QEnc}_{\mathsf{pk}}(|\psi_0\rangle) \end{array} \right\}_{\lambda, x_0, |\psi_0\rangle, x_1, |\psi_1\rangle} \approx_c$$

$$\left\{ (\mathsf{ct}, |\phi\rangle) \;\middle|\; \begin{array}{l} (\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{QFHE.Gen}\left(1^\lambda\right), \\ \mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_{\mathsf{pk}}(x_1), \\ |\phi\rangle \leftarrow \mathsf{QFHE.QEnc}_{\mathsf{pk}}(|\psi_1\rangle) \end{array} \right\}_{\lambda, x_0, |\psi_0\rangle, x_1, |\psi_1\rangle} ,$$

  where $\lambda \in \mathbb{N}$*,* $x_0, x_1 \in \{0,1\}^{\ell(\lambda)}$ *and* $|\psi_0\rangle$*,* $|\psi_1\rangle$ *are* $\ell(\lambda)$*-qubit states.*

- **Compactness:** *There exists a polynomial* $\mathrm{poly}(\cdot)$ *s.t. for every quantum circuit* C *with* ℓ *output qubits and an encryption of an input for* C*, the output size of the evaluation algorithm is* $\ell \cdot \mathrm{poly}(\lambda)$*, where* λ *is the security parameter of the scheme*
- **Measurement-Preserving Homomorphism:** *For every polynomial* $s(\cdot)$ *there exists a negligible function* $\mathrm{negl}(\lambda)(\cdot)$ *such that for every* $\lambda \in \mathbb{N}$*, size-*$s(\lambda)$ *quantum circuit* C*, input* $(x, |\psi\rangle)$ *for* C *which is comprised of a classical string* x *and quantum state* $|\psi\rangle$*, subset* M *of the output qubits of* C*, public and secret key pair* $(\mathsf{pk}, \mathsf{sk}) \in \mathsf{QFHE.Gen}(1^\lambda)$ *and randomness strings* $\left(r_x, r_{|\psi\rangle}\right)$*:*

$$\mathrm{TD}(D_0, D_1) \leq \mathrm{negl}(\lambda) \quad ,$$

  where D_0, D_1 *are the distributions which are defined as follows:*

  - D_0 : *Compute* $|\psi'\rangle \leftarrow C(x, |\psi\rangle)$*, measure the subset of qubits of* $|\psi'\rangle$ *which are in* M *and output the obtained state.*
  - D_1 :
    - *Encrypt* $\mathsf{ct} = \mathsf{QFHE.Enc}_{\mathsf{pk}}(x; r_x)$*,* $|\phi\rangle = \mathsf{QFHE.QEnc}_{\mathsf{pk}}\left(|\psi\rangle; r_{|\psi\rangle}\right)$*.*
    - *Evaluate* $|\hat{\phi}\rangle \leftarrow \mathsf{QFHE.Eval}_{\mathsf{pk}}(C, \mathsf{ct}, |\phi\rangle)$*.*
    - *Measure the* $|M|$ *packets of qubits that correspond to the output qubits in* M *(by compactness, each packet is exactly of size* $\mathrm{poly}(\lambda)$*).*
    - *Decrypt the measured* $|M|$ *packets with* $\mathsf{QFHE.Dec}_{\mathsf{sk}}(\cdot)$*, and decrypt the rest of the qubits with* $\mathsf{QFHE.QDec}_{\mathsf{sk}}(\cdot)$*. Output the obtained state.*

Instantiations. Mahadev [Mah18b] shows how to build quantum FHE based on super-polynomial QLWE modulus and a circular security assumption with respect to a secret key and an additional trapdoor information. Brakerski [Bra18b] subsequently shows how to construct quantum FHE based on polynomial QLWE modulus and a circular security assumption (analogous to the assumptions required for multi-key FHE in the classical setting). The above definition is more specific then the standard definition of QFHE. Specifically, *measurement-preservation* and (statistical) correctness for *every* triplet $(\mathsf{pk}, \mathsf{sk}, r)$ of public and secret keys and randomness r for the encryption algorithm, is not an explicit part of the standard definition. The construction of Brakerski satisfies this more general definition. This follows readily from the main Theorem (4.1) in [Bra18b].

3.3.6 Function-Hiding Secure Function Evaluation

We define two-message function evaluation protocols with sub-exponential statistical circuit privacy and quantum computational input privacy.

Definition 3.12 (2-Message Function Hiding SFE)**.** *A two-message secure function evaluation protocol* $(\mathsf{SFE.Gen}, \mathsf{SFE.Enc}, \mathsf{SFE.Eval}, \mathsf{SFE.Dec})$ *has the following syntax:*

- $\mathsf{sk} \leftarrow \mathsf{SFE.Gen}\left(1^\lambda\right)$: *a probabilistic algorithm that takes a security parameter* 1^λ *and outputs a secret key* sk.
- $\mathsf{ct} \leftarrow \mathsf{SFE.Enc}_{\mathsf{sk}}\left(x\right)$: *a probabilistic algorithm that takes a string* $x \in \{0,1\}^*$*, and outputs a ciphertext* ct.
- $\hat{\mathsf{ct}} \leftarrow \mathsf{SFE.Eval}\left(C, \mathsf{ct}\right)$: *a probabilistic algorithm that takes a (classical) circuit* C *and a ciphertext* ct *and outputs an evaluated ciphertext* $\hat{\mathsf{ct}}$.
- $\hat{x} = \mathsf{SFE.Dec}_{\mathsf{sk}}\left(\hat{\mathsf{ct}}\right)$: *a deterministic algorithm that takes a ciphertext* $\hat{\mathsf{ct}}$ *and outputs a string* $\hat{x}$.

The scheme satisfies the following.

- **Perfect Correctness:** *For any polynomial* $s\left(\cdot\right)$*, for any* $\lambda \in \mathbb{N}$*, size-*$s\left(\lambda\right)$ *circuit* C *and input* x *for* C*,*

$$\Pr\left[\mathsf{SFE.Dec}_{\mathsf{sk}}(\hat{\mathsf{ct}}) = C(x) \;\middle|\; \begin{array}{l} \mathsf{sk} \leftarrow \mathsf{SFE.Gen}(1^\lambda), \\ \mathsf{ct} \leftarrow \mathsf{SFE.Enc}_{\mathsf{sk}}(x), \\ \hat{\mathsf{ct}} \leftarrow \mathsf{SFE.Eval}(C, \mathsf{ct}) \end{array}\right] = 1 \ .$$

- **Quantum Input Privacy:** *For every polynomial* $\ell\left(\cdot\right)$*,*

$$\left\{\mathsf{ct} \;\middle|\; \begin{array}{l} \mathsf{sk} \leftarrow \mathsf{SFE.Gen}(1^\lambda), \\ \mathsf{ct} \leftarrow \mathsf{SFE.Enc}_{\mathsf{sk}}\left(x_0\right) \end{array}\right\}_{\lambda, x_0, x_1} \approx_c \left\{\mathsf{ct} \;\middle|\; \begin{array}{l} \mathsf{sk} \leftarrow \mathsf{SFE.Gen}(1^\lambda), \\ \mathsf{ct} \leftarrow \mathsf{SFE.Enc}_{\mathsf{sk}}\left(x_1\right) \end{array}\right\}_{\lambda, x_0, x_1} ,$$

  where $\lambda \in \mathbb{N}$ *and* $x_0, x_1 \in \{0,1\}^{\ell(\lambda)}$.

- **Sub-exponential Statistical Circuit Privacy:** *There exist unbounded algorithms, probabilistic* $\mathsf{SFE.Sim}$ *and deterministic* $\mathsf{SFE.Ext}$ *and a constant* $\varepsilon \in (0,1)$ *such that:*
  - *For every* $x \in \{0,1\}^*$*,* $\mathsf{ct} \in \mathsf{SFE.Enc}\left(x\right)$*, the extractor outputs* $\mathsf{SFE.Ext}\left(\mathsf{ct}\right) = x$.
  - *For any polynomial* $s(\cdot)$ *the following two ensembles has statistical distance bounded by* $O(2^{-\lambda^\varepsilon})$*,*

$$\left\{\mathsf{SFE.Eval}\left(C, \mathsf{ct}^*\right)\right\}_{\lambda, C, \mathsf{ct}^*} \ , \ \left\{\mathsf{SFE.Sim}\left(\ 1^\lambda, C(\mathsf{SFE.Ext}(1^\lambda, \mathsf{ct}^*))\ \right)\right\}_{\lambda, C, \mathsf{ct}^*} \ ,$$

    where $\lambda \in \mathbb{N}$*,* C *is a* $s(\lambda)$*-size circuit, and* $\mathsf{ct}^* \in \{0,1\}^*$.

Remark. *For the computational security of the input's privacy in the above SFE scheme, we can assume without the loss of generality that the amount of randomness used by the party encrypting the input, is the size of the security parameter* λ*. The reason, is that if the randomness needed is more than that, we can use a PRG on randomness of size the security parameter and obtain a pseudorandom string in the appropriate length. To prove the computational input privacy in the new scheme, we first use the security of the PRG and move from using a pseudorandom string for the randomness of the encryption to using a truly random string. After we moved to using a truly random string, the security proof is identical to that for proving the input privacy in the original scheme.*

Instantiations. Such secure function evaluation schemes are known based on QLWE [OPCPC14, BD18].

4 Defining Post-Quantum Resettable Soundness

In this section, we present our definition of resettable soundness, and show and immediate implication of this definition, regarding the triviality of black-box zero-knowledge arguments with resettable soundness.

4.1 Post-Quantum Resettable Soundness

We present our definition for post-quantum resettable soundness. Our definition deals with giving oracle access to fixed verifier. We shall use $\mathsf{V}(x,\cdot;r)$ to denote the interaction of algorithm V on instance x fixed randomness r (where the input is a partial transcript). Also, to denote the application of V's predicate on a transcript ts we shall write $\mathsf{V}(x,\mathsf{ts};r)$. The definition of resettable soundness is as follows,

Definition 4.1 (Post-Quantum Resettable Soundness)**.** *A classical interactive protocol $\langle\mathsf{P},\mathsf{V}\rangle$ for language $\mathcal{L}$ has resettable soundness against quantum provers, if for any malicious* QPT *resetting prover $\mathsf{rP}=\{\mathsf{rP}_\lambda,|\psi_\lambda\rangle\}_{\lambda\in\mathbb{N}}$ there exists a negligible function $\mu(\cdot)$ such for any security parameter $\lambda\in\mathbb{N}$ and any $x\in\{0,1\}^\lambda\setminus\mathcal{L}$ it holds that,*

$$\Pr_r\left[\mathsf{V}(x,\mathsf{ts};r)=1\;\middle|\;\mathsf{ts}\leftarrow\mathsf{rP}_\lambda^{\mathsf{V}(x,\cdot;r)}(|\psi_\lambda\rangle)\right]\leq\mathrm{negl}(\lambda)\quad,$$

where ts is a transcript of a possible interaction between P,V. $\mathsf{V}(x,\cdot;r)$ is the function that computes V's next message, on instance x and some fixed randomness r, given as input a transcript of a partial interaction.

As a shorthand we shall use $\Pr_r\left[\langle\mathsf{rP},\mathsf{V}(x,\cdot;r)\rangle(x)=1\right]$ to measure the probability of success in the above experiment. We also consider a variant of the above definition where the resetting prover can also query on multiple instances, formally

Definition 4.2 (Post-Quantum Multi-Input Resettable Soundness)**.** *A classical interactive protocol $\langle\mathsf{P},\mathsf{V}\rangle$ for language $\mathcal{L}$ has multi-input resettable soundness against quantum provers, if for any malicious* QPT *resetting prover $\mathsf{rP}=\{\mathsf{rP}_\lambda,|\psi_\lambda\rangle\}_{\lambda\in\mathbb{N}}$ there exists a negligible function $\mu(\cdot)$ such for any security parameter $\lambda\in\mathbb{N}$ it holds that,*

$$\Pr_r\left[\begin{array}{c}x\notin L\\ \mathsf{V}(x,\mathsf{ts};r)=1\end{array}\;\middle|\;(x,\mathsf{ts})\leftarrow\mathsf{rP}_\lambda^{\mathsf{V}(\cdot;r)}(|\psi_\lambda\rangle)\right]\leq\mathrm{negl}(\lambda)\quad,$$

where ts is a transcript of a possible interaction between P,V. $V(\cdot;r)$ is the function that computes V's next message assuming some fixed randomness r, given as input a partial transcript, starting with an instance $x\in\{0,1\}^\lambda$ which the interaction is on.

Remark. *Another variant one might consider is giving the resetting prover the ability to learn if the verifier accepts or rejects a possible transcript by querying it with a full transcript. We note here that assuming that $\langle\mathsf{P},\mathsf{V}\rangle$ is publicly-verifiable this trait is satisfied automatically.*

4.2 Black-Box Zero Knowledge with Resettable Soundness is Trivial

In this section, we prove an analog of the classical claim found in [BGGL01], on the triviality of zero-knowledge protocols with resettable soundness and a black-box simulator. More formally we prove the following claim,

Theorem 4.3. *If a language $\mathcal{L}$ has a post-quantum black-box zero-knowledge, resettably sound protocol, then $\mathcal{L}\in\mathbf{BQP}$.*

Proof. The proof is similar to the proof in classical case, and is described for completeness in Appendix A.1 □

5 Transforming Protocols to Achieve Quantum Resettable Soundness

In this section we show that classical three-message protocols as well as constant-round public-coin protocols can be made resettably sound assuming one-way functions. The transformation is simple and similar to the one from the classical setting [BGGL01], however, having to deal with quantum resetting attacks, the analysis is significantly different. The transformation preserves black-box zero-knowledge; accordingly, we deduce as a corollary that post-quantum black-box zero-knowledge protocols cannot be 3-message or constant-round public-coin, except for trivial languages.

5.1 Quantum Oracle Notations

We rely on a couple of lemmas proved in [DFM20]. We restate them here again, while augmenting some of the notation, to fit with our conventions. Let A^H be a quantum oracle-aided algorithm. For a q-query algorithm, without loss of generality, A can be described as having the following registers, query registers on which we apply the unitary $\mathcal{O}_H$ computing $|x\rangle|y\rangle \to |x\rangle|y \oplus H(x)\rangle$, X, Z which are output registers, and E holds any other internal qubits used by A. More so, the operation of A on its initial state can be described as,

$$\mathsf{A}^H = \mathsf{A}_q \mathcal{O}_H \dots \mathsf{A}_1 \mathcal{O}_H \ ,$$

where A_i is a sequence of unitaries. Like [DFM20] we use the following notation for $i < j \in [q]$

$$\mathsf{A}^H_{i \to j} = \mathsf{A}_j \mathcal{O}_H \dots \mathsf{A}_{i+1} \mathcal{O}_H \ .$$

We also denote $\mathsf{A}^H_{i \to j} = \mathsf{Id}$ for $i \geq j \in [q]$. Assuming A gets as initial input a pure state $|\phi_0\rangle$, we denote,

$$|\phi^H_i\rangle = \mathsf{A}^H_{0 \to i} |\phi_0\rangle \ .$$

For a function H we denote by $H_{x \to \theta}$ the same function where x is remapped to θ:

$$H_{x \to \theta}(x') = \begin{cases} H(x') & x' \neq x \\ \theta & x' = x \end{cases} \ .$$

5.2 Transforming 3 Message Private Coin Protocols

We show that any 3 message interactive protocol $\langle \mathsf{P}, \mathsf{V} \rangle$ can be transformed to a quantum resettably sound one, assuming the existence of quantum secure PRFs. More formally we show the following,

Proposition 5.1 (Compiler For 3 Message Protocols)**.** *Assuming quantum-secure one-way functions, any 3 message protocol $\langle \mathsf{P}, \mathsf{V} \rangle$ with negligible soundness for a language $\mathcal{L}$, can be transformed to into a post-quantum resettably sound protocol $\left\langle \mathsf{P}, \tilde{\mathsf{V}} \right\rangle$. More so, if $\langle \mathsf{P}, \mathsf{V} \rangle$ is (black-box) zero-knowledge then so is $\left\langle \mathsf{P}, \tilde{\mathsf{V}} \right\rangle$.*

Combining proposition 5.1 with theorem 4.3 immediately implies the following corollary,

Corollary 5.2. *If $\mathcal{L}$ has a 3 message post-quantum black-box zero-knowledge protocol, then $\mathcal{L} \in$ **BQP**.*

5.2.1 Single Value Reprogramming

To prove our construction presented in 5.2.2, we shall rely on a lemma by [DFM20].

Lemma 5.3 (Single Value Reprogramming Lemma ([DFM20]))**.** *Let A be a q-query oracle quantum algorithm. Then, for any function $H : \mathcal{X} \to \mathcal{Y}$, any $x \in \mathcal{X}$ and $\theta \in \mathcal{Y}$, and any projection $\Pi_{x,\theta}$ acting on the Z register (which may depend on x, θ), it holds that*

$$\mathop{\mathbb{E}}_{i,b} \left[\left\| (|x\rangle\langle x| \otimes \Pi_{x,\theta}) \left(\mathsf{A}^{H_{x \to \theta}}_{i+b \to q} \right) \left(\mathsf{A}^H_{i \to i+b} \right) (|x\rangle\langle x|) \, |\phi^H_i\rangle \right\|_2^2 \right] \geq \frac{\left\| (|x\rangle\langle x| \otimes \Pi_{x,\theta}) \, |\phi^{H_{x \to \theta}}_q\rangle \right\|_2^2}{(2q+1)^2} \ ,$$

where the expectation is over uniform $(i,b) \in \{0,\ldots,q-1\} \times \{0,1\} \cup \{(q,0)\}$. *We emphasize that first* $|x\rangle\langle x|$ *acts on query register, while the second acts on the* X *register.*

Remark. *We state here the technical lemma and not the existence of a simulator, as done in the multiple values reprogramming in the public-coin case, since unlike [DFM20] we use this lemma to reprogram a non-uniform output function, in our private-coin transform.*

5.2.2 Construction

Fix some language $\mathcal{L}$ with a three-message protocol $\langle \mathsf{P}, \mathsf{V} \rangle$ whose message we denote by (α, β, γ). Assume V uses $m(\lambda)$ bits of randomness. We present the protocol $\left\langle \mathsf{P}, \tilde{\mathsf{V}} \right\rangle$. P is exactly the same, where as $\tilde{\mathsf{V}}$ is described in 1.

Algorithm 1: $\tilde{\mathsf{V}}(x;k)$

1 Use k as a key for $\mathsf{PRF}_k(\cdot)$, a pseudo-random function.
2 Given α compute $\beta = \mathsf{V}(x, \alpha; \mathsf{PRF}_k(\alpha))$.
3 Given a transcript α, β, γ compute $\mathsf{V}(x, (\alpha, \beta, \gamma); \mathsf{PRF}_k(\alpha))$ and output it.

The fact that the protocol preserves completeness and zero-knowledge follows readily, we focus on proving resettable soundness. To show resettable soundness, we show an efficient reduction from a resetting prover rP to a prover $\tilde{\mathsf{P}}$ for the original protocol, which preserves the cheating probability up to a polynomial loss.

Fix a malicious quantum resetting prover rP for a false instance x. Assume that rP makes at most q oracle queries, and has non-uniform advice $|\psi_0\rangle$. Assume rP has registers A, Z, E and query registers. The query registers are for querying a first message α and receiving the corresponding second message β. A, Z will hold the outputted first and third message, and E holds any internal qubits used. Then, $\tilde{\mathsf{P}}$ will perform as follows,

Algorithm 2: $\tilde{\mathsf{P}}(x)$ - Malicious Quantum Prover for $\langle \mathsf{P}, \mathsf{V} \rangle$

1 Sample $(i,b) \leftarrow \{0,\ldots,q-1\} \times \{0,1\} \cup \{(q,0)\}$.
2 Sample $k \leftarrow \{0,1\}^\lambda$.
3 Run $\mathsf{rP}^{\tilde{\mathsf{V}}(x,\cdot;k)}_{0 \to i} |\psi_0\rangle$ and denote the resulting state $|\psi_i^{\tilde{\mathsf{V}}(x,\cdot;k)}\rangle$.
4 Measure the query register to obtain a value α and send it as the first message. Denote the state after measurement by $|\phi_i^{\tilde{\mathsf{V}}(x,\cdot;k)}(\alpha)\rangle$.
5 Upon receiving the second message β, run $\left(\mathsf{rP}^{\tilde{\mathsf{V}}(x,\cdot;k)_{\alpha\to\beta}}_{i+b\to q}\right)\left(\mathsf{rP}^{\tilde{\mathsf{V}}(x,\cdot;k)}_{i\to i+b}\right)|\phi_i^{\tilde{\mathsf{V}}(x,\cdot;k)}(\alpha)\rangle$.
6 Measure A, Z to obtain (α', γ) if $\alpha' = \alpha$ output γ as the third message, otherwise abort.

We show that,

Claim 5.4.

$$\Pr\left[\left\langle \tilde{\mathsf{P}}, \mathsf{V} \right\rangle(x) = 1\right] \geq \frac{1}{(2q+1)^2} \Pr_k\left[\left\langle \mathsf{rP}, \tilde{\mathsf{V}}(x,\cdot;k) \right\rangle(x) = 1\right] - \mathrm{negl}(\lambda) \quad .$$

Proof. We denote by $\tilde{\mathsf{V}}^R$ a version of $\tilde{\mathsf{V}}$ such that $\tilde{\mathsf{V}}$ uses a truly random function R to derive its randomness (i.e it runs $\mathsf{V}(x,\cdot,R(\alpha))$ for a first message α). From the pseudo-randomness of the PRF it holds that,

$$\Pr_k\left[\left\langle \mathsf{rP}, \tilde{\mathsf{V}}(x,\cdot;k) \right\rangle(x) = 1\right] - \mathrm{negl}(\lambda) \leq \mathop{\mathbb{E}}_R\left[\Pr\left[\left\langle \mathsf{rP}, \tilde{\mathsf{V}}^R \right\rangle(x) = 1\right]\right] \tag{1}$$

We also denote $\tilde{\mathsf{P}}^R$ to be the malicious prover that uses $\tilde{V}^R$ (where R is a truly random function) instead of $\mathsf{V}(x,\cdot;k)$ as the oracle for rP. Again by pseudo-randomness of the PRF it holds that,

$$\Pr\left[\left\langle\tilde{\mathsf{P}},\mathsf{V}\right\rangle(x)=1\right]\geq\mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle\tilde{\mathsf{P}}^R,\mathsf{V}\right\rangle(x)=1\right]\right]-\mathrm{negl}(\lambda) \tag{2}$$

We define the event $W(i,b,\alpha,r,R)$ to be the event where after sampling an external verifier's randomness r, sampling i,b by $\tilde{\mathsf{P}}^R$ and measuring α as the first message in stage 4, $\tilde{\mathsf{P}}^R$ succeeds in convincing the external verifier. Then it holds that,

$$\begin{aligned}\mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle\tilde{\mathsf{P}}^R,\mathsf{V}\right\rangle(x)=1\right]\right]&=\mathop{\mathbb{E}}_{r,R}\left[\Pr\left[\left\langle\tilde{\mathsf{P}}^R,\mathsf{V}(x;r)\right\rangle(x)=1\right]\right]\\&=\sum_{\alpha}\mathop{\mathbb{E}}_{r,R}\left[\mathop{\mathbb{E}}_{i,b}\left[\Pr\left[W(i,b,\alpha,r,R)\right]\right]\right]\ .\end{aligned}$$

Also, we note that,

$$\Pr\left[W(i,b,\alpha,r,R)\right]=\left\||\alpha\rangle\langle\alpha|\otimes\Pi^{\alpha}_{\mathsf{V}(x,\cdot;r)}\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}_{i+b\to q}\right)\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R}_{i\to i+b}\right)|\alpha\rangle\langle\alpha||\psi_i^{\tilde{\mathsf{V}}^R}\rangle\right\|^2\ ,$$

where

$$\Pi^{\alpha}_{f}=\sum_{c:f(\alpha,f(\alpha),c)=1}|c\rangle\langle c|\ ,$$

the first $|\alpha\rangle\langle\alpha|$ is applied to the query register, the second $|\alpha\rangle\langle\alpha|$ is applied to the A register, and $\Pi^{\alpha}_{V(x,\cdot;r)}$ is applied to the Z register. Hence, it holds,

$$\begin{aligned}&\mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle\tilde{\mathsf{P}}^R,\mathsf{V}\right\rangle(x)=1\right]\right]=\\&\sum_{\alpha}\mathop{\mathbb{E}}_{r,R}\left[\mathop{\mathbb{E}}_{i,b}\left[\left\||\alpha\rangle\langle\alpha|\otimes\Pi^{\alpha}_{\mathsf{V}(x,\cdot;r)}\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}_{i+b\to q}\right)\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R}_{i\to i+b}\right)|\alpha\rangle\langle\alpha||\psi_i^{\tilde{\mathsf{V}}^R}\rangle\right\|^2\right]\right]\ .\end{aligned}$$

For any fixed α,r,k by the single value reprogramming lemma (5.3), it holds that,

$$\begin{aligned}&\mathop{\mathbb{E}}_{i,b}\left[\left\||\alpha\rangle\langle\alpha|\otimes\Pi^{\alpha}_{\mathsf{V}(x,\cdot;r)}\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}_{i+b\to q}\right)\left(\mathsf{rP}^{\tilde{\mathsf{V}}^R}_{i\to i+b}\right)|\alpha\rangle\langle\alpha||\psi_i^{\tilde{\mathsf{V}}^R}\rangle\right\|^2\right]\geq\\&\frac{\left\|(|\alpha\rangle\langle\alpha|)\otimes\Pi^{\alpha}_{\mathsf{V}(x,\cdot;r)}|\psi_q^{\tilde{\mathsf{V}}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}\rangle\right\|^2}{(2q+1)^2}\ .\end{aligned}$$

Above, $|\psi_q^{\tilde{\mathsf{V}}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}\rangle=\mathsf{rP}^{\tilde{V}^R_{\alpha\to\mathsf{V}(x,\alpha;r)}}|\psi_0\rangle$ (see subsection 5.1 for further explanation on notation, which will be used freely from now on). Hence it holds that,

$$\begin{aligned}
\mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle \tilde{\mathsf{P}}^{R},\mathsf{V}\right\rangle (x)=1\right]\right] &\geq \sum_{\alpha}\mathop{\mathbb{E}}_{r,R}\left[\frac{\left\|(|\alpha\rangle\langle\alpha|)\otimes\Pi^{\alpha}_{\mathsf{V}(x,\cdot;r)}|\psi_q^{\tilde{\mathsf{V}}^{R}_{\alpha\to\mathsf{V}(x,\alpha;r)}}\rangle\right\|^2}{(2q+1)^2}\right]\\
&=\sum_{\alpha}\mathop{\mathbb{E}}_{r,R}\left[\frac{\left\|(|\alpha\rangle\langle\alpha|)\otimes\Pi^{\alpha}_{\tilde{\mathsf{V}}^{R}_{\alpha\to\mathsf{V}(x,\alpha;r)}}|\psi_q^{\tilde{\mathsf{V}}^{R}_{\alpha\to\mathsf{V}(x,\alpha;r)}}\rangle\right\|^2}{(2q+1)^2}\right]\\
&\underset{(*)}{=}\sum_{\alpha}\mathop{\mathbb{E}}_{r,R}\left[\frac{\left\|(|\alpha\rangle\langle\alpha|)\otimes\Pi^{\alpha}_{\tilde{\mathsf{V}}^{R}}|\psi_q^{\tilde{\mathsf{V}}^{R}}\rangle\right\|^2}{(2q+1)^2}\right]\\
&=\mathop{\mathbb{E}}_{R}\left[\frac{\Pr\left[\left\langle \mathsf{rP},\tilde{\mathsf{V}}^{R}\right\rangle (x)=1\right]}{(2q+1)^2}\right],
\end{aligned}$$

where $(*)$ follows for any x,α and uniformly sampled r,R the oracles $\tilde{V}^{R}$ and $\tilde{V}^{R}_{\alpha\to(x,\alpha;r)}$ are perfectly indistinguishable. Thus, it holds

$$\mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle \tilde{\mathsf{P}}^{R},\mathsf{V}\right\rangle (x)=1\right]\right]\geq\mathop{\mathbb{E}}_{R}\left[\frac{\Pr\left[\left\langle \mathsf{rP},\tilde{\mathsf{V}}^{R}\right\rangle (x)=1\right]}{(2q+1)^2}\right].$$

Hence, by combining equations 1,2 with the equation above, the claim follows. □

5.3 Transforming Constant-Round Public-Coin Protocols

We show that any constant-round public-coin protocol can be transformed into a post-quantum resettably sound one, assuming the existence of quantum secure PRFs. More formally, we show,

Proposition 5.5 (Compiler For Public-Coin Protocols)**.** *Assuming quantum-secure one-way functions, any public-coin constant-round protocol with negligible soundness for a language $\mathcal{L}$, $\langle\mathsf{P},\mathsf{V}\rangle$ can be transformed into a post-quantum resettably sound protocol $\left\langle\mathsf{P},\tilde{\mathsf{V}}\right\rangle$. More so, if $\langle\mathsf{P},\mathsf{V}\rangle$ is (black-box) zero-knowledge so is $\left\langle\mathsf{P},\tilde{\mathsf{V}}\right\rangle$.*

The above proposition 5.5 with the above theorem 4.3 immediately imply the following corollary,

Corollary 5.6. *If $\mathcal{L}$ has a constant-round public-coin post-quantum black-box zero-knowledge protocol, $\mathcal{L}\in$* **BQP**.

5.3.1 Multiple Value Reprogramming

For a function $H:\mathcal{X}_0\times(\mathcal{X}\times\mathcal{Y})^*\to\mathcal{Y}$ and given a tuple $\mathbf{x}=(x_0,x_1,\ldots,x_n)$ in $\mathcal{X}_0\times\mathcal{X}^n$, we define the chained function, $\mathbf{h}^{H,\mathbf{x}}=\left(h_1^{H,\mathbf{x}},\ldots,h_n^{H,\mathbf{x}}\right)$ given by

$$h_1^{H,\mathbf{x}}=H\left(x_0,x_1\right)\qquad\text{and}\qquad h_i^{H,\mathbf{x}}:=H\left(x_0,x_1,h_1^{H,\mathbf{x}},\ldots,h_{i-1}^{H,\mathbf{x}},x_i\right)\qquad\text{for } 2\leq i\leq n.$$

We also use the notion of an n-stage quantum algorithm. An n-stage quantum algorithm S has the following syntactic behavior. At each stage it outputs a value x_i and then takes as input a value θ_i. Finally, it outputs some register Z. We denote such an interaction by $\mathbf{x},Z\leftarrow\langle\mathsf{S},\boldsymbol{\Theta}\rangle$. Using the above notations, we state the following lemma from [DFM20]

Lemma 5.7 (Reprogramming Multiple Values ([DFM20], Theorem 7, Remark 12))**.** *Let n be a positive integer, and let $\mathcal{X}_0, \mathcal{X}$ and $\mathcal{Y}$ be finite non-empty sets. There exists a black-box polynomial-time $(n+1)$-stage quantum algorithm S, satisfying the following property. Let A be an arbitrary oracle quantum algorithm that makes q queries to a uniformly random $H : \left(\mathcal{X}_0 \times (\mathcal{X} \times \mathcal{Y})^*\right) \to \mathcal{Y}$ and that outputs a tuple $\mathbf{x} = (x_0, x_1, \ldots, x_n) \in \mathcal{X}_0 \times \mathcal{X}^n$ and a (possibly quantum) output Z. Then, for any $\mathbf{x}^\circ \in \mathcal{X}_0 \times \mathcal{X}^n$ without duplicate entries and for any predicate V:*

$$\begin{aligned}&\Pr_{\boldsymbol{\Theta}}\left[\mathbf{x}{=}\mathbf{x}^\circ \wedge V\left(\mathbf{x}, \boldsymbol{\Theta}, Z\right) : (\mathbf{x}, Z) \leftarrow \left\langle \mathsf{S}^{\mathsf{A}}, \boldsymbol{\Theta}\right\rangle\right] \\ &\geq \frac{n!}{(q+n+1)^{2n}} \Pr_{H}\left[\mathbf{x}{=}\mathbf{x}^\circ \wedge V\left(\mathbf{x}, \mathbf{h}^{H,\mathbf{x}}, Z\right) : (\mathbf{x}, Z) \leftarrow \mathsf{A}^H\right] - \epsilon_{\mathbf{x}^\circ} \quad ,\end{aligned}$$

where $\epsilon_{\mathbf{x}^\circ}$ is equal to $\frac{n!}{|\mathcal{Y}|}$ when summed over all $\mathbf{x}^\circ$, and $\boldsymbol{\Theta}$ is sampled uniformly at random.

5.3.2 Construction

Assume some classical $2n+1$ message public-coin protocol $\langle \mathsf{P}, \mathsf{V}\rangle$ for a language $\mathcal{L}$. We denote the prover messages as $\{\alpha_i\}_{i\in[n+1]}$ and the verifier messages as $\{\beta_i\}_{i\in[n]}$. Then we present the modified protocol $\left\langle \mathsf{P}, \tilde{\mathsf{V}}\right\rangle$. Given some instance to prove, the prover P is exactly the prover of the original protocol. The changes to the verifier are described in 3

Algorithm 3: $\tilde{\mathsf{V}}\left(x; k\right)$

1 Use k as a key for a $\mathsf{PRF}_k\left(\cdot\right)$.
2 Set $\mathsf{ts} = x$
3 For $i \in [n]$:
4 Given a message α_i append it to ts, meaning $\mathsf{ts} = \mathsf{ts}||\alpha_i$
5 Compute $\beta_i = \mathsf{PRF}_k\left(ts\right)$ and send it to the prover.
6 Append β_i to ts, meaning $\mathsf{ts} = \mathsf{ts}||\beta_i$
7 Upon receiving α_n use V's predicate on the transcript $\mathsf{ts}||\alpha_n$. Accept if an only if it accepts.

We note that completeness and zero-knowledge proofs follow readily, and we focus on showing resettable soundness. To show resettable soundness, assume some resetting malicious quantum prover rP which convinces the honest modified verifier $\tilde{\mathsf{V}}$ with probability ε. We construct using it a malicious quantum prover $\tilde{\mathsf{P}}$ for the original protocol, with a related success probability.

First we note that due to the pseudo-randomness of PRF, we can replace $\tilde{\mathsf{V}}$'s aid of a PRF, with the aid of a random function R, with only negligible difference in success probability. We denote this modification by $\tilde{\mathsf{V}}^R$. More so, since $\tilde{\mathsf{V}}^R$ answers the messages themselves, we can replace oracle access $\tilde{\mathsf{V}}^R$ with oracle access to R. Formally, this implies,

$$\begin{aligned}&\Pr_{k}\left[\boldsymbol{\alpha}{=}\boldsymbol{\alpha}^\circ \wedge \Pi\left(\boldsymbol{\alpha}, \mathbf{h}^{\mathsf{PRF}_k,\boldsymbol{\alpha}}, Z\right) : (\boldsymbol{\alpha}, Z) \leftarrow \mathsf{rP}^{\tilde{\mathsf{V}}(x,\cdot;k)}\right] \geq \\ &\Pr_{R}\left[\boldsymbol{\alpha}{=}\boldsymbol{\alpha}^\circ \wedge \Pi\left(\boldsymbol{\alpha}, \mathbf{h}^{R,\boldsymbol{\alpha}}, Z\right) : (\boldsymbol{\alpha}, Z) \leftarrow \mathsf{rP}^{R}\right] - \mathrm{negl}\left(\lambda\right) \quad ,\end{aligned} \tag{3}$$

for any predicate Π.

Then, we note we can view a q-query rP as a quantum algorithm working on query registers and Z, E registers, Z is the output register (outputting an entire transcript), and E holds any internal qubits. Now, applying the multiple value reprogramming lemma (lemma 5.7) for rP^R implies the existence of a simulator algorithm S such that for any $\boldsymbol{\alpha}^\circ$ it holds that,

$$\Pr_{\boldsymbol{\beta}}\left[\boldsymbol{\alpha}=\boldsymbol{\alpha}^{\circ}\wedge\Pi_{\mathsf{V}}\left(\boldsymbol{\alpha},\boldsymbol{\beta},Z\right):\left(\boldsymbol{\alpha},Z\right)\leftarrow\left\langle\mathsf{S}^{\mathsf{rP}},\boldsymbol{\beta}\right\rangle\right]$$
$$\geq\frac{n!}{\left(q+n+1\right)^{2n}}\Pr_{R}\left[\boldsymbol{\alpha}=\boldsymbol{\alpha}^{\circ}\wedge\Pi_{\mathsf{V}}\left(\boldsymbol{\alpha},\mathbf{h}^{R,\boldsymbol{\alpha}},Z\right):\left(\boldsymbol{\alpha},Z\right)\leftarrow\mathsf{rP}^{R}\right]-\epsilon_{\mathbf{x}^{\circ}}\ ,$$

where the predicate Π_{V} is,

$$\sum_{\substack{\boldsymbol{\alpha},\boldsymbol{\beta},z\\ \mathsf{V}\left(x,z\right)=1,\,z\text{ is consistent with }\left(\boldsymbol{\alpha},\boldsymbol{\beta}\right)}}\left|\boldsymbol{\alpha}\right\rangle\left\langle\boldsymbol{\alpha}\right|\otimes\left|\boldsymbol{\beta}\right\rangle\left\langle\boldsymbol{\beta}\right|\otimes\left|z\right\rangle\left\langle z\right|\ .$$

We note that by consistent we mean that the the transcript z without the last message is exactly $\alpha_0,\beta_0,\ldots,\alpha_{n-1},\beta_{n-1}$. By summing over $\boldsymbol{\alpha}^{\circ}$ it holds,

$$\Pr_{\boldsymbol{\beta}}\left[\Pi_{\mathsf{V}}\left(\boldsymbol{\alpha},\boldsymbol{\beta},Z\right):\left(\boldsymbol{\alpha},Z\right)\leftarrow\left\langle\mathsf{S}^{\mathsf{rP}},\boldsymbol{\beta}\right\rangle\right]$$
$$\geq\frac{n!}{\left(q+n+1\right)^{2n}}\Pr_{R}\left[\Pi_{\mathsf{V}}\left(\boldsymbol{\alpha},\mathbf{h}^{R,\boldsymbol{\alpha}},Z\right):\left(\boldsymbol{\alpha},Z\right)\leftarrow\mathsf{rP}^{R}\right]-\frac{n!}{\left|\mathcal{Y}\right|}\underset{(3)}{\geq}$$
$$\frac{\varepsilon}{\mathrm{poly}\left(q\right)}-\mathrm{negl}\left(\lambda\right),$$

assuming $n=O\left(1\right)$. Finally, we note that S^{rP} is exactly a malicious prover for the original protocol. By running the $(n+1)$-stage algorithm S, sending each out as a message α_i and setting each returned message β_i as the input, we achieve a malicious prover for the original protocol with success probability of $\Omega\left(\frac{\varepsilon}{\mathrm{poly}q}\right)$.

5.4 Deterministic-Prefix Resetting Provers

We generalize proposition 5.1 and show that for any protocol can be modified to preserve completeness and (black-box) zero-knowledge while enabling to transform any resetting prover against modified protocol, to deterministic-prefix resetting prover against the original protocol. We will use this in section 6 in our construction of post-quantum constant-round resettably sound zero-knowledge argument.

5.4.1 Definitions

We define the notion of prefix oracle collection,

Definition 5.8 (Prefix Oracle Collection)**.** *A prefix oracle collection* $\mathbb{H}=\{(a,H_a)\,|\,a\in\mathcal{X}_0,H_a:\mathcal{X}\to\mathcal{Y}\}$ *is collection of oracles* H_a *from domain* $\mathcal{X}$ *to range* $\mathcal{Y}$*, parametrized by* $a\in\mathcal{X}_0$*.*

We can define oracle access to such a prefix oracle collection by operating on prefix query registers, where separate registers contain the prefix for the query and the query itself. We note here that we can view resetting prover, as querying an collection of prefix defined oracles such that each query to H is queried to the appropriate oracle $H_a\left(\cdot\right)=H\left(a,\cdot\right)$ (by splitting the original query register). We can also define reprogram such a prefix oracle collection such for a prefix a the oracle H_a is replaced by some arbitrary oracle $\bar{H}$. We denote this by $\mathbb{H}_{a\to\bar{H}}$.

5.4.2 Reducing to Single-Prefix Oracle Access

Using the above notations, we are able to show the following proposition which we use in section 6

Proposition 5.9. *Let $f(x,y,z;r)$ be a multivariate function for non-empty strings x,y,z,r. Consider the oracle aided variant f^R that given x,y,z outputs $f(x,y,z;R(x),R(x,y))$, where R is a random oracle.*

Assume some efficient q-query algorithm A^{f^R} with advice $|\psi\rangle$, such that A^{f^R} outputs x,y,z that satisfy some predicate Π with probability ε. Then there exists an efficient two-stage algorithm B such that B first outputs some value x. Then, given oracle access to $f^R|_x(y,z) = f(x,y,z;R(x),R(x,y))$ it outputs y,z. Then B's output, x,y,z, satisfies the predicate with probability of $\Omega\left(\frac{\varepsilon}{q^2}\right)$.

To show this, we generalize the single value reprogramming lemma (lemma 5.3) from [DFM20]. Formally we show,

Lemma 5.10 (Reprogramming a Single Prefix Oracle)**.** *Let A be a q-query quantum algorithm with advice $|\phi_0\rangle$ for a prefix oracle collection, with output registers A,Z. Then, for any such collection $\mathbb{H} = \{(a,H_a) : a \in \mathcal{X}_0, H_a : \mathcal{X} \to \mathcal{Y}\}$, any $a \in \mathcal{X}_0$, any function $\bar{H} : \mathcal{X} \to \mathcal{Y}$, and any projection $\Pi_{a,\bar{H}}$, it holds that*

$$\mathop{\mathbb{E}}_{i,b}\left[\left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+b\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+b}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\right\|_2^2\right] \geq \frac{\left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}|\phi^{\mathbb{H}_{a\to\bar{H}}}_q\rangle\right\|_2^2}{(2q+1)^2}\ ,$$

where the expectation is over uniform $(i,b) \in \{0,\ldots,q-1\}\times\{0,1\}\cup\{(q,0)\}$. We note that $|a\rangle\langle a|\otimes\mathsf{Id}$ is applied on the prefix query registers, where as $|a\rangle\langle a|\otimes\Pi_{a,\bar{H}}$ is applied on A,Z. Also, $|\phi^{\mathbb{H}}_i\rangle = \mathsf{A}^{\mathbb{H}}_{0\to i}|\phi_0\rangle$.

Proof. (of Lemma 5.10) First, note that for any $0 \leq i \leq q$ it holds that,

$$\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)((\mathsf{Id}-|a\rangle\langle a|)\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle = \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)((\mathsf{Id}-|a\rangle\langle a|)\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\ ,$$

where $(\mathsf{Id}-|a\rangle\langle a|)$ is applied to the prefix query register and Id is applied to the rest of the query. Hence, we can write,

$$\begin{aligned}
\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)|\phi^{\mathbb{H}}_{i+1}\rangle &= \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)|\phi^{\mathbb{H}}_i\rangle \\
&= \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)((\mathsf{Id}-|a\rangle\langle a|)\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle \qquad + \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle \\
&= \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)((\mathsf{Id}-|a\rangle\langle a|)\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle \qquad + \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle \\
&= \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)|\phi^{\mathbb{H}}_i\rangle - \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle \quad + \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\ .
\end{aligned}$$

Rearranging the terms, we can deduce the following identity for any $0 \leq i \leq q$,

$$\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)|\phi^{\mathbb{H}}_i\rangle = \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)|\phi^{\mathbb{H}}_{i+1}\rangle + \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle - \left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\ .$$

Applying $|a\rangle\langle a|\otimes\Pi_{a,\bar{H}}$ on both sides (where $|a\rangle\langle a|\otimes\Pi_{a,\bar{H}}$ act on A,Z), and using the triangle inequality, we can write,

$$\begin{aligned}
\left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)|\phi^{\mathbb{H}}_i\rangle\right\| \leq{}& \left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)|\phi^{\mathbb{H}}_{i+1}\rangle\right\| \\
&+ \left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i\to q}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\right\| \\
&+ \left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+1\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+1}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\right\|\ .
\end{aligned} \tag{4}$$

Applying equation 4 iteratively to the first summand, we can deduce,

$$\begin{aligned}
\left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{0\to q}\right)|\phi_0\rangle\right\| \leq{}& \left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}|\phi_q\rangle\right\| \\
&+ \sum_{\substack{0\leq i<q\\ b\in\{0,1\}}}\left\||a\rangle\langle a|\otimes\Pi_{a,\bar{H}}\left(\mathsf{A}^{\mathbb{H}_{a\to\bar{H}}}_{i+b\to q}\right)\left(\mathsf{A}^{\mathbb{H}}_{i\to i+b}\right)(|a\rangle\langle a|\otimes\mathsf{Id})\,|\phi^{\mathbb{H}}_i\rangle\right\|\ .
\end{aligned}$$

Now, we square, divide by $2q+1$ (the number of terms on the right hand side), and use Jensen inequality for $f(x) = x^2$ to deduce,

$$\frac{\left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} \left(\mathsf{A}_{0\to q}^{\mathbb{H}_{a\to\bar{H}}} \right) |\phi_0\rangle \right\|^2}{2q+1} \leq \left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} |\phi_q\rangle \right\|^2 + \sum_{\substack{0\leq i<q \\ b\in\{0,1\}}} \left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} \left(\mathsf{A}_{i+b\to q}^{\mathbb{H}_{a\to\bar{H}}} \right) \left(\mathsf{A}_{i\to i+b}^{\mathbb{H}} \right) \left(|a\rangle\langle a| \otimes \mathsf{Id} \right) |\phi_i^{\mathbb{H}}\rangle \right\|^2 .$$

Finally, note that

$$\left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} |\phi_q\rangle \right\|^2 = \left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} \left(\mathsf{A}_{q+0\to q}^{\mathbb{H}_{a\to\bar{H}}} \right) \left(\mathsf{A}_{i\to q+0}^{\mathbb{H}} \right) \left(|a\rangle\langle a| \otimes \mathsf{Id} \right) |\phi_0\rangle \right\|^2 .$$

Hence by dividing again by $2q+1$ we can deduce,

$$\mathop{\mathbb{E}}_{i,b} \left[\left\| |a\rangle\langle a| \otimes \Pi_{a,\bar{H}} \left(\mathsf{A}_{i+b\to q}^{\mathbb{H}_{a\to\bar{H}}} \right) \left(\mathsf{A}_{i\to i+b}^{\mathbb{H}} \right) \left(|a\rangle\langle a| \otimes \mathsf{Id} \right) |\phi_i^{\mathbb{H}}\rangle \right\|_2^2 \right] \geq \frac{\left\| \Pi_{x,\bar{H}} |\phi_q^{\mathbb{H}_{a\to\bar{H}}}\rangle \right\|_2^2}{(2q+1)^2} .$$

□

Proof (of Proposition 5.9). Assume that A outputs by measuring registers X, Y, Z. We can again view A as a prefix query algorithm for f^R, where the prefix is defined by the x part in the query register. We present B in 4,5.

Algorithm 4: B_1

1 Sample $(i,b) \leftarrow \{0,\ldots,q-1\} \times \{0,1\} \cup \{(q,0)\}$.
2 Sample $H \leftarrow \mathcal{H}$, where $\mathcal{H}$ is a family of $2(q+1)$-wise independent hashes.
3 Run $\mathsf{A}_{0\to i}^{f^H} |\psi_0\rangle$ where $|\psi_0\rangle$ is the advice for A. Denote the resulting state $|\psi_i^{f^H}\rangle$.
4 Measure the prefix query register to obtain a value x. Denote the state after measurement $|\phi_i^{f^H}(x)\rangle$.
5 Output x, and $i, b, H, |\phi_i^{f^H}(x)\rangle$.

Algorithm 5: $\mathsf{B}_2^{f^R|_x} \left(x, i, b, H, |\phi_i^{f^H}(x)\rangle \right)$

1 Run $\left(\mathsf{A}_{i+b\to q}^{f^H_{x\to f^R|_x}} \right) \left(\mathsf{A}_{i\to i+b}^{f^H} \right) |\phi_i^{f^H}(x)\rangle$, where $\mathsf{A}_{i+b\to q}^{f^H_{a\to f^R|_x}}$ is executed by controlled query to the oracle $f^R|_x$, which is controlled by if the first input is x_1. If not, use k to simulate.
2 Measure the output registers of A to obtain x', y, z. If $x' = x$ output y, z.

We note that we can move to a version of B where it uses a truly random function R instead of H the $2(q+1)$-wise independent hash [Zha12] without harming the success probability of B. We denote this version by B^R. Hence it holds,

$$\Pr\left[\Pi(x,y,z) \,|\, x,y,z \leftarrow \mathsf{B} \right] = \mathop{\mathbb{E}}_{R} \left[\Pr\left[\Pi(x,y,z) \,\middle|\, x,y,z \leftarrow \mathsf{B}^R \right] \right].$$

Following similar lines to Claim 5.4, it holds that,

$$\mathop{\mathbb{E}}_{R} \left[\Pr\left[\Pi(x,y,z) \,\middle|\, x,y,z \leftarrow \mathsf{B}^R \right] \right] = \sum_x \mathop{\mathbb{E}}_{R,R'} \left[\mathop{\mathbb{E}}_{i,b} \left[\left\| |x\rangle\langle x| \otimes \Pi_x(Y,Z) \left(\mathsf{A}_{i+b\to q}^{f^R_{x\to f^{R'}|_x}} \right) \left(\mathsf{A}^{f^R_{i\to i+b}} \left(|x\rangle\langle x| \otimes \mathsf{Id} \right) |\psi_i^{f^R}\rangle \right) \right\|^2 \right] \right] ,$$

where $(|x\rangle\langle x| \otimes \mathsf{Id})$ is applied on the prefix query registers, and $|x\rangle\langle x| \otimes \Pi_x(Y,Z)$ is applied to the output X,Y,Z registers. Also, $\Pi_x = \sum_{y,z:\Pi(x,y,z)=1} |y,z\rangle\langle y,z|$.

Using Lemma 5.10 for any fixed $x, \Pi_x, f^{R'}|_x, f^R$ (where f^R is viewed as a prefix oracle collection), it holds that,

$$\mathop{\mathbb{E}}_{i,b}\left[\left\||x\rangle\langle x| \otimes \Pi_x(Y,Z)\left(\mathsf{A}^{f^R_{x\to f^{R'}|_x}}_{i+b\to q}\right)\left(\mathsf{A}^{f^R_{i\to i+b}}(|x\rangle\langle x| \otimes \mathsf{Id})\,|\psi_i^{f^R}\rangle\right)\right\|^2\right] \geq$$
$$\frac{\left\|(|x\rangle\langle x|) \otimes \Pi_x(Y,Z)\,|\psi_q^{f^R_{x\to f^{R'}|_x}}\rangle\right\|^2}{(2q+1)^2}.$$

Hence,

$$\mathop{\mathbb{E}}_{R}\left[\Pr\left[\Pi(x,y,z)\,\middle|\, x,y,z \leftarrow \mathsf{B}^R\right]\right] \geq$$
$$\sum_x \mathop{\mathbb{E}}_{R,R'}\left[\frac{\left\|(|x\rangle\langle x|) \otimes \Pi_x(Y,Z)\,|\psi_q^{f^R_{x\to f^{R'}|_x}}\rangle\right\|^2}{(2q+1)^2}\right].$$

Then, we note that it holds that the following oracles f^R and $f^R_{x\to f^{R'}|_x}$ are perfectly indistinguishable for any x and truly random functions R, R'. Hence,

$$\mathop{\mathbb{E}}_{R}\left[\Pr\left[\Pi(x,y,z)\,\middle|\, x,y,z \leftarrow \mathsf{B}^R\right]\right] \geq \sum_x \mathop{\mathbb{E}}_{R}\left[\frac{\left\|(|x\rangle\langle x|) \otimes \Pi_x(Y,Z)\,|\psi_q^{f^R}\rangle\right\|^2}{(2q+1)^2}\right] =$$
$$\frac{\Pr\left[\Pi(x,y,z)\,\middle|\, x,y,z \leftarrow \mathsf{A}^{f^R}\right]}{(2q+1)^2}$$

□

We note that proposition 5.9 also implies the possibility to reduce a multi-input resetting prover to a regular resetting prover, assuming sub-exponentially hard PRFs. Formally,

Corollary 5.11. *Assuming sub-exponentially secure PRFs, any resettably sound protocol can be transformed into a multi-input resettably sound one.*

Proof (Sketch). We sketch the outlines of this proof in Appendix A.2. □

6 A Post-Quantum Resettably Sound Zero Knowledge Protocol

In this section we present a post-quantum resettably-sound zero-knowledge protocol. The protocol is also constant-round.

Ingredients and Notation:

- A post-quantum pseudorandom function PRF.
- A post-quantum non-interactive commitment scheme Com.
- A post-quantum compute and compare obfuscator Obf.

- A quantum fully-homomorphic encryption scheme ($\mathsf{QFHE.Gen}, \mathsf{QFHE.Enc}, \mathsf{QFHE.QEnc}, \mathsf{QFHE.Dec}, \mathsf{QFHE.QDec}, \mathsf{QFHE.Eval}$).
- A delayed-input 3-message post-quantum WI proof $(\mathsf{WI.P}, \mathsf{WI.V})$ for **NP**.
- A delayed-input 4-message sub-exponential statistical WI argument system $(\mathsf{sWI.P}, \mathsf{sWI.V})$ for **NP**.
- A 2-message post-quantum input hiding, sub-exponentially statistically function hiding secure function evaluation scheme ($\mathsf{SFE.Gen}$, $\mathsf{SFE.Enc}$, $\mathsf{SFE.Eval}$, $\mathsf{SFE.Dec}$).
- Denote by $\varepsilon \in (0,1)$ a constant such that both the 4-message WI and SFE have sub-exponential statistical security with respect to (in the statistical indistinguishability guarantee in both primitives, the statistical distance is bounded by $O(2^{-\lambda^{\varepsilon}})$).

The protocol is described in Figure 1.

6.1 Quantum Resettable Soundness

Proposition 6.1 (The protocol has quantum resettable soundness)**.** *Let V be the verifier from Protocol 1 and let $\mathsf{V}(x,\cdot;r)$ be the next message function of the verifier, conditioned on the instance being x and the randomness of the verifier being r. For any quantum polynomial-size resetting prover $\mathsf{rP} = \{\mathsf{rP}_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ there exists a negligible function $\mu(\cdot)$ such that for any security parameter $\lambda \in \mathbb{N}$ and $x \in \{0,1\}^\lambda \setminus \mathcal{L}$,*

$$\Pr_r\left[\mathsf{V}(x,\mathsf{ts};r)=1 \,\middle|\, \mathsf{ts} \leftarrow \mathsf{rP}_\lambda^{\mathsf{V}(x,\cdot;r)}(\rho_\lambda)=1\right] \leq \mu(\lambda) \ .$$

Proof. Let $\mathsf{rP} = \{\mathsf{rP}_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ a polynomial-size quantum prover and let $x = \{x_\lambda\}_{\lambda\in\mathbb{N}}$ be a sequence such that $\forall\lambda\in\mathbb{N} : x_\lambda \in \{0,1\}^\lambda \setminus \mathcal{L}$. Denote by ε_λ the probability that rP_λ breaks resettable soundness. We prove soundness by a hybrid argument. We consider a series of hybrid processes with output over $\{0,1\}$, starting from $\mathsf{V}(x,\mathsf{ts};r)$ the output bit distribution of V after receiving the transcript ts from the malicious rP, where rP has quantum oracle access to $\mathsf{V}(x,\cdot;r)$, the verifier next message function. The proof will show that the probability to output 1 is negligible, which proves the soundness of the protocol.

Define by Hyb_0 the process of interaction between $\mathsf{rP}^{\mathsf{V}(x,\cdot;r)}$ and the honest verifier, where the output of Hyb_0 is $\mathsf{V}(x,\mathsf{ts};r)$, and define by Hyb_1 the same process only that the verifier's PRF is swapped with a random function. By the security of the PRF, the outputs of these two processes are computationally indistinguishable.

Consider the next message function of the verifier in the process Hyb_1. $\mathsf{V}(\cdot)$ gets m_1, m_2, m_3 and randomness, where m_1 is the first prover message, m_2 is the prover SFE encryption ct_P and statistical WI second message $\beta_\mathbf{s}$ from step 2b, and m_3 is the rest of the prover's messages (in case we want to give the function only a prefix of the transcript, the messages not included are defined to be $\perp$). The randomness of V for its first message is generated by applying a random function only to m_1, and the randomness for the second (and last) verifier message is by applying a random function only to (m_1, m_2). Considering Proposition 5.9, we can think of the function f from the proposition as our verifier's next message function, on x, y, z from the proposition as m_1, m_2, m_3 and on the predicate Π as the verifier's verdict. It follows by the Proposition that for every prover that breaks resettable soundness with probability ϵ there is a different type of prover that first sends its first message, and only then gets oracle access to the verifier's next message function, executable from the verifier's second message. The Proposition guarantees that such prover breaks resettable soundness with probability at least $\Omega\left(\frac{\varepsilon}{q^2}\right)$, where q is an upper bound on the running time of the original prover. Since the prover is polynomial-time, this probability is noticeable if ε is.

So, we assume without the loss of generality that rP is of the form described above. Since the prover does not see any of the verifier's information before sending its first message, we can, by an averaging argument over the prover's quantum measurements, fix the prover's first message and quantum advice to be deterministic. The reason is that we can take the first message and quantum advice that maximize the prover's probability to successfully cheat. Our setting from now on is thus one where the prover rP sends its

Protocol 1

Common Input: An instance $x \in \mathcal{L}$, security parameter $\lambda := |x|$. Below we denote $\bar{\lambda} = \lambda^{2/\varepsilon}$.
P's private input: A classical witness $w \in \mathcal{R}_{\mathcal{L}}(x)$ for x.

1. **Prover Commitment:** P sends the following,
   - Non-interactive commitments to the witness, and two strings of zeros of length $\bar{\lambda}$:
   $$\mathsf{cmt}_1 \leftarrow \mathsf{Com}(1^\lambda, w), \;\; \mathsf{cmt}_2 \leftarrow \mathsf{Com}(1^\lambda, 0^{\bar{\lambda}}), \;\; \mathsf{cmt}_3 \leftarrow \mathsf{Com}(1^\lambda, 0^{\bar{\lambda}}) \;\; .$$
   - Two independent first messages α_1, α_2 for two independent executions of 3-message, delayed-input WI proofs $(\mathsf{WI.P}, \mathsf{WI.V})$.
   - First message h of a 4-message delayed-input statistical WI argument $(\mathsf{sWI.P}, \mathsf{sWI.V})$, with security parameter $\bar{\lambda}$.
2. **Extractable Commitment to Verifier Secret:** V samples a PRF seed $s \leftarrow \{0,1\}^\lambda$. V's randomness for the first message is generated by applying $\mathsf{PRF}_s(\cdot)$ to the first prover message.
   (a) V computes $u \leftarrow \{0,1\}^\lambda$, $v \leftarrow \{0,1\}^\lambda$, $(\mathsf{pk}, \mathsf{sk}) \leftarrow \mathsf{QFHE.Gen}(1^\lambda)$. V sends
   $$\mathsf{pk}, \;\; \mathsf{ct}_{\mathsf{V}} \leftarrow \mathsf{QFHE.Enc}_{\mathsf{pk}}(u), \;\; \widetilde{\mathbf{CC}} \leftarrow \mathsf{Obf}\Big(\mathbf{CC}\big[\mathsf{QFHE.Dec}_{\mathsf{sk}}(\cdot), v, \mathsf{sk}\big]\Big) \;\; .$$
   V also sends β_1, β_2 following α_1, α_2, and $\alpha_{\mathbf{s}}$ following h.
   (b) P sends,
   - $\mathsf{ct}_{\mathsf{P}} \leftarrow \mathsf{SFE.Enc}(1^{\bar{\lambda}}; 0^\lambda)$ an encryption of 0^λ encrypted with security parameter $\bar{\lambda}$.
   - $\beta_{\mathbf{s}}$ for $h, \alpha_{\mathbf{s}}$ as the last message of $\mathsf{sWI.V}$ in the 4-message WI protocol.
   - A WI proof γ_1, following α_1 and β_1, that $x \in \mathcal{L}$ or, (1) the randomness used to generate ct_{P} is the content of cmt_2[a], and (2) the randomness for $h, \beta_{\mathbf{s}}$ is the content of cmt_3.

   (c) V applies $\mathsf{PRF}_s(\cdot)$ to $(\mathsf{ct}_{\mathsf{P}}, \beta_{\mathbf{s}},$ Prover's first message$)$ to generate randomness for its current message. It sends,
   - $\hat{\mathsf{ct}} \leftarrow \mathsf{SFE.Eval}\Big(\mathbf{CC}\big[\mathsf{Id}(\cdot), u, v\big], \mathsf{ct}_{\mathsf{P}}\Big)$ executed with security parameter $\bar{\lambda}$, where $\mathsf{Id}(\cdot)$ is the identity function.
   - $\gamma_{\mathbf{s}}$, for $h, \alpha_{\mathbf{s}}, \beta_{\mathbf{s}}$, proving that the transcript of the verifier so far is explainable or, cmt_1 is a commitment to a non-witness $z \notin \mathcal{R}_{\mathcal{L}}(x)$. The witness that V uses for the proof is its randomness, that proves that the transcript is explainable.
3. **Final WI by the Prover:** P sends γ_2 which proves that $x \in \mathcal{L}$ or, that cmt_1 is a valid commitment and there exists a string c such that $\widetilde{\mathbf{CC}}(c) \neq \bot$. The witness that P uses for its proofs γ_1, γ_2 is w, which proves $x \in \mathcal{L}$.
4. **Acceptance:** V accepts if the WI statements by the prover are verified.
5. **Aborts:** During the protocol, if either party does not respond, sends a message of an incorrect form or provides a non-convincing WI proof it considered as an abort, and the other party terminates the interaction.

[a] Formally, there are strings r_1, r_2, r_3 such that $\mathsf{ct}_{\mathsf{P}} = \mathsf{SFE.Enc}(r_3; r_2)$, $\mathsf{cmt}_2 = \mathsf{Com}(1^\lambda, r_2; r_1)$.

Figure 1: A post-quantum classical constant-round zero-knowledge argument for $\mathcal{L} \in \mathbf{NP}$, with quantumly-resettable soundness.

first message, the verifier samples true randomness r_1 for its first message and sends that first message to the prover. From that point on, the prover has oracle access to the verifier's next message function, conditioned on that the randomness for the first message was r_1.

Now, by the statistical soundness of the WI proofs that rP gives, it follows that for the first WI messages α_1, α_2 sent in the prover's first message, with overwhelming probability over the randomness in generating the second messages β_1, β_2 (i.e. with overwhelming probability over choosing r_1), there does not exist a false statement along with a proof γ that can be accepted by the verifier of the WI. It follows that unless ε is negligible (in which case, our proof ends), since the statement $x \in \mathcal{L}$ is incorrect in the case of a cheating prover, in the first prover message, the commitments $\mathsf{cmt}_1, \mathsf{cmt}_2, \mathsf{cmt}_3$ are all valid commitments.

Observe that because cmt_1, cmt_2 are consistent with the prover's WI statement, cmt_1 is necessarily a commitment to a non-witness $z \notin \mathcal{R}_{\mathcal{L}}(x)$, and denote by r_z a string s.t. $\mathsf{cmt}_1 = \mathsf{Com}(1^\lambda, z; r_z)$ and by r_{SFE} the string s.t. $\mathsf{cmt}_2 \leftarrow \mathsf{Com}(1^\lambda, r_{\mathsf{SFE}})$. Notice that since these commitments are fixed, we can obtain z, r_z, r_{SFE} as non-uniform classical advice.

Define the following hybrid distributions.

- Hyb_2 : This hybrid process is identical to Hyb_1, with the exception that when the verifier gives its WI (step 2c), V uses the information (z, r_z) as witness for its WI statement, instead of the witness that shows its transcript is explainable.

- Hyb_3 : This hybrid process is identical to Hyb_2, except that in step 2c when the verifier responds with an SFE evaluation, instead of performing an SFE evaluation of the circuit $\mathbf{CC}[\mathsf{Id}(\cdot), u, v]$, V performs an SFE evaluation of $C_\perp$, a circuit that always outputs $\perp$.

- Hyb_4 : This hybrid process is identical to Hyb_3, except that the verifier's randomness for its second message is generated by a PRF and not by a random function. The randomness generated for the first message is still truly random.

- Hyb_5 : This hybrid process is identical to Hyb_4, except that in the verifier's first message in step 2a, instead of sending an actual CC obfuscation, the verifier executes $\mathbf{Sim}^{\mathrm{CC}}$ (from the simulation property of the CC obfuscation) and sends $\mathbf{Sim}^{\mathrm{CC}}(1^{|\mathsf{QFHE.Dec}|}, 1^{|\mathsf{sk}|}, 1^\lambda)$.

Note that $\mathbf{Sim}^{\mathrm{CC}}(1^{|\mathsf{QFHE.Dec}|}, 1^{|\mathsf{sk}|}, 1^\lambda)$ always outputs $\perp$ with overwhelming probability over the randomness of $\mathbf{Sim}^{\mathrm{CC}}$. Thus, by the soundness of the prover's last WI proof in step 3, the probability for the prover to make the verifier accept in Hyb_5 is at most negligible as its last WI statement is necessarily false. To finish the proof it remains to explain why each consecutive pair of the distributions above are statistically indistinguishable (recall that for a pair of distributions over a single bit, they are statistically indistinguishable iff they are computationally indistinguishable).

- $\mathsf{Hyb}_1 \approx_s \mathsf{Hyb}_2$: Observe that due to the fact that the prover committed to its randomness strings (r_{SFE} in cmt_2 and r_{sWI} in cmt_3) in the beginning of the protocol, the number of *possible* messages for the oracle $\mathsf{V}(x, \cdot; r)$ in later stages of the protocol, is restricted.

  More precisely, first, by the soundness of the prover's WI proofs, with overwhelming probability over the randomness for β_1, β_2, rP can send γ_i (for $i \in 1, 2$) that proves a false WI statement with probability 0. Now, this means that the only messages that $\mathsf{V}(x, \cdot; r)$ will respond to are ones that satisfy the WI statements which in turn means that the SFE encryption sent by rP at step 2b is restricted to have the randomness r_{SFE} and β_{s} is restricted to use the randomness r_{sWI} and thus has a single option. Accordingly, the only places where the prover has remaining freedom in choosing its message in step 2b (that is, without the message being rejected by the verifier), is the content of the SFE encryption and the randomness associated with generating the proof γ_1. While the prover has freedom in generating its proof γ_1, the verifier does not apply the random function to this part. Finally, this implies that the only places where the prover has the ability to influence the verifier's randomness for its second message, is by changing the content of the SFE encryption that it sends in step 2b. This is an encryption of a λ-bit string, and thus there are 2^λ many options.

It is now remains to perform a hybrid argument of 2^λ steps. In each step we change the verifier's WI argument response $\gamma_{\mathbf{s}}$ given a prover's valid transcript so far T_i (for $i \in [2^\lambda]$). Specifically, at step i we change the verifier's behavior when given the prover's SFE encryption for he string i: We swap the witness that is used to generate $\gamma_{\mathbf{s}}$, from the witness that proves the verifier's transcript is explainable, to the witness (z, r_z) that proves that cmt_1 contains a non-witness. Note that since the verifier applies a random function on the prover's SFE encryption ct_P and $\beta_{\mathbf{s}}$ to generate its randomness for the second message, then between each T_i's the distributions on $\gamma_{\mathbf{s}}$ are independent of each other, thus we can indeed perform the hybrid argument. By the sub-exponential statistical security of the 4-message WI of the verifier, when executed with security parameter $\bar{\lambda}$ the statistical distance between each consecutive hybrids is bounded by $O(2^{-(\bar{\lambda})^\varepsilon}) = O(2^{-\lambda^{c\cdot\varepsilon}}) \leq O(2^{-\lambda^2})$. Since we have 2^λ hybrids, the overall statistical distance is bounded by $2^\lambda \cdot O(2^{-\lambda^2}) = O(2^{-\lambda^2})$.

- $\mathsf{Hyb}_2 \approx_s \mathsf{Hyb}_3$: This indistinguishability is established in Claim 6.2.
- $\mathsf{Hyb}_3 \approx_s \mathsf{Hyb}_4$: This indistinguishability follows directly from the security of the PRF.
- $\mathsf{Hyb}_4 \approx_s \mathsf{Hyb}_5$: This indistinguishability follows directly from the security of the CC obfuscation and the fact that the target v is chosen uniformly and at random, and either of the processes Hyb_4, Hyb_5 uses the value v after the CC obfuscation.

□

Claim 6.2. *Let* Hyb_2, Hyb_3 *be the hybrid processes described in Proposition 6.1. Then,*

$$\mathsf{Hyb}_2 \approx_s \mathsf{Hyb}_3 \ .$$

Proof. The proof will be based on the SFE statistical circuit privacy, the computational security of the QFHE, and on the computational security of the CC obfuscation.

First we note that by the soundness of the prover's WI proof, with overwhelming probability over choosing the verifier's randomness, that the message ct_P sent from the prover to $\mathsf{V}(x, \cdot; r)$ at step 2b is always a valid SFE encryption with the randomness r_{SFE} (or it is rejected by the verifier). Let dk the secret SFE derived from the randomness r_{SFE}.

Consider the process of prover's execution in Hyb_2, which has at most q steps for some polynomial q. Consider the random variable J over $[q]$ which denotes the first place where the prover sends ct_P such that $u = \mathsf{SFE.Dec}_{\mathsf{dk}}(\mathsf{ct}_\mathsf{P})$ (define $J = q + 1$ if such time step does not exist), this ciphertext ct_P can be sent to either the verifier next message function or the actual verifier in the transcript ts. One can ask what is the probability that $J = j'$ for some $j' \in [q + 1]$, and we denote by j the minimal $j \in [q + 1]$ such that the probability that $J = j$, is noticeable (the fact that the distribution is noticeable is well defined when we recall that the interaction between the prover and verifier defines an asymptotic sequence of experiments). Note that for all $i < j$, because only with a negligible probability the prover sends in query i an SFE ciphertext such that $u = \mathsf{SFE.Dec}_{\mathsf{dk}}(\mathsf{ct}_\mathsf{P})$, we can change these cases: Until before step j, for each of the prover's queries, before we apply the verifier's next message function, we project the quantum state, which is a superposition of classical queries, to being a superposition of queries such that $u \neq \mathsf{SFE.Dec}_{\mathsf{dk}}(\mathsf{ct}_\mathsf{P})$. To do this efficiently we use dk to check the content of the encryption, output 0 or 1 on a register on the side with accordance to whether the content was u or not, and then measure this 1-qubit register. Such process, denoted Hyb'_2, has negligible statistical distance to Hyb_2 due to the fact that for each of the projections, the projected state and the original state have negligible trace distance between them (this follows in turn from the fact that for the indices before i, the prover had only a negligible probability to send ct_P such that $u = \mathsf{SFE.Dec}_{\mathsf{dk}}(\mathsf{ct}_\mathsf{P})$).

Consider the prover's execution in Hyb'_2 until just before time step j. Since the prover's SFE encryption ct_P is valid and can only be encrypted using the secret key dk it follows that the content of ct_P is always defined, and because it cannot be an SFE encryption of the correct u, it is the case that the output of the circuit $C_{u\to v}$ on the content of the prover's encryption is always $\perp$. This means that by the circuit privacy of the SFE we can evaluate the prover's SFE ciphertext with the circuit $C_\perp$ that always outputs $\perp$ instead of the circuit $C_{u\to v}$.

Now to be more precise, we will modify the verifier's SFE evaluations until before step j and call this process Hyb'_3. By the exact same statistical hybrid argument described in the indistinguishability $\mathsf{Hyb}_1 \approx_s \mathsf{Hyb}_2$ in Proposition 6.1, we perform a hybrid argument with 2^λ steps, this time swapping each of the SFE evaluations of $\mathsf{V}(x,\cdot;r)$ until before step j (i.e. for each $i \in [2^\lambda]$, the SFE evaluation for the SFE encryption that contains i) to use the circuit $C_\perp$ rather than the circuit $C_{u\to v}$. Since the SFE security parameter we are using is $\bar{\lambda} := \lambda^{2/\varepsilon}$, similarly to the indistinguishability $\mathsf{Hyb}_1 \approx_s \mathsf{Hyb}_2$, the overall statistical distance between Hyb'_2 and Hyb'_3 is $O(2^{-\lambda^2})$.

We can now define $\tilde{\mathsf{Hyb}}_3$ and further remove the check on ct_P, and not project the prover's queries before time step j to ciphertexts such that $u \neq \mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P})$. Due to the exact same argument as to why Hyb_2 and Hyb'_2 are statistically close, $\tilde{\mathsf{Hyb}}_3$ and Hyb'_3 are statistically close.

Notice that if $j = q+1$ and there is no time step in the prover's execution where it sends a ciphertext containing u with a noticeable probability, then $\tilde{\mathsf{Hyb}}_3 = \mathsf{Hyb}_3$ and we are done. We end our proof by showing that it is necessarily the case that $j = q+1$, by showing that otherwise, we can use the prover to break the computational security of the QFHE.

Observe that the security of the QFHE implies that for every efficient quantum adversary $\mathsf{A}^* = \{\mathsf{A}^*_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$, the probability that A^* finds u given $\mathsf{pk}, \mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_\mathsf{pk}(u)$ for a uniformly random $u \leftarrow \{0,1\}^\lambda$, is negligible - we will assume toward contradiction that rP sends ct_P s.t. $u = \mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P})$ with a noticeable probability at step $j \in [q]$, and get a contradiction with the last property about the hardness of finding a random encrypted u.

Using rP and the fact that $u = \mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P})$ with noticeable probability in step j, we now describe a (non-uniform) algorithm A^* that finds u given $\mathsf{pk}, \mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_\mathsf{pk}(u)$ for $u \leftarrow \{0,1\}_\lambda$ and thus breaks the security of the QFHE. As part of the non-uniform advice of A^*, it will have the secret SFE key dk, which is fixed. First, we swap the setting to an efficient one, where the verifier uses a PRF for its randomness. The processes are indistinguishable by the security of the PRF. Given $\mathsf{pk}, \mathsf{ct} \leftarrow \mathsf{QFHE.Enc}_\mathsf{pk}(u)$, the algorithm A^* will act as the verifier $\mathsf{V}(x,\cdot;r)$ with one change - it will use the simulator $\mathbf{Sim}^{\mathrm{CC}}$ (from the simulation property of the CC obfuscation) and send to rP the following as the first verifier message,

$$\mathsf{pk},\ \mathsf{ct},\ \mathbf{Sim}^{\mathrm{CC}}(1^{|\mathsf{QFHE.Dec}|}, 1^{|\mathsf{sk}|}, 1^\lambda), \alpha_\mathbf{s}, \beta_1, \beta_2\ ,$$

and from that moment on will act as the verifier in $\tilde{\mathsf{Hyb}}_3$ (only that it uses a PRF to generate its randomness and not a truly random function). At step j, A^* will measure the prover's ciphertext and output $\mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P})$.

We now use the simulation property guarantee of the CC obfuscation: Note that the probability that rP outputs ct_P s.t. $\mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P}) = u$ in the simulated setting, where A^* sends $\mathbf{Sim}^{\mathrm{CC}}(1^{|\mathsf{QFHE.Dec}|}, 1^{|\mathsf{sk}|}, 1^\lambda)$ instead of $\widetilde{\mathbf{CC}}$, is negligibly close to the probability that it outputs ct_P s.t. $\mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P}) = u$ in the regular setting where it gets $\widetilde{\mathbf{CC}}$ - this is due to the security of the CC obfuscator. Because we know that rP sends ct_P s.t. $u = \mathsf{SFE.Dec}_\mathsf{dk}(\mathsf{ct}_\mathsf{P})$ with a noticeable probability in step j, it follows that A^* outputs u with the same probability, in contradiction. □

6.2 Quantum Zero-Knowledge

We now prove that our protocol is quantum zero-knowledge (with respect to Definition 3.2). That is, we next describe our simulator and then prove that the simulation is indistinguishable from the output state of V^* in the real interaction.

The proof for ZK follows very similar lines to those of [BS20]. We include the proof for completeness. The simulator Sim is constructed from numerous subroutines described first. In what follows V^* is an arbitrary quantum polynomial-size circuit, $x \in \mathcal{L}$, and ρ is a polynomial-size mixed quantum state as auxiliary input for the verifier.

$\mathsf{Sim}_\mathrm{a}(x, \mathsf{V}^*, \rho)$:

1. **Simulator Actions:** Sim_a interacts with V^* as the honest prover P until the end of the verifier's WI proof (in step 2c of the protocol) with exactly these changes:

- In the beginning of the protocol when the prover sends commitments (step 1), cmt_1 is a commitment to $0^{|w|}$ instead of a commitment to the actual witness w and cmt_2, cmt_3 are commitments to randomness strings r_{SFE}, r_{sWI} rather than to $0^{\bar{\lambda}}$ and $0^{\bar{\lambda}}$.
- When the prover computes its SFE encryption in step 2b and its WI message $\beta_{\mathbf{s}}$ in step 2b it uses the randomness r_{SFE} and r_{sWI} respectively.
- In the first WI proof that the prover gives, the simulator uses a different witness. Specifically, for the first WI statement it uses the witness that proves that (1) the randomness for generating ct_{P} is the content of cmt_2 and (2) the randomness used for generating the messages $h, \beta_{\mathbf{s}}$ in the 4-message WI is the content of cmt_3.

2. **Simulation Verdict:** If at some point V^* aborts or fails in its WI proof until the end of step 2c, $\mathsf{Sim}_{\mathrm{a}}$ outputs the aborting verifier's output. Otherwise, $\mathsf{Sim}_{\mathrm{a}}$ outputs `Fail`.

$\mathsf{Sim}_{\mathrm{na}}(x, \mathsf{V}^*, \rho)$:

1. **Simulation of Initial Commitments and Verifier Message:**

   (a) $\mathsf{Sim}_{\mathrm{na}}$ samples randomness r_{SFE}, r_{sWI} for the SFE and WI with security parameter $\bar{\lambda}$ and sends to V^* the commitments $\mathsf{cmt}_1 \leftarrow \mathsf{Com}(1^\lambda, 0^{|w|})$, $\mathsf{cmt}_2 \leftarrow \mathsf{Com}(1^\lambda, r_{\mathsf{SFE}})$, $\mathsf{cmt}_3 \leftarrow \mathsf{Com}(1^\lambda, r_{\mathsf{sWI}})$. Additionally, the simulator sends honestly generated α_1, α_2 for the upcoming WI proofs, and honestly generated h for the verifier's 4-message statistical WI argument.

   (b) V^* sends pk, $\mathsf{ct}_{\mathsf{V}^*}$, $\widetilde{\mathbf{CC}}$, β_1, β_2 and $\alpha_{\mathbf{s}}$.

2. **Extraction Attempt:**

   (a) $\mathsf{Sim}_{\mathrm{na}}$ first encrypts $\rho^{(1)}$, the inner (quantum) state of the verifier after its first message:

   $$\mathsf{ct}_{\rho^{(1)}} \leftarrow \mathsf{QFHE.QEnc}_{\mathsf{pk}}(\rho^{(1)}) \ .$$

   Let dk the SFE secret key derived from the randomness r_{SFE}. Consider the unified ciphertext $\mathsf{ct}_{\mathsf{V}^*,\rho^{(1)}} = \mathsf{ct}_{\mathsf{V}^*}, \mathsf{ct}_{\rho^{(1)}}$ which encrypts some $\tilde{u}$ on the left (if $\mathsf{ct}_{\mathsf{V}^*}$ is a valid ciphertext) and the state $\rho^{(1)}$ on the right. $\mathsf{Sim}_{\mathrm{na}}$ performs quantum homomorphic evaluation on $\mathsf{ct}_{\mathsf{V}^*,\rho^{(1)}}$ of the circuit $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$, which has four steps:

   i. Encrypts $\tilde{u}$ with $\mathsf{SFE.Enc}_{\mathsf{dk}}(\cdot)$ to get a ciphertext $\mathsf{ct}_{\mathsf{Sim}}$.

   ii. Computes an honestly generated $\beta_{\mathbf{s}}$ for the verifier's WI.

   iii. Computes a WI proof γ_1 for the prover's statement in step 2b of the protocol. The witness used is for proving that the randomness used to generate $\mathsf{ct}_{\mathsf{Sim}}$, $\beta_{\mathbf{s}}$ is the content of cmt_2, cmt_3 respectively (rather than using the witness that shows that $x \in \mathcal{L}$).

   iv. Executes V^* on the entire $\mathsf{ct}_{\mathsf{Sim}}, \beta_{\mathbf{s}}, \gamma_1, \rho^{(1)}$.

   v. Let the output of V^* be $\hat{\mathsf{ct}}_{\mathsf{Sim}}, \gamma_{\mathbf{s}}$ in the interaction register and $\rho^{(2)}$ in the inner state register. Then the circuit decrypts $\tilde{v} := \mathsf{SFE.Dec}_{\mathsf{dk}}(\hat{\mathsf{ct}}_{\mathsf{Sim}})$.

   Given input $\tilde{u}, \rho^{(1)}$, the output of $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$ is $(\mathsf{ct}_{\mathsf{Sim}}, \beta_{\mathbf{s}}, \gamma_1, \hat{\mathsf{ct}}_{\mathsf{Sim}}, \gamma_{\mathbf{s}}, \tilde{v}, \rho^{(2)})$.

   (b) Given the QFHE encryption of the output $(\mathsf{ct}_{\mathsf{Sim}}, \beta_{\mathbf{s}}, \gamma_1, \hat{\mathsf{ct}}_{\mathsf{Sim}}, \gamma_{\mathbf{s}}, \tilde{v}, \rho^{(2)})$ of $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$, let $\mathsf{ct}_{\tilde{v}}$ be the encryption of the part that contains $\tilde{v}$. $\mathsf{Sim}_{\mathrm{na}}$ computes $\mathsf{sk}' = \widetilde{\mathbf{CC}}(\mathsf{ct}_{\tilde{v}})$. If $\mathsf{sk}' = \bot$ then it halts simulation and outputs `Fail`. Otherwise, $\mathsf{Sim}_{\mathrm{na}}$ decrypts the rest of the evaluated ciphertext using sk' to get $\mathsf{ct}_{\mathsf{Sim}}, \beta_{\mathbf{s}}, \gamma_1, \hat{\mathsf{ct}}_{\mathsf{Sim}}, \gamma_{\mathbf{s}}, \rho^{(2)}$. If the proof $h, \alpha_{\mathbf{s}}, \beta_{\mathbf{s}}, \gamma_{\mathbf{s}}$ fails to prove the verifier's WI statement, the simulation fails and the output is `Fail`.

3. **Simulation of the Prover's WI Proof:** $\mathsf{Sim}_{\mathrm{na}}$ gives V^* a WI proof γ_2 using the witness for the second statement, that shows that there exists an input c to $\widetilde{\mathbf{CC}}$ such that $\widetilde{\mathbf{CC}}(c) \neq \bot$, and that cmt_1 is a valid commitment. The witness used is the randomness for the commitment cmt_1 and $c := \mathsf{ct}_{\tilde{v}}$.

4. **Simulation Verdict:** If V^* completed interaction without aborting and gave a convincing WI proof, $\mathsf{Sim}_{\mathrm{na}}$ outputs the verifier's output. Otherwise, $\mathsf{Sim}_{\mathrm{na}}$ outputs `Fail`.

Proof of Simulation Validity. We now turn to prove that there is a simulator Sim such that the simulated output $\mathsf{Sim}(x, \mathsf{V}^*, \rho)$ is computationally indistinguishable from $\mathsf{OUT}_{\mathsf{V}^*}\langle \mathsf{P}, \mathsf{V}^*(\rho)\rangle(x)$. This is done in several steps:

1. **Simulating aborting interactions:** Let $\mathsf{V}^*_{\mathrm{a}}$ be the augmented verifier that is identical to V^*, with the exception that if V^* does not abort, $\mathsf{V}^*_{\mathrm{a}}$ outputs `Fail`. Then the output of $\mathsf{Sim}_{\mathrm{a}}$ is indistinguishable from the output of $\mathsf{V}^*_{\mathrm{a}}$ in a real interaction.

2. **Simulating non-aborting interactions:** Let $\mathsf{V}^*_{\mathrm{na}}$ be the augmented verifier that is identical to V^*, with the exception that if V^* aborts, $\mathsf{V}^*_{\mathrm{na}}$ outputs `Fail`. Then the output of $\mathsf{Sim}_{\mathrm{na}}$ is indistinguishable from the output of $\mathsf{V}^*_{\mathrm{na}}$ in a real interaction.

3. **Combining simulators:** We use previous results from [BS20] to combine the two simulators into one Sim that successfully simulates every verifier.

Proposition 6.3 (Similarity of aborting part)**.** *Let $\mathsf{V}^* = \{\mathsf{V}^*_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ a polynomial-size quantum verifier, and let $\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{a}}}$ be the verifier's output at the end of protocol such that if V^* does not abort, the output is* `Fail`*. Then,*

$$\{\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{a}}}\langle \mathsf{P}(w), \mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}_{\mathrm{a}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w} \ ,$$

where $\lambda \in \mathbb{N}$, $x \in \mathcal{L} \cap \{0,1\}^\lambda$, $w \in \mathcal{R}_{\mathcal{L}}(x)$.

Proof. We establish the above indistinguishability by a hybrid argument. The first hybrid Hyb_0 is the simulation process $\mathsf{Sim}_{\mathrm{a}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w}$. For the next hybrids we let the simulator hold the witness w. The next hybrid Hyb_1 is to change the witness used in the first WI proof by the prover, to use the witness w that proves $x \in \mathcal{L}$. These hybrids are computationally indistinguishable by the computational witness-indistinguishability property. Finally, Hyb_2 is identical to Hyb_1 with the exception that cmt_1 is a commitment to the witness w and not to $0^{|w|}$, and $\mathsf{cmt}_2, \mathsf{cmt}_3$ are commitments to strings of zeros of the same lengths as $r_{\mathsf{SFE}}, r_{\mathsf{sWI}}$ respectively. The last two hybrids are computationally indistinguishable by the computational hiding of the commitment scheme Com. □

Proposition 6.4 (Similarity of non-aborting part)**.** *Let $\mathsf{V}^* = \{\mathsf{V}^*_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ a polynomial-size quantum verifier, and let $\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{na}}}$ be the verifier's output at the end of protocol such that if V^* aborts, the output is* `Fail`*. Then,*

$$\{\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{na}}}\langle \mathsf{P}(w), \mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}_{\mathrm{na}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w} \ ,$$

where $\lambda \in \mathbb{N}$, $x \in \mathcal{L} \cap \{0,1\}^\lambda$, $w \in \mathcal{R}_{\mathcal{L}}(x)$.

Proof. We prove the claim by a hybrid argument, specifically, we consider hybrid distributions, all of which will be computationally indistinguishable.

- Hyb_0 : The output distribution of $\mathsf{Sim}_{\mathrm{na}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)$.
- Hyb_1 : This hybrid process is identical to Hyb_0, with the exception that when the simulator gives any of the WI proofs γ_1, γ_2 in the simulation, it uses the witness w in the proof, that proves the first statement in the OR statement ($x \in \mathcal{L}$) rather than the second statement.
- Hyb_2 : This hybrid process is identical to Hyb_1, with the exception that cmt_1 is a commitment to the witness w rather than to $0^{|w|}$, cmt_2 is a commitment to 0^λ rather than to r_{SFE} and cmt_3 is a commitment to 0^λ rather than to r_{sWI}.

- Hyb_3 : This hybrid process is identical to Hyb_2, only that the prover performs an inefficient check: Given the verifier's first message it checks whether it is explainable, and if not the process outputs `Fail`. If it was explainable, the simulator proceeds to regularly perform quantum homomorphic evaluation as in Hyb_2.

- Hyb_4 : This process is identical to the previous, with the following changes: (1) we don't check that the verifier's first message is explainable but only break the QFHE ciphertext ct_{V} to get u (if ct_{V} is not a valid ciphertext and there is no such u it sets $u = 0^\lambda$). (2) we take the verifier's response out of the homomorphic evaluation - given u, the simulator gives the verifier an SFE encryption $\mathsf{ct}_{\mathsf{Sim}} = \mathsf{SFE.Enc}(u)$, gets the response $\hat{\mathsf{ct}}_{\mathsf{Sim}}$, and continues regularly.

- Hyb_5 : This hybrid process is identical to Hyb_4, with the exception that instead of breaking the QFHE ciphertext ct_{V} of the verifier (and then sending an SFE encryption of that extracted value), the simulator always just sends an SFE encryption of 0^λ. Observe that in this process the simulator's actions are exactly the same as the prover's in the original protocol, so this process is exactly $\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{na}}}\langle \mathsf{P}(w), \mathsf{V}^*(\rho)\rangle(x)$.

We now prove that each pair of consecutive distributions are computationally indistinguishable, and our proof is finished.

- $\mathsf{Hyb}_0 \approx_c \mathsf{Hyb}_1$: This indistinguishability follows from the witness-indistinguishability property of each of the three WI proofs that the simulator gives in the simulation.

- $\mathsf{Hyb}_1 \approx_c \mathsf{Hyb}_2$: This indistinguishability follows from the hiding of the commitments $\mathsf{cmt}_1, \mathsf{cmt}_2, \mathsf{cmt}_3$ that the simulator gives in step 1a of the simulation.

- $\mathsf{Hyb}_2 \approx_s \mathsf{Hyb}_3$: This indistinguishability follows from the soundness of the verifier's WI. Note that because in both processes the simulator starts with sending a commitment to the witness w, the only way for the verifier's WI statement to be correct is that the verifier's transcript is explainable. Now, the only difference between the processes is the executions where the verifier's first message is not explainable, but still, when it proves that it is explainable in the WI later, the proof is accepted by the simulator. By the quantum computational soundness of the WI and by the fact that V^* is a quantum polynomial-time algorithm, the probability that this happens is negligible, and so is the statistical distance between the two outputs of the hybrids.

- $\mathsf{Hyb}_3 \approx_s \mathsf{Hyb}_4$: This indistinguishability follows from the correctness of all three - the CC obfuscator, the QFHE and the SFE, as well as the soundness of the verifier's WI. We next elaborate how each of these is used.

  1. **CC obfuscation correctness:** We first think of a process which is identical to Hyb_3 with the change that after the simulator performs the quantum homomorphic evaluation of the circuit $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$ (recall the output of this circuit is $(\mathsf{ct}_{\mathsf{Sim}}, \gamma_1, \hat{\mathsf{ct}}_{\mathsf{Sim}}, \tilde{v}, \rho^{(2)})$), instead of trying to decrypt through $\widetilde{\mathbf{CC}}$, the simulator decrypts using the secret key sk (which it obtains inefficiently from the first verifier message). The simulator then checks whether $\tilde{v} = v$, and if not, it outputs `Fail`. By the perfect correctness of the CC obfuscation this process is exactly the same as the previous.
  2. **QFHE correctness:** In the previous process we first make sure that there is some value u inside the verifier QFHE ciphertext ct_{V} by checking explainability, and then homomorphically evaluate the circuit $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$. Then, we decrypt the result, check that $\tilde{v} = v$ and halt if not. In this process everything is the same only that after we break the QFHE ciphertext to get u, we execute $C_{\mathsf{SFE.Enc},\mathsf{V}^*,\mathsf{SFE.Dec}}$ out in the open. This process has negligible statistical distance to the previous by the statistical QFHE correctness that holds for every valid ciphertext[1].

[1] The correctness of the QFHE guarantees that for every valid QFHE ciphertext, even one where the encrypted value and randomness were maliciously chosen, the QFHE evaluation has statistical correctness.

3. **SFE correctness and verifier's WI soundness:** In this process we don't check the full explainability of the first verifier message, but still break the QFHE ciphertext ct_{V} to get u ($u = 0^\lambda$ if the ciphertext is invalid). Also, we don't check that $\tilde{v} = v$ after the verifier SFE evaluation. Note that when the verifier is explainable until the end of its SFE evaluation the processes are identical due to the perfect correctness of the SFE. So, the only difference between the outputs of the processes is contained in the cases where the verifier is not explainable, which means that by the soundness of the WI of the verifier the statistical distance is negligible.

- $\mathsf{Hyb}_4 \approx_c \mathsf{Hyb}_5$: This indistinguishability follows from the input privacy property of the SFE encryption. More precisely, let us assume toward contradiction that the distributions are distinguishable and we fix the transcript until the end of step 1b of the simulation by an averaging argument. If the transcript is not explainable then in both processes the outputs are `Fail` with overwhelming probability by the soundness of the verifier's WI, and are statistically indistinguishable. If the transcript is explainable then there is some u that is sent inside the QFHE encryption ct_{V}. This means that a distinguisher that distinguishes between Hyb_4 and Hyb_5 in this case can distinguish between SFE encryptions of u and 0^λ, in contradiction the the SFE input privacy.

□

The rest of the proof follows directly from the results in [BS20]. We give the proof for completeness. In [BS20] it is shown that whenever one can show two simulators $\mathsf{Sim}_{\mathrm{a}}, \mathsf{Sim}_{\mathrm{na}}$ such that for $b \in \{\mathrm{a}, \mathrm{na}\}$, Sim_b successfully simulates V^*_b (that is, the outputs are computationally indistinguishable), then the simulators can be combined into a single simulator Sim. We use this lemma to complete the proof of our simulation's validity.

Lemma 6.5 (Simulator Combiner Lemma, follows from [BS20])**.** *Let (P, V) an efficient classical protocol (argument or proof system) for proving a language $\mathcal{L}$. Consider a probabilistic event over the execution of the protocol called an abort, where it can be efficiently and publicly decided, given the protocol transcript, whether an abort occurred or not.*

Assume there exist two quantum polynomial-time algorithms $\mathsf{Sim}_{\mathrm{a}}$, $\mathsf{Sim}_{\mathrm{na}}$ such that for every quantum polynomial-time verifier $\mathsf{V}^* = \{\mathsf{V}^*_\lambda, \rho_\lambda\}_{\lambda\in\mathbb{N}}$ the following holds:

- *Let $\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{a}}}$ be the verifier's output at the end of protocol such that if V^* does not abort, the output is* `Fail`*. Then the simulator $\mathsf{Sim}_{\mathrm{a}}$ satisfies,*

$$\{\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{a}}}\langle \mathsf{P}(w), \mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}_{\mathrm{a}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w} \ ,$$

  where $\lambda \in \mathbb{N}$, $x \in \mathcal{L} \cap \{0,1\}^\lambda$, $w \in \mathcal{R}_{\mathcal{L}}(x)$.

- *Let $\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{na}}}$ be the verifier's output at the end of protocol such that if V^* aborts, the output is* `Fail`*. Then the simulator $\mathsf{Sim}_{\mathrm{na}}$ satisfies,*

$$\{\mathsf{OUT}_{\mathsf{V}^*_{\mathrm{na}}}\langle \mathsf{P}(w), \mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}_{\mathrm{na}}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w} \ ,$$

  where $\lambda \in \mathbb{N}$, $x \in \mathcal{L} \cap \{0,1\}^\lambda$, $w \in \mathcal{R}_{\mathcal{L}}(x)$.

Then, there exists a simulator Sim such that every verifier V^*,

$$\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle \mathsf{P}(w), \mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}(x, \mathsf{V}^*_\lambda, \rho_\lambda)\}_{\lambda,x,w} \ .$$

The lemma is proved by citing the correct claims from [BS20].

Proof. For a quantum auxiliary input ρ, instance in the language $x \in \{0,1\}^\lambda \cap \mathcal{L}$ and witness $w \in \mathcal{R}_{\mathcal{L}}(x)$, define the following probabilities.

- $a(x, \rho)$: The probability that in the simulation $\mathsf{Sim}_{\mathrm{a}}(x, \mathsf{V}^*, \rho)$, the verifier V^* aborted.

- $b(x,\rho)$: The probability that in the simulation $\mathsf{Sim}_{\mathrm{na}}(x,\mathsf{V}^*,\rho)$, the verifier V^* aborted.
- $c(x,\rho,w)$: The probability that the interaction $\langle \mathsf{P}(w),\mathsf{V}^*(\rho)\rangle(x)$ was aborting.

Observe that because $\mathsf{Sim}_{\mathrm{a}}$ successfully simulates aborting interactions, in particular the distance between $a(x,\rho)$ and $c(x,\rho,w)$ is negligible. Also, because $\mathsf{Sim}_{\mathrm{na}}$ successfully simulates non-aborting interactions, the distance between $b(x,\rho)$ and $c(x,\rho,w)$ is negligible. By triangle inequality it follows that also $a(x,\rho)$ and $b(x,\rho)$ are negligibly close, and the simulators $\mathsf{Sim}_{\mathrm{a}}$ and $\mathsf{Sim}_{\mathrm{na}}$ satisfy the statement of Corollary 3.1 from [BS20].

Define $\mathsf{Sim}(x,\mathsf{V}^*,\rho)$ exactly the same way it is defined in [BS20], Subection 3.2. By Proposition 3.5 from [BS20] it follows that,

$$\{\mathsf{OUT}_{\mathsf{V}^*_\lambda}\langle \mathsf{P}(w),\mathsf{V}^*_\lambda(\rho_\lambda)\rangle(x)\}_{\lambda,x,w} \approx_c \{\mathsf{Sim}(x,\mathsf{V}^*_\lambda,\rho_\lambda)\}_{\lambda,x,w} \quad .$$

□

7 Quantum Resettable Soundness and Unobfuscatable Functions

In this section we prove a claim establishing a connection between quantum resettable soundness and quantum unobfuscatable functions (as defined in 7.4). This connection shows that constructing such protocol as below, should be as hard as implying the aforementioned impossibility of quantum virtual black-box obfuscation schemes (as defined in 7.3). More formally we show,

Theorem 7.1. *If there exists a post-quantum resettably sound, zero-knowledge classical argument $\langle \mathsf{P},\mathsf{V}\rangle$ for* **NP** *and quantum one-way functions, then there exists an unobfuscatable function family as defined in 7.4*

7.1 Definitions

Definition 7.2 (Quantum Obfuscation Scheme)**.** *A quantum virtual black-box obfuscator for the classical circuit class* $\mathbf{C}$ *is a quantum algorithm $\mathcal{O}$ and a* QPT $\mathcal{J}$ *such that,*

1. **Polynomial Expansion (Compactness)***: For every circuit $C\in\mathbf{C}$, $\mathcal{O}(C)$ is an m-qubit quantum state with $m=\mathrm{poly}(|C|)$.*
2. **Functional Equivalence***: For every circuit $C\in\mathbf{C}$ and every input x,*

$$\mathrm{TD}\left(\mathcal{J}\left(\mathcal{O}(C)\otimes|x\rangle\langle x|\right),|C(x)\rangle\langle C(x)|\right)\leq \mathrm{negl}(|C|) \quad .$$

Definition 7.3 (Quantum Virtual Black-Box Obfuscation Scheme ([ABDS20]))**.** *A quantum obfuscation scheme is virtual black-box if for every non-uniform* QPT *adversary $\mathsf{A}=\{\mathsf{A}_\lambda,\rho_\lambda\}_\lambda$, there exists a* QPT *circuit family $\mathsf{Sim}=\{\mathsf{Sim}_\lambda\}$ (with superposition access to its oracle) such that for all circuits $C\in\mathbf{C}$, $|C|=\lambda$*

$$\left|\Pr\left[\mathsf{A}_\lambda\left(\mathcal{O}(C),\rho_\lambda\right)=1\right]-\Pr\left[\mathsf{Sim}^C_\lambda\left(1^{|C|},\rho_\lambda\right)=1\right]\right|\leq \mathrm{negl}(\lambda) \quad .$$

Definition 7.4 (Quantum Virtual Black-Box Unobfuscateable Circuit Family)**.** *A circuit family $\mathbf{C}=\{C_k\}_k$ parameterized by some parameter k is unobfuscatable for the a relation $\mathcal{R}_{\mathbf{C}}$ if,*

- **Quantum Black-Box Unlearnable***: For any* QPT *algorithm A and quantum advice ρ it holds that,*

$$\Pr_k\left[(k,z)\in\mathcal{R}_{\mathbf{C}}\,\middle|\, z\leftarrow\mathsf{A}^{C_k}(\rho)\right]=\mathrm{negl}(\lambda) \quad .$$

- **Quantum Non Black-Box Learnable** *There exist a* QPT *extractor Ext such that for any obfuscation scheme $(\mathcal{O},\mathcal{J})$, holding compactness and functional equivalence,*

$$\Pr_k\left[(k,z)\in\mathcal{R}_{\mathbf{C}}\,\middle|\, \begin{array}{c}\rho\leftarrow\mathcal{O}(C_k)\\ z\leftarrow\mathsf{Ext}\left((\mathcal{J},\rho),1^{|C|}\right)\end{array}\right]\geq 1-\mathrm{negl}(\lambda) \quad .$$

As noted in [ABDS20], and as their name suggests, unobfuscatable families cannot be obfuscated according to definition 7.3 (in fact this is true, even for inefficient obfuscators, as long as they satisfy polynomial expansion).

7.2 Useful Quantum Algorithms Lemmas

We shall use the following lemmas in proving our construction. First, we state lemma proved in [ABDS20]. Informally, the lemma states that any quantum circuit with almost classical output, can be transformed to an input recovering circuit with the same functionality; namely, one that in addition to computing the (almost classical) output also recovers the initial quantum input.

Lemma 7.5 (Input Recovering Lemma ([ABDS20]))**.** *Let C be a quantum circuit. There exists an* **input-recovering circuit** *C_{rec} such that for any input ρ and classical string x*

$$\mathrm{TD}\left(C_{rec}\left(\rho\right), \rho \otimes |x\rangle\langle x|\right) \leq 2\sqrt{\mathrm{TD}\left(C\left(\rho\right), |x\rangle\langle x|\right)} \ .$$

Secondly, we state the one-way to hiding lemma from [AHU19]. Informally, the lemma asserts that any quantum distinguisher between two classical oracles can be turned into one that finds inputs on which the two oracle differ (with related probability).

Lemma 7.6 (One-Way to Hiding Lemma ([AHU19]))**.** *There exists an oracle-aided QPT algorithm* B *such that for any d-query oracle-aided quantum circuit* A *and (classical) functions $H, G : \mathcal{X} \to \mathcal{Y}$,*

$$\left|\Pr\left[\mathsf{A}^{H}=1\right]-\Pr\left[\mathsf{A}^{G}=1\right]\right| \leq 2d\sqrt{\Pr\left[\exists x \in T : H(x) \neq G(x) \,|\, T \leftarrow \mathsf{B}^{H}\left(\mathsf{A}\right)\right]} \ .$$

7.3 Construction

Let $\mathsf{PRG} : \{0,1\}^n \to \{0,1\}^{2n}$ be a quantum secure length doubling pseudo-random generator. Define $\mathcal{L}$ to be the image set of PRG. Let $\langle \mathsf{P}, \mathsf{V} \rangle$ be a d-round argument for **NP** and in particular for $\mathcal{L}$, where V uses m bits of randomness. Define the following circuit family $\mathbf{C} = \{C_k\}_{k \in \{0,1\}^{2n+m}}$ that implements the honest verifier V next message functionality on some $y \in \mathcal{L}$ statement (to be later defined). More formally, interpret k as the triplet (x, r, s) where x is an input to the PRG, r is the randomness for V and s is some secret string. The functionality of C_k is as follows,

1. Given a special input, $\mathtt{ST}$, C_k outputs $y = \mathsf{PRG}\left(x\right)$.
2. Given a partial prover side transcript $\mathsf{ts}_{\mathsf{P}}^{i} = (p_1, \ldots, p_i)$ for $i < d$, containing prover messages, C_k executes V on the statement $y \in \mathcal{L}$ with randomness r and outputs v_i, the i^{th} verifier message.
3. Given a full prover side transcript $\mathsf{ts}_{\mathsf{P}}^{d} = (p_1, \ldots, p_d)$, C_k it computes the full transcript of the interaction ts and the predicate $\mathsf{V}\left(y, ts_P^d; r\right)$. If the predicate accepts it outputs s, otherwise it outputs $\perp$.

The relation we define for this function family $\mathcal{R}_{\mathbf{C}} = \left\{(k, s) \,\middle|\, k \in \{0,1\}^{2n+m}, k = (x, r, s)\right\}$. We show that the above construction is indeed an unobfuscatable function family.

Black-Box Unlearnability*:* Assume some QPT A that is able to learn the relation $\mathcal{R}_{\mathbf{C}}$ with some probability ε. Fix some $C_k = C_{x,r,s}$, where k was sampled uniformly at random. We now rely one the one-way to hiding lemma from [AHU19] (Lemma 7.6) to transform A into an adversary B that outputs an accepting transcript with related success probability. We denote by C'_k the circuit that implement exactly the same functionality as C_k with the only difference being that for any full prover side transcript $\mathsf{ts}_{\mathsf{P}}^{d}$ it is queried upon, it outputs $\perp$. Then, we note that by the one-way to hiding lemma there exists a QPT B such that,

$$\Pr\left[C_{x,r,s}\left(\mathsf{ts}_{\mathsf{P}}^{d}\right)=s \,\middle|\, \mathsf{ts}_{\mathsf{P}}^{d}\leftarrow \mathsf{B}^{C_{x,r,s}}\right]\geq$$
$$\frac{1}{\mathrm{poly}\left(|C_k|\right)}\left|\Pr\left[s\neq\perp \,\middle|\, s\leftarrow \mathsf{A}^{C_k}\right]-\Pr\left[s\neq\perp \,\middle|\, s\leftarrow \mathsf{A}^{C'_k}\right]\right|=$$
$$\frac{\varepsilon}{\mathrm{poly}\left(|C_k|\right)} \quad .$$

We then note that B can be easily transformed into a quantum resetting prover for the statement $y\in\mathcal{L}$ against some uniformly sampled $\mathsf{V}\left(y,\cdot;r\right)$ using the resetting access (we can simulate black-box access to $C_{x,r,s}$ for a uniformly sampled s, and generate a corresponding ts). More so, ts is accepting with probability of $\Omega\left(\frac{\varepsilon}{\mathrm{poly}(|C_k|)}\right)$. Finally, we note that by the pseudo-randomness of the PRG, B succeeds with roughly the same probability for a uniform $y\in\{0,1\}^{2n}$ (otherwise B could be turned to adversary for the PRG). Hence, since with overwhelming probability for a uniformly random y, the statement $y\in\mathcal{L}$ is false, B only succeeds with negligible probability, implying $\varepsilon=\mathrm{negl}\left(|C_k|\right)$. □

Non Black-Box Learnability: We aim to construct an extractor Ext such that given any quantum obfuscation of C_k, Ext extracts with probability negligibly close to 1. The extractor is presented in 6

Algorithm 6: $\mathsf{Ext}\left(\rho\right)$ - An extractor for the relation $\mathcal{R}_{\mathbf{C}}$

1 Apply $\mathcal{J}_{rec}\left(\rho\otimes|\mathtt{ST}\rangle\langle\mathtt{ST}|\right)$, where $\mathcal{J}_{rec}$ is the input-recovering version of $\mathcal{J}$ (Lemma 7.5) the evaluation algorithm. Measure the result as y and denote the residual state as ρ'_0.
2 Run $\mathsf{Sim}\left(y,\mathsf{V}^*,\rho'_0\right)$ for the verifier V^* as described in 7, obtaining ts,b,ρ'.
3 Run $\mathcal{J}\left(\rho'\otimes|\mathsf{ts}\rangle\langle\mathsf{ts}|\right)$ measure the result as s'.
4 Output s'.

Algorithm 7: $\mathsf{V}^*\left(\rho'_0\right)$ - A quantum verifier for the protocol $\langle\mathsf{P},\mathsf{V}\rangle$

1 Set ts_{P} to be an empty list.
2 For $i\in[1,\ldots,d-1]$:
3 Accept prover message p_i and append it to ts_{P}.
4 Run $\mathcal{J}_{rec}\left(\rho'_{i-1}\otimes|\mathsf{ts}_{\mathsf{P}}\rangle\langle\mathsf{ts}_{\mathsf{P}}|\right)$ and measure the output v_i. Denote the residual state ρ'_i.
5 Output v_i as the i^{th} message.
6 Upon receiving the final message p_d append it to ts_{P} and apply $\mathcal{J}_{rec}\left(\rho'_d\otimes|\mathsf{ts}_{\mathsf{P}}\rangle\langle\mathsf{ts}_{\mathsf{P}}|\right)$.
7 Measure the acceptance bit of the result b (using the projections $\Pi_\perp=\sum_{s:s\neq\perp}|s\rangle\langle s|$). Denote the residual state and σ
8 Output b,σ.

Claim 7.7. *For any quantum obfuscation scheme $(\mathcal{O},\mathcal{J})$, it holds that,*

$$\Pr_{k}\left[(k,s)\in\mathcal{R}_{\mathbf{C}}\,\middle|\,\begin{array}{c}\rho\leftarrow\mathcal{O}\left(C_k\right)\\ s\leftarrow\mathsf{Ext}\left(\rho,\mathcal{J}\right)\end{array}\right]\geq 1-\mathrm{negl}\left(|C_k|\right) \quad .$$

Proof. Assume $k=(x,r,s)$. First, note that due to functional equivalence of the obfuscation scheme,

$$\mathrm{TD}\left(\mathcal{J}\left(\rho\otimes|\mathtt{ST}\rangle\langle\mathtt{ST}|\right),|\mathsf{PRG}\left(x\right)\rangle\langle\mathsf{PRG}\left(x\right)|\right)\leq\mathrm{negl}\left(|C_k|\right) \quad .$$

Hence, with overwhelming probability $y=\mathsf{PRG}\left(x\right)$. More so, by the input recovering lemma $\mathrm{TD}\left(\rho,\rho'_0\right)\leq\mathrm{negl}\left(|C_k|\right)$

We then argue the following claim,

Claim 7.8. *for any* $\tau \in \left\{\rho'_0, \cdots, \rho'_{d-1}, \sigma\right\}$, $\mathrm{TD}(\tau, \rho) \leq \mathrm{negl}(|C_k|)$

Proof. Denote $\sigma = \rho'_d$. Then, the claim follows by induction, where the base case was proved above for ρ'_0 and the induction step is repeating the above reasoning to show that $\mathrm{TD}\left(\rho'_i, \rho'_{i+1}\right)$. The induction step then follows from the triangle inequality. We emphasize here that we rely on the fact the protocol is classical, to be able to use the input recovering lemma. □

Thus, using functional equivalence for ρ and Claim 7.8 it holds,

$$\left\{(\mathsf{ts}, b) \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \end{array}\right\}_{x,s\in\{0,1\}^n, r\in\{0,1\}^m} \approx_s \tag{5}$$
$$\left\{(\mathsf{ts}, b) \,|\, (\mathsf{ts}, b) \leftarrow \langle \mathsf{P}, \mathsf{V}(\cdot; r)\rangle(\mathsf{PRG}(x))\right\}_{x\in\{0,1\}^n, r\in\{0,1\}^m} .$$

Also, since for any x, $\mathsf{PRG}(x) \in L$ is a true statement, then from zero-knowledge it holds,

$$\left\{(\mathsf{ts}, b, \sigma) \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \end{array}\right\}_{x,s\in\{0,1\}^n, r\in\{0,1\}^m} \approx_c \tag{6}$$
$$\left\{(\mathsf{ts}, b, \sigma) \,|\, (\mathsf{ts}, b, \sigma) \leftarrow \mathsf{Sim}(\mathsf{PRG}(x), \mathsf{V}^*, \rho'_0)\right\}_{x\in\{0,1\}^n, r\in\{0,1\}^m} ,$$

where Sim is the zero-knowledge simulator of $\langle \mathsf{P}, \mathsf{V}\rangle$.

We also note that after the interaction of V^* with the prove P the final state σ holds $\mathrm{TD}(\sigma, \rho) \leq \mathrm{negl}(|C_k|)$ (following Claim 7.8). Hence due to functional equivalence for ρ it holds that,

$$\Pr_{x,r,s}\left[s' = C_{x,r,s}(\mathsf{ts}) \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \\ s' \leftarrow \mathcal{M}(\mathcal{J}(\sigma \otimes |\mathsf{ts}\rangle\langle\mathsf{ts}|)) \end{array}\right] \geq 1 - \mathrm{negl}(|C_k|) ,$$

and that

$$\Pr_{x,r,s}\left[s' = C_{x,r,s}(\mathsf{ts}') \equiv s \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \\ (\mathsf{ts}', b') \leftarrow \langle \mathsf{P}, \mathsf{V}(x, \cdot; r)\rangle(\mathsf{PRG}(x)) \\ s' \leftarrow \mathcal{M}(\mathcal{J}(\sigma \otimes |\mathsf{ts}'\rangle\langle\mathsf{ts}'|)) \end{array}\right] \geq 1 - \mathrm{negl}(|C_k|) .$$

Hence, due to 5 it holds that,

$$\Pr_{x,r,s}\left[s' = s \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \\ s' \leftarrow \mathcal{M}(\mathcal{J}(\sigma \otimes |\mathsf{ts}\rangle\langle\mathsf{ts}|)) \end{array}\right] \geq 1 - \mathrm{negl}(|C_k|) .$$

Then, using 6 it holds that,

$$\Pr_k\left[(k, s) \in \mathcal{R}_{\mathbf{C}} \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_k) \\ s \leftarrow \mathsf{Ext}(\rho, \mathcal{J}) \end{array}\right] = \Pr_{x,r,s}\left[s' = s \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \mathsf{Sim}(y, \mathsf{V}^*, \rho'_0) \\ s' \leftarrow \mathcal{M}(\mathcal{J}(\sigma \otimes |\mathsf{ts}\rangle\langle\mathsf{ts}|)) \end{array}\right] \geq$$
$$\Pr_{x,r,s}\left[s' = s \;\middle|\; \begin{array}{c} \rho \leftarrow \mathcal{O}(C_{x,r,s}) \\ (y, \rho'_0) \leftarrow \mathcal{M}(\mathcal{J}_{rec}(\rho \otimes |\mathtt{ST}\rangle\langle\mathtt{ST}|)) \\ (\mathsf{ts}, b, \sigma) \leftarrow \langle \mathsf{P}, \mathsf{V}^*(\rho'_0)\rangle(y) \\ s' \leftarrow \mathcal{M}(\mathcal{J}(\sigma \otimes |\mathsf{ts}\rangle\langle\mathsf{ts}|)) \end{array}\right] - \mathrm{negl}(|C_k|) \geq 1 - \mathrm{negl}(|C_k|) .$$

□

References

[ABB+17] Erdem Alkim, Nina Bindel, Johannes Buchmann, Özgür Dagdelen, Edward Eaton, Gus Gutoski, Juliane Krämer, and Filip Pawlega. Revisiting tesla in the quantum random oracle model. In *International Workshop on Post-Quantum Cryptography*, pages 143–162. Springer, 2017.

[ABDS20] Gorjan Alagic, Zvika Brakerski, Yfke Dulek, and Christian Schaffner. Impossibility of quantum virtual black-box obfuscation of classical circuits. *CoRR*, abs/2005.06432, 2020.

[ABG+20] Amit Agarwal, James Bartusek, Vipul Goyal, Dakshita Khurana, and Giulio Malavolta. Post-quantum multi-party computation. *IACR Cryptol. ePrint Arch.*, 2020:1395, 2020.

[AF16] Gorjan Alagic and Bill Fefferman. On quantum obfuscation. *CoRR*, abs/1602.01771, 2016.

[AHU19] Andris Ambainis, Mike Hamburg, and Dominique Unruh. Quantum security proofs using semi-classical oracles. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology – CRYPTO 2019*, pages 269–295, Cham, 2019. Springer International Publishing.

[AP20] Prabhanjan Ananth and Rolando L. La Placa. Secure software leasing. *CoRR*, abs/2005.05289, 2020.

[Bar01] Boaz Barak. How to go beyond the black-box simulation barrier. In *Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science*, FOCS '01, page 106, USA, 2001. IEEE Computer Society.

[BCC88] Gilles Brassard, David Chaum, and Claude Crépeau. Minimum disclosure proofs of knowledge. *J. Comput. Syst. Sci.*, 37(2):156–189, 1988.

[BD18] Zvika Brakerski and Nico Döttling. Two-message statistically sender-private ot from lwe. In *Theory of Cryptography Conference*, pages 370–390. Springer, 2018.

[BDF+10] Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. *IACR Cryptol. ePrint Arch.*, 2010:428, 2010.

[BFJ+20] Saikrishna Badrinarayanan, Rex Fernando, Aayush Jain, Dakshita Khurana, and Amit Sahai. Statistical ZAP arguments. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part III*, volume 12107 of *Lecture Notes in Computer Science*, pages 642–667. Springer, 2020.

[BGGL01] Boaz Barak, Oded Goldreich, Shafi Goldwasser, and Yehuda Lindell. Resettably-sound zero-knowledge and its applications. In *42nd Annual Symposium on Foundations of Computer Science, FOCS 2001, 14-17 October 2001, Las Vegas, Nevada, USA*, pages 116–125. IEEE Computer Society, 2001.

[BGI+12] Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6:1–6:48, 2012.

[BKP18] Nir Bitansky, Yael Tauman Kalai, and Omer Paneth. Multi-collision resistance: a paradigm for keyless hash functions. In Ilias Diakonikolas, David Kempe, and Monika Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 671–684. ACM, 2018.

[BKP19] Nir Bitansky, Dakshita Khurana, and Omer Paneth. Weak zero-knowledge beyond the black-box barrier. In Moses Charikar and Edith Cohen, editors, *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019*, pages 1091–1102. ACM, 2019.

[Blu86] Manuel Blum. How to prove a theorem so no one else can claim it. In *Proceedings of the International Congress of Mathematicians*, volume 1, page 2. Citeseer, 1986.

[BLV06] Boaz Barak, Yehuda Lindell, and Salil P. Vadhan. Lower bounds for non-black-box zero knowledge. *J. Comput. Syst. Sci.*, 72(2):321–391, 2006.

[BP12] Nir Bitansky and Omer Paneth. From the impossibility of obfuscation to a new non-black-box simulation technique. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 223–232. IEEE Computer Society, 2012.

[BP13] Nir Bitansky and Omer Paneth. On the impossibility of approximate obfuscation and applications to resettable cryptography. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 241–250. ACM, 2013.

[BP15] Nir Bitansky and Omer Paneth. On non-black-box simulation and the impossibility of approximate obfuscation. *SIAM J. Comput.*, 44(5):1325–1383, 2015.

[Bra18a] Zvika Brakerski. Quantum fhe (almost) as secure as classical. In *Annual International Cryptology Conference*, pages 67–95. Springer, 2018.

[Bra18b] Zvika Brakerski. Quantum fhe (almost) as secure as classical. In *Annual International Cryptology Conference*, pages 67–95. Springer, 2018.

[BS20] Nir Bitansky and Omri Shmueli. Post-quantum zero knowledge in constant rounds. In Konstantin Makarychev, Yury Makarychev, Madhur Tulsiani, Gautam Kamath, and Julia Chuzhoy, editors, *Proccedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing, STOC 2020, Chicago, IL, USA, June 22-26, 2020*, pages 269–279. ACM, 2020.

[CCY20] Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. A black-box approach to post-quantum zero-knowledge in constant rounds. *IACR Cryptol. ePrint Arch.*, 2020:1384, 2020.

[CGGM00] Ran Canetti, Oded Goldreich, Shafi Goldwasser, and Silvio Micali. Resettable zero-knowledge (extended abstract). In F. Frances Yao and Eugene M. Luks, editors, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 235–244. ACM, 2000.

[CMS19] Alessandro Chiesa, Peter Manohar, and Nicholas Spooner. Succinct arguments in the quantum random oracle model. In *Theory of Cryptography Conference*, pages 1–29. Springer, 2019.

[COP+14] Kai-Min Chung, Rafail Ostrovsky, Rafael Pass, Muthuramakrishnan Venkitasubramaniam, and Ivan Visconti. 4-round resettably-sound zero knowledge. In *Theory of Cryptography Conference*, pages 192–216. Springer, 2014.

[COPV13] Kai-Min Chung, Rafail Ostrovsky, Rafael Pass, and Ivan Visconti. Simultaneous resettability from one-way functions. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 60–69. IEEE, 2013.

[COSV12] Chongwon Cho, Rafail Ostrovsky, Alessandra Scafuro, and Ivan Visconti. Simultaneously resettable arguments of knowledge. In *Theory of Cryptography Conference*, pages 530–547. Springer, 2012.

[COV17] Wutichai Chongchitmate, Rafail Ostrovsky, and Ivan Visconti. Resettably-sound resettable zero knowledge in constant rounds. In Yael Kalai and Leonid Reyzin, editors, *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part II*, volume 10678 of *Lecture Notes in Computer Science*, pages 111–138. Springer, 2017.

[CPS13] Kai-Min Chung, Rafael Pass, and Karn Seth. Non-black-box simulation from one-way functions and applications to resettable security. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 231–240. ACM, 2013.

[CPS16] Kai-Min Chung, Rafael Pass, and Karn Seth. Non-black-box simulation from one-way functions and applications to resettable security. *SIAM Journal on Computing*, 45(2):415–458, 2016.

[DFM20] Jelle Don, Serge Fehr, and Christian Majenz. The measure-and-reprogram technique 2.0: Multi-round fiat-shamir and more. In Daniele Micciancio and Thomas Ristenpart, editors, *Advances in Cryptology - CRYPTO 2020 - 40th Annual International Cryptology Conference, CRYPTO 2020, Santa Barbara, CA, USA, August 17-21, 2020, Proceedings, Part III*, volume 12172 of *Lecture Notes in Computer Science*, pages 602–631. Springer, 2020.

[DFMS19] Jelle Don, Serge Fehr, Christian Majenz, and Christian Schaffner. Security of the fiat-shamir transformation in the quantum random-oracle model. In Alexandra Boldyreva and Daniele Micciancio, editors, *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part II*, volume 11693 of *Lecture Notes in Computer Science*, pages 356–383. Springer, 2019.

[DFS04] Ivan Damgård, Serge Fehr, and Louis Salvail. Zero-knowledge proofs and string commitments withstanding quantum attacks. In *Annual International Cryptology Conference*, pages 254–272. Springer, 2004.

[DGS09] Yi Deng, Vipul Goyal, and Amit Sahai. Resolving the simultaneous resettability conjecture and a new non-black-box simulation strategy. In *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 251–260. IEEE, 2009.

[DNRS03] Cynthia Dwork, Moni Naor, Omer Reingold, and Larry J. Stockmeyer. Magic functions. *J. ACM*, 50(6):852–921, 2003.

[ES15] Edward Eaton and Fang Song. Making existential-unforgeable signatures strongly unforgeable in the quantum random-oracle model. In Salman Beigi and Robert König, editors, *10th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2015, May 20-22, 2015, Brussels, Belgium*, volume 44 of *LIPIcs*, pages 147–162. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015.

[FGJ18] Nils Fleischhacker, Vipul Goyal, and Abhishek Jain. On the existence of three round zero-knowledge proofs. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part III*, volume 10822 of *Lecture Notes in Computer Science*, pages 3–33. Springer, 2018.

[FLS99] Uriel Feige, Dror Lapidot, and Adi Shamir. Multiple noninteractive zero knowledge proofs under general assumptions. *SIAM J. Comput.*, 29(1):1–28, 1999.

[FP96] Christopher A. Fuchs and Asher Peres. Quantum-state disturbance versus information gain: Uncertainty relations for quantum information. *Phys. Rev. A*, 53:2038–2045, Apr 1996.

[FS86] Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In Andrew M. Odlyzko, editor, *Advances in Cryptology - CRYPTO '86, Santa Barbara, California, USA, 1986, Proceedings*, volume 263 of *Lecture Notes in Computer Science*, pages 186–194. Springer, 1986.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.

[GHKW17] Rishab Goyal, Susan Hohenberger, Venkata Koppula, and Brent Waters. A generic approach to constructing and proving verifiable random functions. In Yael Kalai and Leonid Reyzin, editors, *Theory of Cryptography - 15th International Conference, TCC 2017, Baltimore, MD, USA, November 12-15, 2017, Proceedings, Part II*, volume 10678 of *Lecture Notes in Computer Science*, pages 537–566. Springer, 2017.

[GJJM20] Vipul Goyal, Abhishek Jain, Zhengzhong Jin, and Giulio Malavolta. Statistical zaps and new oblivious transfer protocols. In Anne Canteaut and Yuval Ishai, editors, *Advances in Cryptology - EUROCRYPT 2020 - 39th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Zagreb, Croatia, May 10-14, 2020, Proceedings, Part III*, volume 12107 of *Lecture Notes in Computer Science*, pages 668–699. Springer, 2020.

[GK96a] Oded Goldreich and Ariel Kahan. How to construct constant-round zero-knowledge proof systems for NP. *J. Cryptol.*, 9(3):167–190, 1996.

[GK96b] Oded Goldreich and Hugo Krawczyk. On the composition of zero-knowledge proof systems. *SIAM J. Comput.*, 25(1):169–192, February 1996.

[GKVW20] Rishab Goyal, Venkata Koppula, Satyanarayana Vusirikala, and Brent Waters. On perfect correctness in (lockable) obfuscation. In Rafael Pass and Krzysztof Pietrzak, editors, *Theory of Cryptography - 18th International Conference, TCC 2020, Durham, NC, USA, November 16-19, 2020, Proceedings, Part I*, volume 12550 of *Lecture Notes in Computer Science*, pages 229–259. Springer, 2020.

[GKW17] Rishab Goyal, Venkata Koppula, and Brent Waters. Lockable obfuscation. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 612–621. IEEE, 2017.

[GM11] Vipul Goyal and Hemanta K. Maji. Stateless cryptographic protocols. In Rafail Ostrovsky, editor, *IEEE 52nd Annual Symposium on Foundations of Computer Science, FOCS 2011, Palm Springs, CA, USA, October 22-25, 2011*, pages 678–687. IEEE Computer Society, 2011.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM J. Comput.*, 18(1):186–208, 1989.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity or all languages in np have zero-knowledge proof systems. *J. ACM*, 38(3):690–728, July 1991.

[Goy13] Vipul Goyal. Non-black-box simulation in the fully concurrent setting. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 221–230. ACM, 2013.

[GS09a] Vipul Goyal and Amit Sahai. Resettably secure computation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 54–71. Springer, 2009.

[GS09b] Vipul Goyal and Amit Sahai. Resettably secure computation. In Antoine Joux, editor, *Advances in Cryptology - EUROCRYPT 2009, 28th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cologne, Germany, April 26-30, 2009. Proceedings*, volume 5479 of *Lecture Notes in Computer Science*, pages 54–71. Springer, 2009.

[HI19] Akinori Hosoyamada and Tetsu Iwata. 4-round luby-rackoff construction is a qprp. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 145–174. Springer, 2019.

[JKMR09] Rahul Jain, Alexandra Kolla, Gatis Midrijanis, and Ben W. Reichardt. On parallel composition of zero-knowledge proofs with black-box quantum simulators. *Quantum Inf. Comput.*, 9(5&6):513–532, 2009.

[KLS18] Eike Kiltz, Vadim Lyubashevsky, and Christian Schaffner. A concrete treatment of fiat-shamir signatures in the quantum random-oracle model. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology - EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Tel Aviv, Israel, April 29 - May 3, 2018 Proceedings, Part III*, volume 10822 of *Lecture Notes in Computer Science*, pages 552–586. Springer, 2018.

[KNY20] Fuyuki Kitagawa, Ryo Nishimaki, and Takashi Yamakawa. Secure software leasing from standard assumptions. *IACR Cryptol. ePrint Arch.*, 2020:1314, 2020.

[Kob03] Hirotada Kobayashi. Non-interactive quantum perfect and statistical zero-knowledge. In Toshihide Ibaraki, Naoki Katoh, and Hirotaka Ono, editors, *Algorithms and Computation, 14th International Symposium, ISAAC 2003, Kyoto, Japan, December 15-17, 2003, Proceedings*, volume 2906 of *Lecture Notes in Computer Science*, pages 178–188. Springer, 2003.

[KP01] Joe Kilian and Erez Petrank. Concurrent and resettable zero-knowledge in poly-loalgorithm rounds. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 560–569, 2001.

[KRR17] Yael Tauman Kalai, Guy N. Rothblum, and Ron D. Rothblum. From obfuscation to the security of fiat-shamir for proofs. In Jonathan Katz and Hovav Shacham, editors, *Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part II*, volume 10402 of *Lecture Notes in Computer Science*, pages 224–251. Springer, 2017.

[LS19] Alex Lombardi and Luke Schaeffer. A note on key agreement and non-interactive commitments. *IACR Cryptology ePrint Archive*, 2019:279, 2019.

[LZ19] Qipeng Liu and Mark Zhandry. Revisiting post-quantum fiat-shamir. In *Annual International Cryptology Conference*, pages 326–355. Springer, 2019.

[Mah18a] Urmila Mahadev. Classical homomorphic encryption for quantum circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 332–338. IEEE, 2018.

[Mah18b] Urmila Mahadev. Classical homomorphic encryption for quantum circuits. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 332–338. IEEE, 2018.

[MR01a] Silvio Micali and Leonid Reyzin. Min-round resettable zero-knowledge in the public-key model. In Birgit Pfitzmann, editor, *Advances in Cryptology - EUROCRYPT 2001, International Conference on the Theory and Application of Cryptographic Techniques, Innsbruck, Austria, May 6-10, 2001, Proceeding*, volume 2045 of *Lecture Notes in Computer Science*, pages 373–393. Springer, 2001.

[MR01b] Silvio Micali and Leonid Reyzin. Min-round resettable zero-knowledge in the public-key model. In *International Conference on the Theory and Applications of Cryptographic Techniques*, pages 373–393. Springer, 2001.

[OPCPC14] Rafail Ostrovsky, Anat Paskin-Cherniavsky, and Beni Paskin-Cherniavsky. Maliciously circuit-private fhe. In *Annual Cryptology Conference*, pages 536–553. Springer, 2014.

[OV12] Rafail Ostrovsky and Ivan Visconti. Simultaneous resettability from collision resistance. *Electron. Colloquium Comput. Complex.*, 19:164, 2012.

[PTW11] Rafael Pass, Wei-Lung Dustin Tseng, and Douglas Wikström. On the composition of public-coin zero-knowledge protocols. *SIAM J. Comput.*, 40(6):1529–1553, 2011.

[Reg05] Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '05, page 84–93, New York, NY, USA, 2005. Association for Computing Machinery.

[TU16] Ehsan Ebrahimi Targhi and Dominique Unruh. Post-quantum security of the fujisaki-okamoto and OAEP transforms. In Martin Hirt and Adam D. Smith, editors, *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part II*, volume 9986 of *Lecture Notes in Computer Science*, pages 192–216, 2016.

[Unr14] Dominique Unruh. Quantum position verification in the random oracle model. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II*, volume 8617 of *Lecture Notes in Computer Science*, pages 1–18. Springer, 2014.

[Unr15] Dominique Unruh. Non-interactive zero-knowledge proofs in the quantum random oracle model. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II*, volume 9057 of *Lecture Notes in Computer Science*, pages 755–784. Springer, 2015.

[Unr16a] Dominique Unruh. Collapse-binding quantum commitments without random oracles. In Jung Hee Cheon and Tsuyoshi Takagi, editors, *Advances in Cryptology - ASIACRYPT 2016 - 22nd International Conference on the Theory and Application of Cryptology and Information Security, Hanoi, Vietnam, December 4-8, 2016, Proceedings, Part II*, volume 10032 of *Lecture Notes in Computer Science*, pages 166–195, 2016.

[Unr16b] Dominique Unruh. Computationally binding quantum commitments. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II*, volume 9666 of *Lecture Notes in Computer Science*, pages 497–527. Springer, 2016.

[VDGC97] Jeroen Van De Graaf and C Crepeau. *Towards a formal definition of security for quantum protocols*. Université de Montréal, 1997.

[Wat02] John Watrous. Limits on the power of quantum statistical zero-knowledge. In *43rd Symposium on Foundations of Computer Science (FOCS 2002), 16-19 November 2002, Vancouver, BC, Canada, Proceedings*, page 459. IEEE Computer Society, 2002.

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

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

[WZ17] Daniel Wichs and Giorgos Zirdelis. Obfuscating compute-and-compare programs under lwe. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 600–611. IEEE, 2017.

[Zha12] Mark Zhandry. How to construct quantum random functions. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 679–687. IEEE Computer Society, 2012.

[Zha15] Mark Zhandry. Secure identity-based encryption in the quantum random oracle model. *International Journal of Quantum Information*, 13(04):1550014, 2015.

[Zha18] Mark Zhandry. How to record quantum queries, and applications to quantum indifferentiability. *IACR Cryptol. ePrint Arch.*, 2018:276, 2018.

A Missing Proofs

A.1 Proof of Theorem 4.3

Theorem 4.3. *If a language $\mathcal{L}$ has a post-quantum black-box zero-knowledge, resettably sound protocol, then $\mathcal{L} \in \mathbf{BQP}$.*

Proof. We describe a decider for the language $\mathcal{L}$ which will be based on the simulator Sim of protocol $\langle \mathsf{P}, \mathsf{V} \rangle$. Assume V uses m bits of randomness. The decider is described in 8. We aim to show that indeed $\mathcal{L}$ recognizes

Algorithm 8: $\mathsf{D}(x)$ - A decider for the language $\mathcal{L}$

1 Sample some randomness $r \leftarrow \{0,1\}^m$.
2 Run $\mathsf{Sim}^{\mathsf{V}(x,\cdot;r)}(x)$ and measure an output transcript ts.
3 Compute $\mathsf{V}(x, \mathsf{ts}; r)$ and output it as output.

the language $\mathcal{L}$.

Correctness: Fix some $x \in \mathcal{L}$, then by the zero-knowledge guarantee for the simulator Sim it holds that for any r,

$$\left\{ \mathsf{ts} \,\middle|\, \mathsf{ts} \leftarrow \mathsf{Sim}^{\mathsf{V}(x,\cdot;r)}(x) \right\} \approx_c \{ \mathsf{ts} \,|\, \mathsf{ts} \leftarrow \langle \mathsf{P}, \mathsf{V}(x;r) \rangle (x) \} \quad .$$

Then, with overwhelming probability over r, by the correctness of $\langle \mathsf{P}, \mathsf{V} \rangle$ it holds that,

$$\mathop{\mathbb{E}}_{r} [\mathsf{V}(x, \mathsf{ts}; r) = 1 \,|\, \mathsf{ts} \leftarrow \langle \mathsf{P}, \mathsf{V}(x;r) \rangle (x)] \geq 1 - \mathrm{negl}(\lambda) \quad .$$

Hence,

$$\begin{aligned} \Pr[b = 1 \,|\, b \leftarrow \mathsf{D}(x)] =& \mathop{\mathbb{E}}_{r} \left[\mathsf{V}(x, \mathsf{ts}; r) = 1 \,\middle|\, \mathsf{ts} \leftarrow \mathsf{Sim}^{\mathsf{V}(x,\cdot;r)}(x) \right] \\ \geq& \mathop{\mathbb{E}}_{r} [\mathsf{V}(x, \mathsf{ts}; r) = 1 \,|\, \mathsf{ts} \leftarrow \langle \mathsf{P}, \mathsf{V}(x;r) \rangle (x) - \mathrm{negl}(\lambda)] \\ \geq& 1 - \mathrm{negl}(\lambda) \quad . \end{aligned}$$

Soundness: Let there be some $x \notin \mathcal{L}$, we aim to bound,

$$\Pr[b = 1 \,|\, b \leftarrow \mathsf{D}(x)] \quad .$$

To bound the above probability, assume some malicious quantum resettable prover rP, that given some black-box access to some randomly sampled $\mathsf{V}(x, \cdot; r)$, simply runs Sim using this access, and outputs the same transcript. Then, by resettable soundness it holds that,

$$\begin{aligned} \Pr[b = 1 \,|\, b \leftarrow \mathsf{D}(x)] =& \mathop{\mathbb{E}}_{r} \left[\mathsf{V}(x, \mathsf{ts}; r) = 1 \,\middle|\, \mathsf{ts} \leftarrow \mathsf{Sim}^{\mathsf{V}(x,\cdot;r)}(x) \right] = \Pr_{r} \left[\mathsf{V}(x, \mathsf{ts}; r) = 1 \,\middle|\, \mathsf{ts} \leftarrow \mathsf{rP}^{\mathsf{V}(x,\cdot;r)} \right] \\ \leq& \mathrm{negl}(\lambda) \quad . \end{aligned}$$

This concludes the proof of Theorem 4.3 □

Remark. *While the above proof is described with negligible completeness and correctness errors, the same proof holds for any completeness error c and resettable soundness error s such that $s < 1 - c - \mathrm{poly}(\lambda)$.*

Remark. *Also, while the above assumes implicitly a strict quantum polynomial time simulator Sim, the proof extend to simulators with expected quantum polynomial time, by running the simulator for time T where $\frac{1}{T} \geq 1 - c - s$, for completeness error c and resettable soundness error s (following Markov's inequality).*

A.2 Proof of Corollary 5.11

Corollary 5.11. *Assuming sub-exponentially secure PRFs, any resettably sound protocol can be transformed into a multi-input resettably sound one.*

Proof Sketch (of Corollary 5.11). Assume some resettably sound protocol $\langle \mathsf{P}, \mathsf{V} \rangle$. We modify V to $\tilde{\mathsf{V}}$ such that $\tilde{\mathsf{V}}$ with inner randomness k receives the instance x and first message a and interacts by following $\mathsf{V}(x, \cdot; \mathsf{PRF}_k(x))$ with the first message a. We claim that any multi-input resetting prover against $\left\langle \mathsf{P}, \tilde{\mathsf{V}} \right\rangle$ can be transformed into a resetting prover against V.

Assume some mP multi-input prover against $\tilde{\mathsf{V}}$. We first consider the variant where $\tilde{\mathsf{V}}^R$ uses a random oracle R instead of a PRF to derive the randomness. We wish to claim as before that mP succeeds with the same (up to negligible difference) against $\tilde{\mathsf{V}}^R$. However, if we used polynomially secure PRFs we couldn't argue that. Note that to rely on the pseudo-randomness of the PRF we have to describe a polynomial distinguisher between oracle access to the PRF or a random function. However, such a reduction needs to decide if the instance x outputted by mP is indeed false. However, this cannot necessarily be done in polynomial time.

To circumvent the above problem, we use sub-exponentially secure PRFs instead. Then our reduction can find the witness for x if such one exists, in $O\left(2^{|x|}\right)$ time. Then if we set the security of the PRF to be secure against adversaries with running time of $O\left(2^{2|x|}\right)$. Using such PRFs we can claim that,

$$\Pr\left[\left\langle \mathsf{mP}, \tilde{\mathsf{V}} \right\rangle = 1\right] \geq \mathop{\mathbb{E}}_{R}\left[\Pr\left[\left\langle \mathsf{mP}, \tilde{\mathsf{V}}^R \right\rangle\right]\right] - \mathrm{negl}(\lambda) \quad .$$

Then, using Proposition 5.9 for $\mathsf{mP}^{\tilde{\mathsf{V}}^R}$ (when viewing $\mathsf{mP}^{\tilde{V}^R}$ as A^{f^R} for $f^R(x, y, z; R(x) R(x, y)) = \mathsf{V}(x, z; R(x))$ and the predicate being outputting an accepting transcript on a false instance) we can argue that there exists B such that B first outputs an instance x and first message a, then gets oracle access to $f^R|_x(\cdot) = \mathsf{V}(x, \cdot; R(x))$ and outputs some transcript ts on x such that ts is accepting and $x \notin \mathcal{L}$ with only a polynomial multiplicative loss compared to the success probability of mP. Then note that $\mathsf{V}(x, \cdot; r)$ and $\mathsf{V}(x, \cdot; R(x))$ are perfectly indistinguishable oracles for uniformly sampled r, R. Hence, we can change the oracle given to B to be $\mathsf{V}(x, \cdot; r)$ without changing the success probability of B.

However, this is still not a regular resetting prover, since the instance x B interacts on is non-deterministic, and we get the oracle corresponding to the instance the first stage outputted. To fix this, consider for an instance x and first message a the residual purified state after B's first stage outputting it $|\psi(x, a)\rangle$. By averaging, there exists some instance $\bar{x}, \bar{a}$ such that $\bar{x}, \bar{a}$ maximizes the success probability of B. Our resetting prover will prove against $\bar{x}$ with first message $\bar{a}$, and use $|\psi(\bar{x}, \bar{a})\rangle$ to execute the second stage of B (while using the oracle access given to $\mathsf{V}(\bar{x}, \cdot; r)$ to simulate the access to f^R) □