Fiat-Shamir for Proofs Lacks a Proof Even in the Presence of Shared Entanglement

Frédéric Dupuis[1], Philippe Lamontagne[2,1], and Louis Salvail[1]

[1]Université de Montréal (DIRO), Montréal, Canada
[2]National Research Council Canada, Ottawa, Canada

We explore the cryptographic power of arbitrary shared physical resources. The most general such resource is access to a fresh entangled quantum state at the outset of each protocol execution. We call this the *Common Reference Quantum State (CRQS)* model, in analogy to the well-known *Common Reference String (CRS)*. The CRQS model is a natural generalization of the CRS model but appears to be more powerful: in the two-party setting, a CRQS can sometimes exhibit properties associated with a Random Oracle queried once by measuring a maximally entangled state in one of many mutually unbiased bases. We formalize this notion as a *Weak One-Time Random Oracle* (WOTRO), where we only ask of the m–bit output to have *some* randomness when conditioned on the n–bit input.

We show that when $n - m \in \omega(\lg n)$, any protocol for WOTRO in the CRQS model can be attacked by an (inefficient) adversary. Moreover, our adversary is efficiently simulatable, which rules out the possibility of proving the computational security of a scheme by a fully black-box reduction to a cryptographic game assumption. On the other hand, we introduce a non-game quantum assumption for hash functions that implies WOTRO in the CRQ$ model (where the CRQS consists only of EPR pairs). We first build a statistically secure WOTRO protocol where $m = n$, then hash the output.

The impossibility of WOTRO has the following consequences. First, we show the fully-black-box impossibility of a *quantum* Fiat-Shamir transform, extending the impossibility result of Bitansky *et al.* (TCC 2013) to the CRQS model. Second, we show a fully-black-box impossibility result for a strenghtened version of quantum lightning (Zhandry, Eurocrypt 2019) where quantum *bolts* have an additional parameter that cannot be changed without generating new bolts. Our results also apply to 2–message protocols in the plain model.

Contents

1 Introduction

Cryptographic protocols can sometimes only be proven secure if some of their components are assumed to be ideal. For example, some protocols that make use of cryptographic hash functions can be proven secure if they are modelled as ideal random functions provided as a black box; this is called the *random oracle model (ROM)*. Another, but weaker, idealized resource is the *common random string model (CRS)*, in which the participants get a freshly generated random string at the outset of each protocol execution. Several cryptographic applications have their most efficient protocols proven secure when provided access to such extra resources, as all known protocols in the plain model are either inefficient, or do not satisfy all security requirements.

The Random Oracle Model (ROM). Introduced by Bellare and Rogaway [BR93] as a way to idealize cryptographic hash functions, the model has been shown to provide formal security proofs for a wide variety of cryptographic protocols that are not known to be secure under standard assumptions in the plain model. A random oracle models a hash function as one whose value for every input is chosen uniformly and independently at random and afresh before each protocol execution. This is meant to model the assumption that a hash function is random, and that looking at its source code yields nothing useful beyond its input-output behaviour. Rigorous security proofs for practical and efficient applications like *Full Domain Hash signatures (FDH-Signatures)*, *Optimal Asymmetric Encryption Padding (OAEP)*, Schnorr's signatures [Sch89; Seu12], and Fischlin's NIZK-PoK [Fis05] are easy to obtain in the ROM but are still missing in the plain model. The random oracle is a powerful primitive that provides all the main properties of a cryptographic hash function at once: collision resistance, preimage resistance, and pseudorandomness. It also has properties that can never be satisfied by any hash function: programmability, (query) extractability (also known as observability), and freshness.

Common Reference String Model. A CRS is nothing more than a fresh random string that materializes upon each protocol execution (freshness) and to which all players have access. This model was originally proposed by Blum, Feldman, and Micali [BFM88] to help remove interaction in zero-knowledge proof systems. In [Blu+91], the model was shown to allow for non-interactive zero-knowledge for all NP languages. The works of [Can01; CF01; DN02] extend its use as a resource enabling universally composable cryptographic primitives. The common reference string model comes in two main flavours. The weakest consists of a random and uniform string of polynomial length (in the security parameter) while the strongest consists of a string of polynomial length picked from some efficiently sampleable distribution. The first flavour will be denoted by the *CR$ model (i.e. the Common Random String Model)* while the second flavour will be denoted by the *CRS model (i.e. the Common Reference String Model)*.

A customary application of both the CRS model and the ROM is the removal of interaction in interactive proof systems. As mentioned above, the CRS model was originally designed for that purpose [BFM88]. Notice that a random oracle is a much more powerful resource than a CRS, since it provides random access to an exponential number of them. However, a random oracle is an immaterial resource as its properties could never be satisfied by any efficient local process. This is in sharp contrast to a CRS, which can be implemented in practice: we only need a way to publish fresh and public random strings of polynomial length. Unfortunately, some basic and useful cryptographic primitives are only known to be securely realizable in the ROM.

When Entanglement Behaves Like a Random Oracle. In order to see why entanglement could outperform a CRS in some settings, consider the following scenario where it seems to provide as much randomness as the random oracle. Suppose Alice prepares n EPR pairs of qubits and sends half of each pair to Bob. Each can then view their n qubits as an access to a weak random oracle implementing a random function $f : \{0,1\}^n \to \{0,1\}^n$. The value $f(a)$ can be obtained the following way. To each possible value $a \in \{0,1\}^n$, we associate a publicly known orthonormal basis θ_a for n qubits. The value of $f(a)$ is simply defined as the outcome of the measurement of the n qubits owned by each party in basis θ_a. Notice that this weak random oracle can be queried in only one place by each party, as after the measurement is performed, the entangled pairs have collapsed to a classical state. However, when both parties measure in the same basis θ_a they obtain the same

uniformly distributed outcome. Moreover, when the bases $\{\theta_a\}_{a\in\{0,1\}^n}$ are chosen to be mutually unbiased [Sch60; WF89][1], the value $f(a)$ does not provide any information about $f(a')$ for any $a \neq a'$.

In this particular setting, n EPR pairs seem to contain as much randomness as a random oracle. It is therefore tempting to believe that an entangled state of polynomial size could in certain cases provide a cryptographic resource tantamount to the random oracle when only one query (or just a few) has to be made by each player. Such a resource, which we call a *Weak One-Time Random Oracle* (WOTRO), would be a powerful primitive for removing interaction in procotols, even if it only provides *some* randomness: that the value $f(a)$ is not a deterministic function of a. The above scheme can be made non-interactive if Alice and Bob share EPR pairs ahead of time. This motivates our study of a model in which parties have access to a pre-shared entangled state.

The CRQS and CRQ$ Models. In this paper, we consider models where a quantum state plays the role of a common random string in a situation involving two parties. In the CRQS (*Common Reference Quantum State*) model, each party receives one half of a fixed pure quantum state at the beginning of each protocol execution. The shared quantum state is of polynomial size and can be generated by some polynomial size quantum circuit. In the *CRQ$* model, each player is given halves of polynomially many (in the security parameter) maximally entangled pairs of qubits (or qudits in general). Although we could allow a CRQS or a CRQ$ to be shared between more than two parties, in this work we only consider the two-party case. Notice that the meaning of *common* in CRQS and CRQ$ is narrower than for a CRS and CR$: even though a CRQS is common to both parties involved in a protocol, it is completely unknown to anybody else, as both players share a pure state. Even though a CRQS is obviously more difficult to deploy in practice than a CRS, it remains a physical resource, unlike the random oracle. Establishing limits on what a CRQS can provide would therefore contribute to a better understanding of the cryptographic power provided by the sharing of a *physical* resource between the parties involved in a protocol.

WOTRO in the CRQS model? We investigate the question of whether or not WOTRO has a secure instantiation in the CRQS model. Like the CRS and ROM, quantum entanglement is known to allow the reduction of interaction, but it also enables tasks that would be classically impossible using only a CRS. Watrous [Wat03] showed that every language in PSPACE has 3-message proof systems. Another example would be nonlocal games such as the magic square game [Ara02; Ara03; BBT05], where a pair of entangled non-interacting provers can win a game that would classically require them to communicate.

The CRQS model provides quantum non-local correlations[2] between the prover and the verifier. Non-local correlations are often idealized by non-local (PR) boxes [PR97]. One PR box takes the first party's input $a \in \{0,1\}$ and the second party's input $b \in \{0,1\}$ to provide $u \in \{0,1\}$ and $v \in \{0,1\}$ such that $u \oplus v = a \wedge b$ to the first and second party respectively. EPR pairs achieve this functionality with probability of success $\cos^2(\frac{\pi}{8})$ while any CRS would not be able to provide the correct answer with probability better than

[1] $\{\theta_a\}_{a\in\{0,1\}^n}$ is a set of mutually unbiased bases for n qubits if for all $|u\rangle \in \theta_a$ and $|v\rangle \in \theta_{a'}$ with $a \neq a'$, we have $|\langle u|v\rangle|^2 = 2^{-n}$. There are $2^n + 1$ mutually unbiased bases for n qubits.

[2] In quantum mechanics, a *non-local correlation* is the name given to the statistics of local measurements applied to distinct parts of a quantum states when they cannot be explained by a local realistic theory. Non-local correlations here (quantum or not) means also that they do not allow for any form of communication as they must be compatible with special relativity.

$\frac{3}{4}$. It is not too hard to see that access to sufficiently many PR boxes allows for a secure implementation for WOTRO (see details in Section 4.2). While PR boxes are not physical objects, the question we are addressing here is whether non-local *quantum* correlations can be harnessed to provide a functionality akin to the use of a random oracle queried once through the use of a CRQS.

One might argue that the CRQS model is not currently realistic given the technological difficulties associated with distributing and coherently storing quantum entanglement (although this is rapidly improving). However, we ask a more fundamental question on the power of setup assumptions. Are there physically realizable setup assumption that allows to solve problems that, in the classical model, appear to require a random oracle.

The Fiat-Shamir Transform. One very useful primitive that needs an idealized cryptographic resource like WOTRO for establishing its security is the Fiat-Shamir transform, also known as the Fiat-Shamir *heuristic*, introduced in the pioneering work of Fiat and Shamir in [FS86] as a way to transform identification schemes of a given form into practical digital signature schemes. More generally, the FS-transform is a simple and efficient primitive allowing to convert sound interactive proof systems of a particular form into non-interactive arguments for the same language. Its primary use is to remove interaction in Σ–protocols.

Σ–protocols [Cra96; Dam10] are public-coin 3-message proof systems where, from public input $x \in \{0,1\}^*$, the prover sends a *commitment* $a \in \{0,1\}^n$ to the verifier as the first message. The verifier then replies with a random *challenge* $c \in_R \{0,1\}^m$ (called *public coins*) before the prover sends the answer $z(x,a,c)$ that the verifier can check for consistency. Henceforth, Σ–protocols with commitments of size n and public coins of size m will be denoted by $\Sigma_{n,m}$*–protocols*. These proof systems can be proofs of knowledge, like their use in identification schemes, or proofs of language membership. In this paper, Σ–protocols are always considered perfectly correct and special sound. Special soundness[3] for proofs of knowledge means that from any two successful conversations with the same commitment $(a,c,z(x,a,c))$ and $(a,c',z(x,a,c'))$ with $c \neq c'$, one can efficiently extract a witness π for $x \in L$. For proofs of language membership, special soundness means that when $x \notin L$ and for each commitment a, there exists at most one challenge $\mathsf{c}(a)$ for which a third message $\tilde{z}$ can ever be found such that $(a,\mathsf{c}(a),\tilde{z})$ is accepted by the verifier.

The Fiat-Shamir transform applied to a Σ–protocol is implemented using hash function $h_r : \{0,1\}^* \to \{0,1\}^m$ picked according to CR\$ r. The prover then sends $(a, h_r(a), z(x,a,h_r(a)))$ to the verifier. In other words, the verifier's challenge or public coin c in the Σ–protocol is replaced by $c = h_r(a)$[4]. It is straightforward to see that when h_r is modelled by a random oracle, the transform applied to a Σ–protocol produces a sound argument The family of hash functions $\mathcal{H} = \{h_r\}_{r \in_D \{0,1\}^{\ell(n)}}$, for D an efficiently sampleable distribution over $\{0,1\}^{\ell(n)}$, is a *sound* $\Sigma_{n,m}$*–universal instantiation* of the Fiat-Shamir transform if h_r converts the special soundness of any $\Sigma_{n,m}$–protocol (as a proof of language membership) into a non-interactive argument. Notice that when the hash function is modelled by a random oracle, the prover and the verifier only have to query the oracle once at the same point. Replacing the random oracle with a secure instantiation of WOTRO would thus provide a sound universal Fiat-Shamir transform.

[3] Special soundness is called *optimal soundness* in [BLV06].

[4] Some works include the public instance x as input to h_r, our results remain untouched if we include it. We leave it out for simplicity.

The Fiat-Shamir Transform in the ROM and QROM. As mentioned above, the Fiat-Shamir transform was shown secure in the ROM by Pointcheval and Stern [PS96] in 1996. The soundness of the Fiat-Shamir transform is straightforward in the ROM. The challenging part was to show that it also provides non-interactive proofs of knowledge. The same was shown to hold in the quantum random oracle (QROM) independently and differently by Don, Fehr, Majenz, and Schaffner in [Don+19] and by Liu and Zhandry in [LZ19].

Known Impossibility Results for the Fiat-Shamir Transform. The Fiat-Shamir transform does not guarantee computational soundness for all Σ–protocols in the CRS model. In particular, Goldwasser and Kalai have shown that the Fiat-Shamir transform applied to some (contrived) Σ–protocols is not sound for any instantiation of the hash function (i.e. instantiated using a CRS) [GK03]. However, this impossibility result requires the Σ–protocol to be a proof of knowledge.

Ambainis, Rosmanis, and Unruh [ARU14] have shown that the Fiat-Shamir transform cannot preserve the soundness of every Σ–protocol against quantum adversaries, even when it is instantiated with a random oracle. More precisely, they construct a proof system, which can be either a proof of knowledge or an argument of language membership, which is sound classically but unsound against quantum adversaries. The same holds true when the Fiat-Shamir transform is applied to these proof systems. In effect, their attack is against the underlying Σ–protocol rather than against a physical instantiation of a random oracle. Their results do not contradict the positive results of [Don+19; LZ19] since they show that the Fiat-Shamir transform preserves soundness in the QROM when the underlying Σ–protocol is sound against quantum adversaries.

Impossibility results for Σ–protocols used as proofs of language membership are not known to be as strong as for proofs of knowledge. One reason being that for language membership, the Fiat-Shamir transform is only asked to provide computational soundness to a Σ–protocol with statistical soundness whereas for a proof of knowledge the Σ–protocol is an argument. Remember that a *cryptographic game* [HH09] is a standard way to define computational assumptions by requiring that no adversary can win an interactive game against a *challenger* with probability that is not overwhelmingly close to some constant value [HH09]. An assumption that can be formulated as a cryptographic game with an efficient challenger is called a *falsifiable assumption* [GW11; Nao03]. Known impossibility results for the Fiat-Shamir transform applied to Σ–protocols for proofs of language membership are about the impossibility of *black-box* reducing its computational soundness to a cryptographic game.

In [Bit+13], Bitansky *et al.* provide two results on the impossibility of establishing the computational soundness of the Fiat-Shamir transform in the CRS model. First, if a language $L \notin BPP$ has an honest-verifier zero-knowledge (HVZK) Σ–protocol (with small enough challenges) then the soundness of the Fiat-Shamir transform applied to it cannot be established by a black-box reduction[5] to a falsifiable assumption[6]. This impossibility result

[5]The security of protocol Π is black-box reduced to an assumption expressed as a game if there exists an oracle polynomial-time machine $\mathcal{R}^{P^*}$ that, with oracle access to any successful adversary P^* for protocol Π, wins the game.

[6]The reason why the result applies in the CRS model is because [Bit+13; Dac+12] show how to get, from such a Fiat-Shamir transform, a 2-message zero-knowledge proof system for L where the verifier simply sends the identity of the hash function to the prover as first message. This is equivalent to non-interactive schemes in the CRS model. These proofs systems are shown impossible by an extension of the impossibility result for 2-round zero-knowledge for non-trivial languages by Goldreich and Oren [GO94].

applies even to Fiat-Shamir transforms tailor-made for specific Σ–protocols. Second, they show the impossibility of black-box reducing the computational soundness of any universal instantiation of the Fiat-Shamir transform to a cryptographic game, even a non falsifiable one where the challenger is not required to run in polynomial time.

Positive results & related work. A series of results have been focusing on achieving soundness of the Fiat-Shamir transform from a cryptographic assumptions that cannot be black-box reduced to cryptographic games. Barak, Lindell and Vadhan [BLV06] introduce the notion of *entropy preserving* hash functions (such function families are rather said to *ensure conditional entropy* therein) and show that assuming their existence, there is no constant-round auxiliary-input zero-knowledge proof system for non-trivial languages. The proof of this result implies the computational soundness of Fiat-Shamir using entropy preserving hash functions. Later, Dodis, Ristenpart and Vadhan [DRV12] gave a construction for entropy preserving hash functions assuming the existence of robust randomness condensers with some extra properties, but without providing any candidate construction. Canetti, Goldreich, and Halevi [CGH04] introduce *correlation intractable* families of hash functions. Correlation intractability is related to entropy preservation as the latter implies the former. Therefore, a consequence of [Bit+13] is that correlation intractability cannot be proven by black-box reduction to a game. In [KRR17], Kalai, Rothblum, and Rothblum provide a construction for correlation intractable family of hash functions from a subexponentially secure indistinguishability obfuscator, an exponentially secure input-hiding obfuscator for the class of multi-bit point functions, and the existence of a subexponentially secure puncturable PRF[7]. The subexponential indistinguishable security of the IO-obfuscator and the exponential security of the multi-bit point functions obfuscator allow to evade the impossibility result of [Bit+13]. In [Can+18], Canetti, Chen, Holmgren, Lombardi, Rothblum, and Rothblum show how to construct a universal instance of the Fiat-Shamir transform using correlation intractable hash functions built from a strong version of KDM-encryption. The resulting Fiat-Shamir transform also has security black-box reducible to a cryptographic game with subexponential security.

The concept of shared entanglement as a setup was considered in previous works. In [CVZ20], Coladangelo, Vidick, and Zhang have shown how to design non-interactive zero-knowledge arguments for QMA (i.e. quantum NP), with preprocessing. Morimae and Yamakawa [MY22] use a similar setup for classical verifiability of NIZK arguments for QMA. The preprocessing is essentially what we call here a CRQ$. Non-interactivity is obtained from pre-shared EPR pairs used as a teleportation channel. This can be viewed as a quantum version of the work of Peikert and Shiehian [PS19] and, as such, is not a $\Sigma_{n,m}$–universal instantiation of the Fiat-Shamir transform. The ability of a CRQS to provide zero-knowledge against quantum dishonest verifiers has been investigated in [DFS04]. It was shown that a CRQS allows quantum zero-knowledge implementations of a Σ–protocols against a relaxed form of honest verifiers, called *non-oblivious*.

A model called CRQS was recently[8] introduced in [MNY24] as a trusted setup for provably computationally secure quantum bit commitment (i.e. without relying on complexity assumptions). In the model of [MNY24], a setup algorithm samples a classical key k and distributes copies of a quantum state $|\psi_k\rangle$ to each party. A similar model,

[7]Notice that the result of [KRR17] is very general as it allows to apply securely the Fiat-Shamir transform to any public-coin 3-message proof systems, not only to Σ–protocols as we define them. Some of their assumptions can be relaxed a little when the Fiat-Shamir transform is applied to Σ–protocols.

[8]The authors of [MNY24] were aware of our work but decided to use the same naming scheme, arguing that our model is mode akin to the quantum version of correlated randomness rather than CRS.

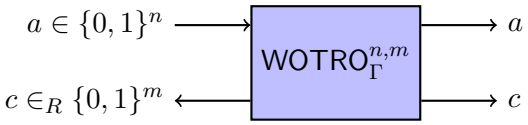

Figure 1: $\mathsf{WOTRO}^{n,m}$ as a box. The prover on the left puts a chosen $a \in \{0,1\}^n$ into it, the box chooses $c \in_R \{0,1\}^m$, and outputs (a, c) to the verifier on the right-hand side.

called "unclonable common random state", is independently introduced in [Qia24] for the same task of unconditionally secure quantum bit commitment. In the context of a negative result, the more general the model is, the stronger is the impossibility result. The models of [MNY24; Qia24] are a special case of ours by considering a CRQS of the form $|\Psi\rangle = \sum_k \sqrt{p(k)}|\psi_k\rangle|\psi_k\rangle$ where $p(k)$ denotes the probability to pick key k.

1.1 Our Contributions

We introduce a cryptographic primitive called a *Weak One-Time Random Oracle*, denoted $\mathsf{WOTRO}^{n,m}$ and defined by the box given in Fig. 1, which takes place between a "prover" who controls the interfaces on the left-hand side of the box, and a "verifier" who controls the interfaces on the right. A protocol instantiating $\mathsf{WOTRO}^{n,m}$ is secure if for any function $f(\cdot)$, the adversary can't produce an output of the form $(a, f(a))$ on the verifier's interface. We ask whether this primitive has a secure non-interactive instantiation in the CRQS model. Our main contribution is showing that, despite the evidence to the contrary presented above, this primitive has no statistically secure implementation in the CRQS model. Our impossibilities also apply for two-message protocols in the plain model (and even in the CRQS model) since the CRQS could be prepared by the verifier.

Theorem 1 (informal version of Theorem 8) *If $n - m \in \omega(\lg n)$, there is no statistically secure non-interactive protocol for $\mathsf{WOTRO}^{n,m}$ in the CRQS model.*

For any protocol in the CRQS model, we construct an (inefficient) attack that will make the verifier accept an output of the form $(a, f(a))$ for a function f chosen at random. Our attack is a novel use of the operator Chernoff bound of Ahlswede and Winter [AW02].

What about WOTRO protocols that provide computational security by relying on a hardness assumption? We show that such a protocol could not be proven secure using reductions that treat the adversary as a black-box (i.e. a CPTP map). We call this a quantum fully-black-box (or f-BB) reduction.

Theorem 2 (informal version of Corollary 1) *If $n - m \in \omega(\lg n)$, there is no protocol for $\mathsf{WOTRO}^{n,m}$ whose security can be established by a quantum f-BB reduction to a cryptographic game assumption, unless that assumption is false.*

The statement above and its proof are similar to the impossibility results of [Bit+13; BGW12] in the context of Fiat-Shamir in the CRS model. We rely on a technique formalized by Wichs in [Wic13]. We show that the input/output behaviour of our attacker against any WOTRO protocol can be *simulated* efficiently by a quantum algorithm. This means that no reduction can exist that breaks the security of a cryptographic game assumption using only the input/output behaviour of a successful adversary against WOTRO, unless the assumption is false. Otherwise, the reduction together with the simulator for the attack would yield an efficient algorithm for breaking the assumption.

While WOTRO implies Fiat-Shamir, the other direction does not hold. Still, we can use the attack from our impossibility of WOTRO to obtain a similar result ruling out any universal instantiation of Fiat-Shamir in the CRQS model.

Theorem 3 (informal version of Theorem 10) *For $n - m \in \omega(\lg n)$, there is no $\Sigma_{n,m}$–universal instantiation of the Fiat-Shamir transform whose security can be established by quantum f-BB reduction to a cryptographic game assumption, unless that assumption is false.*

Interestingly, our impossibility is more general than the classical one [Bit+13], even when restricted to classical shared resources. A CRQS can capture as a special case asymmetric setups such as giving the verifier the trapdoor to some primitive the prover uses or pre-computed randomized oblivious transfers. We obtain this generality "for free" by considering the cryptographic primitive WOTRO instead of a family of cryptographic hash functions, as in [Bit+13].

Studying the WOTRO primitive instead of Fiat-Shamir directly has another advantage in that our impossibility result also applies to any cryptographic task which (black-box) implies WOTRO. For instance, we introduce a strengthened variant of Zhandry's quantum lightning [Zha19] that implies WOTRO. Quantum lightning (QL) is a primitive that produces a quantum state and an associated serial number such that no adversary can produce two states with the same serial number (hence the name "lightning"). A consequence of this property is that serial numbers are highly unpredictable. A natural question is whether some form of metadata can be embedded into quantum lightning such that changing the value of this metadata requires creating a new lightning state. This metadata could for example contain ownership information and it would thus be impossible, even to the emitter of the state, to change the owner of a state without generating an entirely new state. It could also serve to encode a denomination for quantum bank notes, such that not even the emitting bank could change the denomiation of an existing quantum note.

We introduce a variant of quantum lightning that allows such metadata by adding a classical input to the state generation procedure. We call this variant *typed quantum lightning* (tQL) which is secure if the serial numbers remain unpredictable conditioned on the input. We show that this variant implies WOTRO and thus inherits the same black-box impossibility.

Theorem 4 (informal version of Corollary 2) *There is no quantum f-BB reduction from the security of a tQL scheme to the security of a cryptographic game assumption when type length n and serial length m satisfy $n - m \in \omega(\lg n)$, unless that assumption is false.*

Why would tQL be a reasonable assumption? Clearly it is a very powerful primitive, but how much of a leap is it from "vanilla" quantum lightning? While we do not have a definitive answer to that question, we can show that QL implies tQL with small types. More precisely, we construct in Section 8.2 a tQL scheme from regular QL for types of $O(\lg(n))$ bits.

Instantiating WOTRO from a non-game assumption. We show that it is possible to construct a WOTRO protocol for which security is based on a cryptographic assumption that does not fit the game formalism. Our result is based on a new hardness assumption on cryptographic hash functions called *collision-shelters*. Intuitively, a family of hash functions is a collision-shelter if no adversary can produce many collisions *in superposition*.

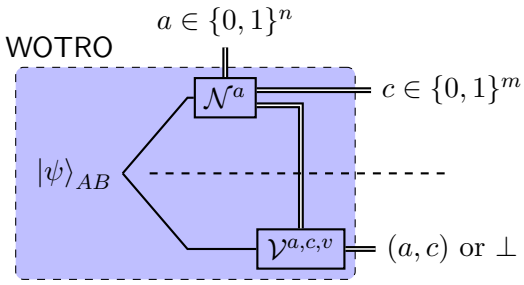

Figure 2: A $\mathsf{WOTRO}^{n,m}$ protocol in the CRQS model as described in Definition 6. The prover's actions are above the dashed line and the verifier's actions are below. The classical wire crossing the dashed line represents the classical message sent from the prover to the verifier.

As such it is an intrinsically quantum definition which cannot be framed as a game since no challenger can verify that an adversary breaks the assumption. Using this assumption, we show how to construct a secure $\mathsf{WOTRO}^{n,m}$ protocol in the CRQ$ model. We first prove the security of a construction for $\mathsf{WOTRO}^{n,n}$ similar to the one based on EPR pairs and mutually unbiased bases sketched earlier. The proof involves computing bounds on the optimal probability of distinguishing between states from many mutually unbiased bases and might be of independent interest. A $\mathsf{WOTRO}^{n,m}$ protocol for $m < n$ is obtained by hashing the output with a collision-shelter hash function.

Theorem 5 (informal version of Theorem 12) *Under the collision-shelter assumption, there are secure instantiations of $\mathsf{WOTRO}^{n,m}$ in the CRQ$ model.*

2 Technical Overview

We call *Weak One-Time Random Oracle*, denoted $\mathsf{WOTRO}^{n,m}$, the following simple non-interactive primitive. To any $a \in \{0,1\}^n$, it provides a challenge $c \in \{0,1\}^m$ *avoiding* with good probability any function $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$. We say that an implementation of $\mathsf{WOTRO}^{n,m}$ *avoids* function c if no (efficient) dishonest prover is able to produce (a,c) such that $c = \mathsf{c}(a)$. An implementation of $\mathsf{WOTRO}^{n,m}$ is said to be κ–secure if it behaves like a random oracle when the prover is honest and avoids any function c with probability at least κ, when the prover is dishonest. It is easy to see that any non-interactive κ–secure implementation of $\mathsf{WOTRO}^{n,m}$ can be used to implement the Fiat-Shamir transform with computational soundness error upper-bounded by $1-\kappa$ (see Section 4.1). Any implementation of $\mathsf{WOTRO}^{n,m}$ that avoids any function $\mathsf{c}(\cdot)$ would be a powerful cryptographic primitive to remove interaction. An implementation $\Pi^{n,m}_{\mathsf{WRO}} = (\mathsf{P}', \mathsf{V}')$ of $\mathsf{WOTRO}^{n,m}$ in the CRQS model is defined by two families of efficient POVMs $\mathsf{P}' = \{\mathcal{P}^a\}_a$ and $\mathsf{V}' = \{\mathcal{V}^{a,c,v}\}_{a,c,v}$ with $a \in \{0,1\}^n$, $c \in \{0,1\}^m$, and v is an auxiliary string announced to V'. $\Pi^{n,m}_{\mathsf{WRO}} = (\mathsf{P}', \mathsf{V}')$ is executed as folllows:

1. Upon input $a \in \{0,1\}^n$, P' applies POVM $\mathcal{P}^a := \{P^a_{c,v}\}_{c,v}$ to register P of the CRQS to get classical outcome (c,v). P' then announces (a,c,v) to V'.

2. V' applies POVM $\mathcal{V}^{a,c,v} := \{V_0^{a,c,v}, V_1^{a,c,v}\}$ to register V of the CRQS and accepts iff classical outcome 1 is obtained.

An adversary $\mathcal{A}$ against $\Pi^{n,m}_{\mathsf{WRO}}$ takes no input and applies a POVM $\mathcal{A} := \{\mathcal{A}_{a,c,v}\}_{a,c,v}$ to register P of the CRQS to obtain a along with the message (c,v). Notice that as defined, $\Pi^{n,m}_{\mathsf{WRO}}$ requires the message transmitted to V' to be classical. This can be done without

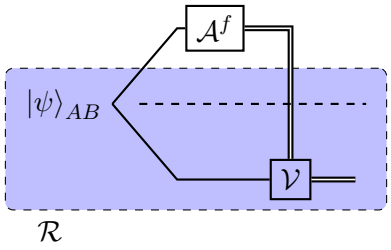

Figure 3: The interface between the f-BB reduction $\mathcal{R}$ and the WOTRO adversary $\mathcal{A}^f$. The reduction simulates the CRQS $|\psi\rangle_{AB}$ and the verifier. It provides register A to the adversary $\mathcal{A}^f$ and receives its classical outcome.

loss of generality as a protocol asking P' to send a quantum message can be transformed into one where P' only sends a classical message by adding to the CRQS enough EPR pairs for the quantum message to be teleported. The security of the original protocol remains untouched by this transformation.

2.1 The impossibility of $\mathsf{WOTRO}^{n,m}$ in the CRQS model.

For main contribution (Theorem 1), we construct an (inefficient) adversary $\mathcal{A}^f$ that picks a random function $f : \{0,1\}^n \to \{0,1\}^m$ such that the verifier will always accept the outcome $(a, f(a))$ in the protocol. $\mathcal{A}^f$ mounts its attack using the prover's honest POVM operators $\mathcal{P}^a_{c,v}$ using the following *attack operators*: $X^f_a := \sum_v \mathcal{P}^a_{f(a),v}$. The success of this attack relies on two crucial facts. First, the quantum operator Chernoff bound of Ahlswede and Winter allows us to show that the $\{X^f_a\}_{a\in\{0,1\}^n}$ (almost) form a POVM with overwhelming probability over the choice of f. Second, since the attack uses the honest prover operators, the verifier will accept with the same probability as with the honest prover.

As for the impossibility of computationally secure WOTRO (Theorem 2), we use a proof strategy similar to that of Bitansky *et al.* in [Bit+13; BGW12] when proving that there exists no black-box reduction from any successful adversary against the entropy preserving property of a family of hash functions to a cryptographic game. That is, we show that our adversary against $\mathsf{WOTRO}^{n,m}$ is *simulatable* by an efficient quantum circuit. The main difference here is that our quantum circuit is stateless while it is stateful in [Bit+13; BGW12]. This prevents the security of $\mathsf{WOTRO}^{n,m}$ to be established by a reduction to any cryptographic game that treats the adversary as a black-box (according to Definition 9), as if there was such a reduction the game would also be won using the efficient simulator (in other words, the game assumption would be false).

2.2 Quantum Black-Box Reductions.

We should make precise what we mean by quantum black-box reductions. The classical notion of black-box reductions is uncontroversial; the reduction is an efficient algorithm R having black-box (i.e. input/output) access to an adversary $\mathcal{A}$ breaking a scheme. In the quantum setting, the reduction itself can be quantum, i.e. be an oracle access quantum circuit (Definition 8), or it can be classical; and the adversary $\mathcal{A}$ can also be quantum or classical. Firstly, for any reasonable definition of fully black-box, R should not be able to tell if $\mathcal{A}$ is quantum or classical. It has been argued that if R is quantum, R can be purified as a unitary circuit, and since unitary circuits are reversible, R should have access to $\mathcal{A}$ and its inverse $\mathcal{A}^*$ in order to preserve this property. While this is sometimes called quantum black-box, it is closer to the classical *semi-black-box* notion, where the

underlying primitive is treated as a black-box, but where the reduction can depend on the adversary [BBF13]. These quantum reductions have been called *quantum semi-black-box reductions* in [CX22].

While this type of reduction is very useful (to allow rewinding, for instance), we argue that it cannot truly be considered "black-box". Most realistic models of quantum computation (including all current prototypes for quantum computers) include irreversible operations as part of their native gate sets (for control flow if nothing else). Given a quantum algorithm written in, say, openQASM, one would need access to the source code to get a unitary circuit that can be run backwards; it requires "opening the box" to the same extent as running a homomorphically-encrypted version of an adversary.

Notice also that unlike (fully) black-box reductions, semi-black-box reductions require the adversary to be efficient, otherwise the reduction implemented as a quantum circuit with $\mathcal{A}$ and $\mathcal{A}^*$ gates, would potentially need to feed exponentially-many auxiliary input wires to these unitaries. (Fully) Black-box reductions should never be affected by how the adversary is implemented. Basically, a blackbox reduction remains efficient even when the adversary is not (when oracle calls to $\mathcal{A}$ are at unit cost). This is a crucial property of fully-blackbox reductions.

In view of the above, we adopt the following definition of quantum fully black-box reductions (f-BB): QPT algorithms R that have oracle access to a CPTP map implementing the adversary $\mathcal{A}$. This is the true quantum analogue of the kind of black-box reduction in impossibility results such as [Bit+13]. Note that by weakening the notion of black-box, for example by giving reversible access to the adversary, the set of possible reductions increases, so black-box impossibility or separation results become harder to find. In particular, the result of [Bit+13] is not known to, and probably does not, hold for this kind of reductions. Our results of Section 6 therefore strictly generalizes [Bit+13] since we consider the quantum variant of classical f-BB reductions. We expect that extending our impossibility result to the semi-blackbox setting would require completely new techniques. On a positive note, the impossibility that we prove here could be avoided when the reductions considered are not f-BB. Notice, however, that most security proofs in post-quantum cryptography proceed by f-BB reductions. Most relevant to what we are doing here is the security of the Fiat-Shamir transform in the QROM as shown in [Don+19; LZ19]. In these two papers, the soundness of the Fiat-Shamir transform is established by f-BB reductions. Rewinding the adversary is only required to extract the witness, and since this is only needed to show that the Fiat-Shamir transform is a proof of knowledge in the QROM, this is irrelevant for the impossibility result we provide here.

2.3 Σ–universal quantum Fiat-Shamir cannot be f-BB reduced to a game.

We then show that the black-box impossibility of WOTRO (indirectly) implies that the soundness of any $\Sigma_{n,m}$–universal quantum Fiat-Shamir transform cannot be established under the same conditions. As our basic impossibility result is about the security of a cryptographic primitive rather than a property of a family of hash functions (as in [Bit+13; BGW12]), we follow a different path. First, let us discuss what distinguishes $\mathsf{WOTRO}^{n,m}$ from a $\Sigma_{n,m}$–universal quantum Fiat-Shamir transform in the CRQS model.

Consider a Fiat-Shamir transform in the CRQS model. The general form of such a protocol is similar to the generic WOTRO protocol outlined above. I.e. the pover performs a POVM specified by a, the first message in protocol Σ and the verifier performs a binary outcome POVM on its part of the CRQS as specified by the prover's message, along with additional checks according to Σ (see Section 6 for details).

Although this is providing something very close to $\mathsf{WOTRO}^{n,m}$ in its inner workings, it

may not need to avoid all functions to be a computationally sound Σ–universal implementation of the Fiat-Shamir transform. It only needs to avoid functions $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$ such that for some Σ–protocol for some language L, there exists $x \notin L$ for which upon commitment a, only challenge $\mathsf{c}(a)$ has a third message z such that $(a, \mathsf{c}(a), z)$ is a valid transcript. We show that this relaxation on the functions to be avoided by any $\Sigma_{n,m}$–universal Π^{QFS} leads to the same impossibility result than for $\mathsf{WOTRO}^{n,m}$.

The proof follows from the existence of a Σ–protocol $\Sigma^f = (\mathsf{P}, \mathsf{V}^f)$ for membership to the empty language, where $f : \{0,1\}^n \to \{0,1\}^m$ is a random oracle. Although Σ^f only requires V^f to have access to the oracle $f(\cdot)$ to run the protocol honestly, the adversary $\mathcal{A}^f$ has also access to $f(\cdot)$ to mount its attack against the soundness of $\Pi^{\mathsf{QFS}}[\Sigma^f]$. This is essentially the same adversary defined as the one against $\mathsf{WOTRO}^{n,m}$ described above. Notice that if the soundness of Π^{QFS} was f-BB reducible to game $\mathcal{G}$ then there would be an efficient algorithm $\mathcal{B}^f$, having oracle access to $f(\cdot)$, that wins game $\mathcal{G}$. The strategy used for $\mathsf{WOTRO}^{n,m}$ can then be applied. A possibly inefficient adversary $\mathcal{A}^f$ is defined that almost all the time breaks the soundness of $\Pi^{\mathsf{QFS}}[\Sigma^f]$. We finally show that both the adversary $\mathcal{A}^f$ and V^f can be simulated by an efficient stateful simulator. As before, this prevents the soundness of Π^{QFS} to be established by f-BB reduction to a cryptographic game unless the game is trivial.

2.4 A quantum assumption allowing for $\mathsf{WOTRO}^{n,m}$.

We introduce a strong variant of collision resistant families of hash functions allowing for a computationally sound Σ–universal implementation of the Fiat-Shamir transform in the CRQ$ model. We call $\mathbb{G}_\Gamma^{n,m} := \{G_s^n\}_s \subset \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ a *collision-shelter* if, for any target function $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$, no efficient quantum adversary can produce any state with inverse-polynomial trace distance to a state of the form

$$|\psi\rangle_{AX} = \sum_a \alpha_a |a\rangle_A \otimes \sum_{x : G^n(a,x) = \mathsf{c}(a)} \beta_x^a |x\rangle_X \ ,$$

that contains collisions to $\mathsf{c}(a)$ when a is measured.

In order to show that collision-shelters are sufficient for a sound Σ–universal Fiat-Shamir transform in the CRQ$ model, we start with the weak random oracle implemented using n shared EPR pairs from the introduction. We modify the scheme slightly to get an unconditionnally $\frac{1}{4}$–secure[9] implementation $\Pi_{\mathsf{WRO}}^{n,n} = (\mathsf{P}', \mathsf{V}')$ of $\mathsf{WOTRO}^{n,n}$ in the CRQ$ model. This forms the basis upon which $\mathsf{WOTRO}^{n,m}$, with $m < n$, is constructed using a collision-shelter. We prove that $\Pi_{\mathsf{WRO}}^{n,n}$ is $\frac{1}{4}$–secure using shared maximally entangled pairs of qutrits as the CRQ$ to allow the use of a particular set $\{\theta_a\}_{a \in \{0,1,2\}^n}$ of mutually unbiased bases, introduced by Wootters and Fields [WF89]. The set $\{\theta_a\}_a$ is shown to prevent any adversary $\mathcal{A} := \{\mathcal{A}_{a,c,v}\}_{a,c,v}$ from observing $\mathcal{A}_{a,\mathsf{c}(a),v} \otimes \mathcal{V}_1^{a,\mathsf{c}(a),v}$ with probability better than $\frac{3}{4}$ when the CRQ$ is measured by P' and V'. This result may be of independent interest and is made possible as $\mathcal{A}$'s success probability is given by an instance of a Weil sum that can be upper bounded by Deligne's resolution of one of Weil's conjectures [Del74].

A protocol $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}] = (\mathsf{P}'', \mathsf{V}'')$ for $\mathsf{WOTRO}^{n,m}$ with $m < n$ can then be constructed using a collision-shelter $\mathbb{G}_\Gamma^{n,m}$ in the obvious way. Upon input $a \in \{0,1,2\}^n$, P'' runs P' upon input a to get $(c', v) \in \{0,1\}^n \times \{0,1\}^n$. P'' announces (a, c', v) to V''. The challenge produced by $\Pi_{\mathsf{WRO}}^{n,m}$ is simply set to $c := G_s^n(a, c') \in \{0,1\}^m$ for s a CRS. V'' simply runs V' on (a, c', v) and accepts if V' accepts. It is not difficult to see that if $\mathbb{G}_\Gamma^{n,m}$ is a collision-shelter then no efficient adversary $\mathcal{A}$ can do better against $\Pi_{\mathsf{WRO}}^{n,m}$ than an unconditional

[9] By $\frac{1}{4}$–secure, we really mean $\left(\frac{1}{4} - \mathsf{negl}(n)\right)$–secure.

adversary against $\Pi^{n,n}_{\mathsf{WRO}}$. As a result, $\Pi^{n,m}_{\mathsf{WRO}}$ avoids all functions with probability $\frac{1}{4}$. Negligible soundness error can then be achieved by parallel repetitions.

2.5 WOTRO and Quantum Lightning.

Quantum lightning (QL), introduced by Zhandry [Zha19], is a quantum cryptographic task allowing anyone to generate quantum states of which they can make exactly one copy (called the *uniqueness property*). The original construction of Zhandry based on an ad hoc assumption was shown insecure by Roberts [Rob21].

Informally, a QL scheme consists of a quantum algorithm Gen instructing how to construct bolts $|\mathsf{bolt}\rangle$ and of a verification algorithm Ver that on input $|\mathsf{bolt}\rangle$ returns a serial number $s \in \{0,1\}^n$ without disturbing state $|\mathsf{bolt}\rangle$ such that no efficient adversary can create two valid states with the same serial number. For this to hold, there must be uncertainty in the serial number of newly created bolts: for every QPT adversary $\mathcal{A}$, $|\mathsf{bolt}\rangle \leftarrow \mathcal{A}(\mathsf{Gen})$ must satisfy $H_\infty(\mathsf{Ver}(|\mathsf{bolt}\rangle)) \in \omega(\lg n)$, otherwise polynomially many tries would give two bolts with the same serial number, contradicting uniqueness. Note that an efficient reduction does not necessarily exist in the other direction: an adversary could for example produce two valid states with identical serial numbers that each have maximal min-entropy. Such an adversary appears useless for producing a single lightning state with low min-entropy in the serial number.

We introduce a variant of quantum lightning where the bolt generation procedure accepts an input. *Typed quantum lightning* (tQL) is a new primitive similar to QL where Gen takes an additional parameter (or type) $a \in \{0,1\}^n$. Intuitively, security asks that when we fix the type a, the resulting scheme still produces unpredictable serial numbers. This is formalized by requiring that the conditional min-entropy $H_\infty(S \mid A)$ is large. We show that a tQL scheme with type length n and serial number length m implies the existence of a protocol for $\mathsf{WOTRO}^{n,m}$. The scheme asks the prover to generate a typed QL state with type a and teleport that state to the verifier using EPR pairs from a CRQ$, the verifier accepts if the teleported state is a valid tQL state. A consequence if this scheme is that no tQL scheme satisfying $n - m \in \omega(\lg n)$ can have its security be f-BB reducible to a cryptographic game assumption.

3 Notations & Preliminaries

We use $n \in \mathbb{N}$ as the security parameter throughout the paper. We use $\mathsf{poly}(n)$ to denote a polynomial in n. A function $f : \mathbb{N} \to \mathbb{N}$ is said to be *negligible* if for all polynomials $p(\cdot)$ and for $n \in \mathbb{N}$ sufficiently large, $f(n) \leq 1/p(n)$. We denote a negligible function by $\mathsf{negl}(\cdot)$. We use "QPT" as a shorthand for *quantum polynomial time*. We use $\ln(\cdot)$ and $\lg(\cdot)$ to respectively denote the base e and 2 logarithms. To denote a Hilbert space of dimension 2^n, we write $\mathscr{H}_n$.

For a set A, its cardinality is denoted $|A|$ and its complement $\bar{A}$. We write $x \in_R A$ to indicate that x is chosen uniformly at random from A.

We often use the notation $f(\cdot)$ to denote functions as a way to differentiate them from variables. If $f(\cdot,\cdot)$ is a function of two arguments, we denote by $f(x,\cdot)$ the function of one argument defined by restricting the first argument to value x. For two sets A and B, we denote the set of functions from A to B as $A \to B$. Let $\mathcal{F}^{n,m}$ be the set of functions $\{0,1\}^n \to \{0,1\}^m$. We will often view m as a function $m(n)$ of the security parameter n. To simplify the notation, when n and m are clear in the context, we will write $\mathcal{F} := \mathcal{F}^{n,m}$.

For a random variable X, $\mathbb{E}[X]$ denotes its expected value and for $X(r)$ a random variable function of r, $\mathbb{E}_r[X]$ denotes its expected value when r is picked at random. Let $\Delta(A, B) = \frac{1}{2}\sum_a |\Pr[A = a] - \Pr[B = a]|$ denote the statistical distance between the distribution of two random variables A and B with the same domain. For an operator $A \in \mathbb{C}^{n\times n}$, $\|A\|_1 = \mathrm{tr}(\sqrt{A^*A})$ denotes its trace norm.

3.1 Σ–protocols and the Fiat-Shamir Transform

Let $R \subseteq \{0,1\}^* \times \{0,1\}^*$ denote an arbitrary efficiently computable binary relation such that if $(x, w) \in R$ then $|w| \leq p(|x|)$ for some polynomial $p(\cdot)$. We call x a *public instance* and π a *witness* for x. The condition above ensures that the witness of any public instance can be conveyed efficiently. From the binary relation R, we define the language $L_R = \{x \,|\, (\exists\pi)[(x, \pi) \in R]\} \in \mathbf{NP}$ of public instances with witnesses for them.

Definition 1 (Σ–protocol [Dam10]) A Σ–protocol $\Sigma = (\mathsf{P}, \mathsf{V})$ for a binary relation R is a 3-message protocol with *conversation alphabet* $\{0,1\}$. On public input $x \in L_R$ and on private input π to P such that $(x, \pi) \in R$, the protocol structure is as follows:

- The prover sends a message $a = \mathsf{P}_1(x, \pi) \in \{0,1\}^n$ called the *commitment*.
- The verifier sends a *challenge* $c \in \{0,1\}^m$.
- The prover sends a *reply* $z = \mathsf{P}_2(a, x, \pi, c) \in \{0,1\}^*$, and the verifier outputs $\mathsf{V}(x, a, c, z) \in \{\texttt{accept}, \texttt{reject}\}$.

Moreover, the protocol satisfies the following requirements:

Random public coins: The challenge $c \in \{0,1\}^m$ is chosen uniformly at random in $\{0,1\}^m$ without any extra processing (i.e. no need for private information to generate c).

Perfect correctness: When $x \in L_R$, V accepts P with probability 1.

Special soundness: When $x \in L_R$, given two accepting conversations for the same commitment (a, c, z) and (a, c', z') with $c \neq c'$, there exists a PPT algorithm E such that $(x, \mathsf{E}(a, c, z, c', z')) \in R$.

□

We should mention here that Σ–protocols are also often used as a synonym of 3-message public-coins protocols (as in [KRR17; PS19], for instance) irrespectively of whether the proof system satisfies perfect correctness or special soundness. However, since we are proving a negative result, there is no loss in generality in adopting the more restrictive definition of [Dam10].

By special soundness, if $x \notin L_R$ then for any commitment $a \in \{0,1\}^n$, there is at most one challenge $c \in \{0,1\}^m$ such that for some response z, (a, c, z) is an accepting conversation. For some Σ–protocol Σ_L for a language L and some $x \notin L$, we call the function that maps a to this one challenge c the *bad challenge* function.

In the ROM, the Fiat-Shamir transform $\Pi^{\mathsf{FS}}[\Sigma] = (\mathsf{P}^{\mathsf{FS}}, \mathsf{V}^{\mathsf{FS}})$ applied to a Σ–protocol $\Sigma = (\mathsf{P}, \mathsf{V})$ with first message length n and challenge length m for a proof of language membership is a non-interactive argument where, on public input $x \in L$ and random oracle $H : \{0,1\}^n \to \{0,1\}^m$,

1. P^{FS} runs $a = \mathsf{P}(x, w)$ computes $c = H(a)$ and $z = \mathsf{P}_2(a, x, w, c)$, and sends (a, c, z) to V^{FS}.

2. V^{FS} rejects if $c \neq H(a)$, otherwise outputs $\mathsf{V}(x, a, c, z)$.

In the CRS model, the protocol is the same with the random oracle replaced with a family of cryptographic hash functions $\mathcal{H} = \{h_r\}_r$ where $h_r : \{0,1\}^n \to \{0,1\}^m$ is sampled using a CRS.

3.2 Black-Box Impossibility Results

The following define what is meant by a cryptographic game assumption.

Definition 2 ([HH09; Bit+13]) A *cryptographic game* is a tuple $\mathcal{G} = (\Gamma, c)$ composed of an interactive Turing machine Γ and a constant $c \in [0,1]$. On security parameter $n \in \mathbb{N}$, the challenger $\Gamma(1^n)$ interacts with an adversary $\mathcal{A}_n$ and outputs a bit b. The output of this interaction is denoted by $\langle \mathcal{A}_n \rightleftharpoons \Gamma(1^n) \rangle$. The *advantage* of the family of adversaries $\mathcal{A} = \{\mathcal{A}_n\}_{n \in \mathbb{N}}$ in game $\mathcal{G}$ is defined as

$$\mathsf{Adv}_{\mathcal{A},\mathcal{G}}(n) = \Pr[\langle \mathcal{A}_n \rightleftharpoons \Gamma(1^n) \rangle = 1] - c \ .$$

A cryptographic game $\mathcal{G}$ is secure if for all PPT adversary $\mathcal{A}$, the advantage $\mathsf{Adv}_{\mathcal{A},\mathcal{G}}(n)$ is $\mathsf{negl}(n)$. The communication can be classical or quantum. □

Intuitively, a protocol Π for WOTRO has its security reducible to a cryptographic game assumption $\mathcal{G}$ if there exists an efficient way to transform any successful adversary $\mathcal{A}$ against Π into a challenger winning game $\mathcal{G}$. If this transformation works only provided the standard input-output behaviour of $\mathcal{A}$ then we say that the security of Π is f-BB reducible to game $\mathcal{G}$. Quantum black-box reductions are defined formally in Section 5.3.

In this paper, we show the impossibility of black-box reducing the security of cryptographic primitive, called WOTRO, to any cryptographic game. Our proof uses the general technique of simulatable attacks formalized by Wichs [Wic13] and applied by [Bit+13] to the Fiat-Shamir transform. An inefficient adversary $\mathcal{A}$ against some primitive is *simulatable* if there exists a simulator Sim such that no efficient algorithm can distinguish between $\mathcal{A}$ and Sim from black-box query access. A cryptographic task having a simulatable attack cannot be black-box reduced to a secure cryptographic game since the reduction $\mathsf{R}^{(\cdot)}$ cannot distinguish between the inefficient $\mathcal{A}$ and the efficient Sim, which means that $\mathsf{R}^{\mathsf{Sim}}$ would yield an efficient algorithm for the game $\mathcal{G}$ with non-negligible advantage, contradicting the assumption.

4 A Simple Non-Interactive Primitive

In this paper, we consider a simple non-interactive cryptographic primitive, called a *weak one-time random oracle* ($\mathsf{WOTRO}^{n,m}$) and illustrated in Fig. 1 where the prover inputs $a \in \{0,1\}^n$ into the box and gets $c \in \{0,1\}^m$ as output while the verifier inputs nothing and gets (a, c) as output. An implementation of this primitive is a protocol taking place between the prover and the verifier. The verifier V is a machine that takes no input, interacts with the prover in the way prescribed by the protocol, and either accepts and outputs (a, c) or rejects and outputs $\perp$. In an honest implementation, the prover is a machine P taking as input an $a \in \{0,1\}^n$ and interacts with the verifier as specified by the protocol, in such a way that the verifier accepts and outputs the same c. The strings

a and c can then be determined from the transcript of the protocol. We can then view the whole protocol in the honest case as a conditional distribution $\Pi(c|a)$ that tells us the probability of getting the challenge c given that the prover was given a as input.

In a dishonest implementation, the prover $\tilde{\mathsf{P}}$ takes no input at all (it is free to choose a) and might behave in a way that will cause the verifier to reject. The protocol is then simply a joint probability distribution $\tilde{\Pi}_{\tilde{\mathsf{P}}}(a, c, v)$, representing the distribution one obtains when $\tilde{\mathsf{P}}$ runs the protocol with the honest verifier V, and where $v \in \{0, 1\}$ is 1 when the verifier accepts and 0 if he rejects.

We now define correctness and security of an implementation. In a correct implementation of this primitive, $\Pi(c|a)$ will reflect exactly the same distribution over a and c given by the ideal box, namely c will be uniformly distributed and independent of a, and the verifier always accepts when the prover is honest:

Definition 3 (ϵ-correctness) A protocol Π is a *ϵ-correct* implementation of $\mathsf{WOTRO}_{\Gamma}^{n,m}$ if for all $a \in \{0,1\}^n$ the conditional distribution $\Pi(c|a)$ is $(1-\epsilon)$-close (in statistical distance) to the uniform distribution over c and if V accepts with probability at least $1-\mathsf{negl}(n)$ when the prover is honest. Π is said to be *statistically correct* if it is $(1-\mathsf{negl}(n))$–correct. □

As for our security definition, it will be rather weak (hence the "weak" in the name of the primitive): we will only require that in a secure implementation, a dishonest prover $\tilde{\mathsf{P}}$ cannot steer the choice of c towards a deterministic function of a. Rather than require that c be almost uniform and independent, we will only demand that there be *some* randomness left in this choice.

Definition 4 (δ–avoiding) For $0 \leq \delta \leq 1$, we say that a tuple of random variables (A, C, V) taking values in $\{0,1\}^n \times \{0,1\}^m \times \{0,1\}$ *δ–avoids* the function $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$ if

$$\Pr\left[V = 1 \wedge C = \mathsf{c}(A)\right] \leq 1 - \delta \ .$$

□

This then leads to the following definition of security for an implementation of WOTRO.

Definition 5 (δ–security) A protocol is a statistically (resp. computationally) δ–secure implementation of $\mathsf{WOTRO}_{\Gamma}^{n,m}$ if for all dishonest provers (resp. all QPT dishonest provers) $\tilde{\mathsf{P}}$, the random variable tuple (A, C, V) with joint distribution $\tilde{\Pi}_{\tilde{\mathsf{P}}}(a, c, v)$ δ–avoids all functions $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$. We say that a protocol for WOTRO is *statistically (resp. computationally) secure* if it is statistically (resp. computationally) $(1 - \mathsf{negl}(n))$–secure. □

Basic Facts About WOTRO. Observe that there is a trivial perfectly secure 2–message protocol for WOTRO where P sends a and V sends a uniformly random c. Therefore, we will focus on non-interactive (or 1–message) implementations of WOTRO. A secure non-interactive WOTRO protocol provides enough conditional randomness for sound instantiation of the Fiat-Shamir transform when applied to public-coin special-sound 3–message interactive proofs (Σ–protocols).

In the plain model, there is no secure WOTRO protocol as the honest prover program defines the output c as a function of a that can never be avoided. In the CR$ model, there exists a simple statistically $(1 - \mathsf{negl}(m-n))$–secure one-message protocol when $m > n$, a statistically $\frac{1}{e}$–secure protocol when $m = n$ and there is no protocol for $m < n$ whose computational security can be black-box reduced to a cryptographic game assumption, as a consequence of [Bit+13]. A detailed examination of these facts is provided in Appendix C.

WOTRO in the CRQS Model. Since the object of study is the (im)possibility of the WOTRO primitive in the CRQS model, we present a general form for a 1–message WOTRO protocol in this model.

Definition 6 (WOTRO in the CRQS model) A $\mathsf{WOTRO}^{n,m}$ protocol $\Pi = (\mathsf{P}, \mathsf{V})$ in the CRQS model consists of

- A CRQS $\Psi_{PV} \in \mathcal{D}(\mathscr{H}_{PV})$
- A mapping of $a \in \{0,1\}^n$ to an efficient POVM $\mathcal{N}^a = \{N^a_{y,w}\}_{(y,w)\in\{0,1\}^{m\times\ell}}$ on register P.
- A mapping of $a \in \{0,1\}^n$, $y \in \{0,1\}^m$ and $w \in \{0,1\}^\ell$ to an efficient POVM $\mathcal{V}^{a,y,w} = \{V_0^{a,y,w}, V_1^{a,y,w}\}$ on register V.

On input $a \in \{0,1\}^n$:

1. P applies POVM $\mathcal{N}^a$ on register P of Ψ_{PV} to obtain y and an auxiliairy verification string w and sends (a, y, w) to the verifier.
2. V applies POVM $\mathcal{V}^{a,y,w}$ on register V of Ψ_{PV}, accepts and outputs (a, y) if the result is 1, and rejects and outputs nothing if the result is 0.

The conditional distribution of y given a is $\Pi(y|a) = \sum_w \mathrm{tr}(N^a_{y,w} \otimes V_1^{a,y,w}\Psi_{PV})$ for $y \neq \perp$ and $\Pi(\perp|a) = \sum_{y,w} \mathrm{tr}(N^a_{y,w} \otimes V_0^{a,y,w}\Psi_{PV})$. We say that Π is δ–*correct* if $\frac{1}{2^n}\sum_{a\in\{0,1\}^n}\sum_{y\neq\perp}\Pi(y|a) \geq \delta$. □

Note that requiring the auxiliary verification string w to be classical is not a restriction since the CRQS can contain EPR pairs for the teleportation of an arbitrary quantum state from the prover to the verifier.

4.1 WOTRO to Implement the Fiat-Shamir Heuristic

Let R_L be a relation for a language L and let $\Sigma_L = (\mathsf{P}_L, \mathsf{V}_L)$ be a Σ–protocol for R_L with commitments in Γ^n and challenges in Γ^m. Consider a secure implementation $\Pi_{\mathsf{WOTRO}} = (\mathsf{P}_{\mathsf{WOTRO}}, \mathsf{V}_{\mathsf{WOTRO}})$ of $\mathsf{WOTRO}^{n,m}_\Gamma$. We construct a non-interactive zero-knowledge proof (argument) system for L by applying the Fiat-Shamir transform to Σ_L using the protocol Π_{WOTRO} as the instantiation of the hash function.

Protocol $\Pi_{\mathsf{WOTRO}}[\Sigma_L]$
Setup: A Σ–protocol $\Sigma_L = (\mathsf{P}_L, \mathsf{V}_L)$ where $\mathsf{P}_L = (\mathsf{P}^1_L, \mathsf{P}^2_L)$ with commitments of size n and challenges of size m and a protocol $\Pi_{\mathsf{WOTRO}} = (\mathsf{P}_{\mathsf{WOTRO}}, \mathsf{V}_{\mathsf{WOTRO}})$ for $\mathsf{WOTRO}^{n,m}_\Gamma$.

Prover message: on public input $x \in L$ and private input w

1. compute $a \leftarrow \mathsf{P}^1_L(x, w)$,
2. compute $c \leftarrow \mathsf{P}_{\mathsf{WOTRO}}(a)$,
3. compute $z = \mathsf{P}^2_L(a, x, w, c)$ and
4. send z to the verifier.

Verification: on public input $x \in L$ and upon reception of z,

1. compute $(a, c) \leftarrow \mathsf{V}_{\mathsf{WOTRO}}()$
2. if $\mathsf{V}_{\mathsf{WOTRO}}$ rejected, output `reject` else output $\mathsf{V}_L(x, a, c, z)$.

Theorem 6 *If Σ_L is a Σ–protocol for language L and if Π_{WOTRO} is a statistically (resp. computationally) $(1-\delta)$–secure and correct implementation of WOTRO, then $\Pi_{\mathsf{WOTRO}}[\Sigma_L]$ is a statistically (resp. computationally) sound (with soundness error δ) and perfectly correct non-interactive proof system for language membership in L.*

Proof We first show correctness. By the correctness of Π_{WOTRO}, it holds that the challenge $c \in \Gamma^m$ produced by Π_{WOTRO} is uniformly distributed. When both parties are honest, the probability that V_L accepts when c is taken as the output of Π_{WOTRO} in protocol $\Pi_{\mathsf{WOTRO}}[\Sigma_L]$ is the same as the probability that V_L accepts in an execution of Σ_L. Since Σ–protocols are perfectly correct by definition, this probability is one.

Now for soundness, again by the definition of Σ–protocols, protocol Σ_L satisfies special soundness. That is, for $x \notin L$, for any commitment $a \in \Gamma^n$, there exist at most one challenge $c \in \Gamma^m$ that leads to an accepting conversation. Let $\mathsf{c} : \Gamma^n \to \Gamma^m$ be the function that maps commitment a to this unique challenge c that makes V_L accept. If Π_{WOTRO} is a statistically $(1-\delta)$–secure implementation of $\mathsf{WOTRO}^{n,m}$, then the output of Π_{WOTRO} $(1-\delta)$–avoids any function for any dishonest $\tilde{\mathsf{P}}_{\mathsf{WOTRO}}$. The probability that V for protocol $\Pi_{\mathsf{WOTRO}}[\Sigma_L]$ accepts when $x \notin L$ is equal to the probability that $\mathsf{V}_{\mathsf{WOTRO}}$ accepts output (A, C) *and* that V_L accepts on input (x, A, C, Z) for some Z. By special soundness, this probability is at most the probability that $\tilde{\mathsf{P}}_{\mathsf{WOTRO}}$ can make $\mathsf{V}_{\mathsf{WOTRO}}$ accept the output $(A, \mathsf{c}(A))$. By the statistical $(1-\delta)$–security of Π_{WOTRO}, this probability is at most δ.

The reasoning for computational soundness is the same, but where we instead restrict to QPT adversarial provers $\tilde{\mathsf{P}}_{\mathsf{WOTRO}}$ against Π_{WOTRO}. ■

4.2 WOTRO from Non-Local Correlations

A non-local box (PR box) is a hypothetical device distributed between two parties such that party A inputs $x \in \{0,1\}$ into the device and gets an output $u \in \{0,1\}$ and party B inputs $y \in \{0,1\}$ and gets $v \in \{0,1\}$. The input/output behaviour of the PR box is described by

$$\Pr[u, v \mid x, y] = \begin{cases} \frac{1}{2} & \text{if } u \oplus v = x \wedge y \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Let $\mathcal{C} : \{0,1\}^n \to \{0,1\}^N$ be an error correcting code with minimum distance ϵn (for any distinct $x, x' \in \{0,1\}^n$, the Hamming distance between $\mathcal{C}(x)$ and $\mathcal{C}(x')$ is at least ϵn). Let $\{h_r : \{0,1\}^N \to \{0,1\}^m\}_{r \in \mathcal{R}}$ be a universal$_2$ family of hash functions. The $\mathsf{WOTRO}^{n,m}$ protocol is as follows:

1. On CR\$ r, and using N PR boxes,
2. Prover: on input $a \in \{0,1\}^n$, compute codeword $x := \mathcal{C}(a)$ and input x into its interface of the N PR boxes. Let $u \in \{0,1\}^N$ be the result. Send (a, x, u) to the verifier and use $(a, h_r(u))$ as output.
3. Verifier: On reception of a, x, check that $x = \mathcal{C}(a)$. Pick $y \in \{0,1\}^N$ uniformly at random and input y into its interface of the N PR boxes. Let $v \in \{0,1\}^N$ be the result. Check that $u \oplus v = x \wedge y$. If any of the checks failed, output $\perp$, otherwise output $(a, h_r((x \wedge y) \oplus v))$.

Theorem 7 *The above protocol avoids every function* $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$.

PROOF We begin by describing the most general strategy for an adversary $\mathcal{A}$ against the protocol. $\mathcal{A}$ can input arbitrary values in the PR boxes in any order an such that input bits can depend on the CR\$ r and on the boxes' outputs to previous inputs. Let $\hat{x} \in \{0,1\}^N$ and $\hat{u} \in \{0,1\}^N$ denote the input and output bits to the N PR boxes, respectively. $\mathcal{A}$ is then free to choose a, x and u adaptively based on $\hat{x}$ and $\hat{u}$ and send (a, x, u) to the verifier. Since the verifier checks that $x = \mathcal{C}(a)$ and aborts otherwise, we can assume that x is indeed the codeword that corresponds to a.

We show that $\mathcal{A}$ has little freedom in the choice of a due to the error-correcting code and input/output behaviour of the PR boxes. Since $\mathcal{C}$ has minimal distance ϵn, there is at most one codeword x_0 such that $d(x_0, \hat{x}) \leq \frac{\epsilon}{2}n$. Let $a_0 = \mathcal{C}^{-1}(x_0)$. If $\mathcal{A}$ tries to send $(a, x = \mathcal{C}(a), u)$ for any $a \neq a_0$, then the verifier will abort with overwhelming probability as the following argument shows. Let (y, v) denote the input/output pair of the verifier. Then,

$$\begin{aligned}
&\Pr[x \wedge y = u \oplus v] \\
&= \Pr[x \wedge y = u \oplus (\hat{u} \oplus \hat{x} \wedge y)] \\
&= \Pr[x \wedge y \oplus u = \hat{x} \wedge y \oplus \hat{u}] \\
&= \prod_i \Pr[x_i \wedge y_i \oplus u_i = \hat{x}_i \wedge y_i \oplus \hat{u}_i] \ .
\end{aligned}$$

Now, consider the set of positions where $\hat{x}$ and x differ: $\mathcal{S} = \{i : \hat{x}_i \neq x_i\}$. For any $i \in \mathcal{S}$,

- When $y_i = 0$, the expression becomes $u_i = \hat{u}_i$.
- When $y_i = 1$, the expression becomes $x_i \oplus u_i = \hat{x}_i \oplus \hat{u}_i$ and it is satisfied when $u_i \neq \hat{u}_i$.

Since y is chosen independently and uniformly at random by the verifier, for every $i \in \mathcal{S}$, the expression $x_i \wedge y_i \oplus u_i = \hat{x}_i \wedge y_i \oplus \hat{u}_i$ has probability $\frac{1}{2}$ of not being satisfied. Therefore since $|\mathcal{S}| \geq \frac{\epsilon}{2}n$ whenever $x \neq x_0$, the verifier rejects with probability at least $2^{-\frac{\epsilon}{2}n}$.

Finally, since $\mathcal{A}$ is obligated to send a_0 and x_0 as described above and u that satisfies $u \oplus v = x_0 \wedge y$ as argued above, the output of the verifier satisfies

$$\begin{aligned}
\Pr[c = \mathsf{c}(a_0)] &= \Pr[h_r(u) = \mathsf{c}(a_0)] = \Pr[h_r(v \oplus x_0 \wedge y) = \mathsf{c}(a_0)] \\
&= \Pr[v \oplus x_0 \wedge y \in h_r^{-1}(\mathsf{c}(a_0))] = \frac{|h_r^{-1}(\mathsf{c}(a_0))|}{2^{-N}}
\end{aligned}$$

since $v \oplus x_0 \wedge y$ is uniformly distributed. On average over the choice of h_r, the above expression equals 2^{-m} because the universal$_2$ condition implies $\mathbb{E}_r |h_r^{-1}(z)| = 2^{N-m}$ for any $z \in \{0,1\}^m$. ■

It is well known that quantum mechanics can approximate the correlations of "noisy" PR boxes with success probability of around 85% whereas the best classical strategies can only achieve 75% success probability. We show in Section 5 that WOTRO is impossible in the CRQS model, i.e. in a model where PR boxes can be approximated with probability 85%. A natural question is: what is the level of noise at which WOTRO is no longer possible?

We must point out to the reader that our results do not imply that WOTRO is impossible using 85% PR boxes. There is a fundamental difference between (noisy) PR boxes

– which capture classical correlations – and entangled states: the latter can be measured coherently with a collapse happening on the other end. Our impossibility crucially relies on the adversary's ability to perform a coherent measurement on its register. In other words, if there is a secure WOTRO protocol using 85% PR boxes, then this protocol is no longer secure when the boxes are instantiated using EPR pairs. The question above thus remains open.

5 Impossibility of WOTRO in the CRQS Model

In this section, we prove our main result: there exists no protocol for WOTRO in the CRQS model with statistical security or with computational security established by black-box reduction to a cryptographic game, even a non falsifiable one. Our black-box impossibility result is proven using a similar technique as [Bit+13; BGW12]. In Section 5.1, we define an inefficient adversary that breaks completely any protocol implementing $\mathsf{WOTRO}^{n,m}$ with $n - m \in \omega(n)$ in the CRQS model. We call this adversary the Chernoff adversary or the Chernoff attack. In Section 5.3, we define what we mean by the security of WOTRO to be established by quantum black-box reduction to crypto game. We generalize to the quantum case this standard way of proving the security of cryptographic protocols. In Section 5.4, we show how to efficiently simulate the attack described in Section 5.1. We then conclude that the security of any protocol for $\mathsf{WOTRO}^{n,m}$ in the CRQS model cannot be established by a quantum black reduction to crypto games. As a consequence, feeding the reduction with the simulator of the Chernoff adversary instead of the Chernoff adversary will win the game while running efficiently. It follows that the game is trivial if such a reduction existed.

5.1 The Chernoff Attack Against Any Implementation of WOTRO

In this section, we show that there exist (inefficient) attacks against any 1–message $\mathsf{WOTRO}^{n,m}$ protocol in the CRQS model for m (sufficiently) smaller than n. The following definition describes a general strategy for an attack against WOTRO.

Definition 7 An attack $\mathcal{A}_n^f$ against a $\mathsf{WOTRO}^{n,m}$ protocol (Definition 6) is characterized by a *target function* $f : \{0,1\}^n \to \{0,1\}^m$ and a (possibly inefficient) POVM $\{P^f_{a,y,w}\}_{(a,y,w)\in\{0,1\}^{n\times m\times \ell}}$. The adversary performs this POVM on register P of CRQS Ψ_{PV} and sends the result (A, Y, W) to the verifier. We say that this attack *hits function* f except with probability $\epsilon(\mathcal{A}_n^f)$ if

$$1 - \epsilon(\mathcal{A}_n^f) = \Pr[Y = f(A) \wedge \ \mathsf{V} \text{ ACCEPTS}] = \sum_{a,w} \operatorname{tr}\left((P^f_{a,y,w} \otimes V_1^{a,f(a),w})\Psi_{PV}\right) \ .$$

□

We construct an attack whose success is based on the Chernoff bound for operators proven by Ahlswede and Winter in [AW02] and stated below. For operators A and B and $0 \le \eta \le 1$, the notation $A \in [(1-\eta)B; (1+\eta)B]$ means that $A \ge (1-\eta)B$ and $A \le (1+\eta)B$.

Lemma 1 (Operator Chernoff bound) *Let $X_1, \ldots, X_M$ be i.i.d. random variables taking values in the operators $\mathcal{D}(\mathcal{H})$ on the D–dimensional Hilbert space $\mathcal{H}$ such that $0 \le X_j \le \mathbb{1}$, with $A = \mathbb{E}[X_j] \ge \alpha\mathbb{1}$, and let $0 < \eta \le 1/2$. Then*

$$\Pr\left[\frac{1}{M}\sum_{j=1}^{M} X_j \notin [(1-\eta)A; (1+\eta)A]\right] \le 2D \exp\left(-M\frac{\alpha\eta^2}{2\ln 2}\right). \tag{2}$$

Our general attack strategy picks a random f and crafts a measurement on its part of the CRQS such that the measurement outcome (A, Y, W) satisfies:

1. $Y = f(A)$ and
2. $\mathsf{V}(A, Y, W)$ accepts with approximately the same probability as in an honest execution of the protocol.

Such a measurement is not efficiently implementable in general. We call this attack the "Chernoff adversary" based on the crucial use of the Chernoff bound in building this measurement.

Theorem 8 (Chernoff adversaries) *Let $n, m \in \mathbb{N}$ such that $m < n$. Let $\Pi^{n,m}_{\mathsf{WOTRO}}$ be a δ–correct $\mathsf{WOTRO}^{n,m}$ protocol described by CRQS Ψ_{PV} and POVM family $\mathcal{N}^a = \{N^a_{y,w}\}_{(y,w)\in\{0,1\}^{m\times\ell}}$ and $\mathcal{V}^{a,y,w} = \{V_0^{a,y,w}, V_1^{a,y,w}\}$ for $a \in \{0,1\}^n$, $y \in \{0,1\}^m$ and $w \in \{0,1\}^\ell$. Let $\mathcal{F} = \{f : \{0,1\}^n \to \{0,1\}^m\}$ be the set of boolean functions from n-bit strings to m-bit strings. Let $2^k = \dim P$ be the dimension of the prover's register of Ψ_{PV}.*

1. *Let $\eta = \sqrt{2\ln 2(n+k)\frac{2^m}{2^n}}$. Then, there exists $\mathcal{F}^* \subseteq \mathcal{F}$ such that $\Pr_{f\in_R\mathcal{F}}[f \in \mathcal{F}^*] \geq 1 - \mathsf{negl}(n)$ and for each $f \in \mathcal{F}^*$,*

$$\left\{ P^f_{a,w} := \frac{N^a_{f(a),w}}{(1+\eta)2^{n-m}} \right\}_{(a,w)\in\{0,1\}^{n\times\ell}} ,$$

   together with $P^f_\perp = \mathbb{1} - \sum_{a,w} P^f_{a,w}$, form a POVM on the prover's register P.

2. *Suppose $\{P^f_{a,w}\}_{a,w}$ can be completed as a POVM, then let $\mathcal{A}^f_n$ be the attack where the adversary applies POVM $\{P^f_{a,w}\}_{a,w} \cup \{P^f_\perp\}$ to its register, upon outcome (A, W) sets $Y = f(A)$ or $Y = \perp$ if outcome $\perp$ is observed and sends (A, Y, W) to the verifier. The probability of error $\epsilon(\mathcal{A}^f_n)$ of this attack satisfies $1 - \mathbb{E}_f[\epsilon(\mathcal{A}^f_n)] \geq \delta - \mathsf{negl}(n-m)$ where δ is the probability that V accepts in an honest execution.*

Proof Let 2^k be the dimension of register P. Consider the subset of measurement operators $N^a_{f(a),w}$ from the honest POVM $\mathcal{N}^a$ that yield the intended outcome for the cheating prover, i.e. on input a gives outcome $y = f(a)$, and define the operators $X^f_a = \sum_{w\in\{0,1\}^\ell} N^a_{f(a),w}$. We have that

$$\mathbb{E}_f[X^f_a] = \mathbb{E}_f \sum_{y,w\in\{0,1\}^{m\times\ell}} I_{\mathcal{E}^a_y}(f) \cdot N^a_{y,w} = \sum_{y,w\in\{0,1\}^{m\times\ell}} \mathbb{E}_f[I_{\mathcal{E}^a_y}(f)] \cdot N^a_{y,w} = \frac{\mathbb{1}_P}{2^m}$$

where $I_{\mathcal{E}^a_y}$ is the indicator function for the event $\mathcal{E}^a_y = \{f \mid y = f(a)\}$ which has probability $\frac{1}{2^m}$ for any y and a since every value for $f(a)$ is equally likely.

Applying the Chernoff bound with $D = 2^k$, $M = 2^m$, and $\alpha = \frac{1}{2^m}$ to the weighted sum over a of the operators X^f_a, we have

$$\Pr_f\left[\frac{1}{2^n}\sum_{a\in\{0,1\}^n} X^f_a \not\leq (1+\eta)\frac{\mathbb{1}}{2^m}\right] \leq 2^{k+1}\exp\left(-\frac{1}{2\ln 2}\cdot\frac{2^n}{2^m}\cdot\eta^2\right) .$$

This bound becomes negligible in n if we choose $\eta = \sqrt{2\ln 2(n+k)\frac{2^m}{2^n}} < \frac{1}{2}$. Therefore, except with probability $\mathsf{negl}(n)$,

$$\frac{1}{2^n}\sum_{a\in\{0,1\}^n} X^f_a = \frac{1}{2^n}\sum_{\substack{a\in\{0,1\}^n \\ w\in\{0,1\}^\ell}} N^a_{f(a),w} \leq (1+\eta)\frac{\mathbb{1}}{2^m} . \tag{3}$$

Define the ensemble of operators $P^f_{a,w}$ by

$$P^f_{a,w} := \frac{N^a_{f(a),w}}{(1+\eta)2^{n-m}} \ .$$

Then, when (3) holds, the set of operators $\{P^f_{a,w}\}_{a,w}$ forms a POVM when completed with $P^f_\perp = \mathbb{1} - \sum_{a,w} P^f_{a,w}$.

This gives rise to an attack $\mathcal{A}^f_n$ where the adversary applies the above POVM to obtain (a,w) and sets $y = f(a)$ to send (a,y,w) to the verifier. The probability of error $\epsilon(\mathcal{A}^f_n)$ for this attack corresponds to the probability of obtaining outcome "$\perp$" or of being rejected by the verifier. We have,

$$1 - \epsilon(\mathcal{A}^f_n) = \sum_{(a,w)\in\{0,1\}^{n\times\ell}} \mathrm{tr}\left(\left(P^f_{a,w} \otimes V_1^{a,f(a),w}\right)\Psi_{PV}\right) \ .$$

Recall that $\Pi(y|a) = \sum_w \mathrm{tr}(N^a_{y,w} \otimes V_1^{a,y,w}\Psi_{PV})$ is the probability that the verifier accepts the outcome $y \neq \perp$ for the prover input a in an honest execution and that by the δ–correctness of Π_{WOTRO}, $\frac{1}{2^n}\sum_a \sum_{y\neq\perp} \Pi(y|a) \geq \delta$. By (3), the above probability can, on average over f, be upper-bounded as follows:

$$\begin{aligned}
\mathbb{E}_f[1-\epsilon(\mathcal{A}^f)] &= \frac{1}{(1+\eta)2^{n-m}} \cdot \mathbb{E}_f \left[\sum_{(a,w)\in\{0,1\}^{n\times\ell}} \mathrm{tr}\left(\left(N_{a,f(a),w} \otimes V_1^{a,f(a),w}\right)\Psi_{PV}\right)\right] \\
&= \frac{1}{(1+\eta)2^{n-m}} \frac{1}{2^m} \sum_{a\in\{0,1\}^n} \sum_{(y,w)\in\{0,1\}^{m\times\ell}} \mathrm{tr}\left(\left(N_{a,y,w} \otimes V_1^{a,y,w}\right)\Psi_{PV}\right) \\
&= \frac{1}{(1+\eta)2^n} \sum_{a\in\{0,1\}^n} \sum_{y\neq\perp} \Pi(y|a) \\
&= \frac{\delta}{(1+\eta)} \\
&\geq (1-\eta)\delta \ ,
\end{aligned}$$

which is approximately δ since η is $\mathsf{negl}(n-m)$. ■

5.2 Oracle Access Quantum Circuits

Establishing the security of Π by *black-box reduction* to a cryptographic game $\mathcal{G} = (\Gamma, c)$ is defined by a (classical or quantum but efficient) machine $\mathsf{M}^{\mathcal{A}}$ such that $\mathsf{M}^{\mathcal{A}}(1^n)$ produces a quantum circuit (with oracle access) made out of some universal set of quantum gates together with *oracle access* to the standard interface of any adversary $\mathcal{A} = \{\mathcal{A}_n\}_n$ against Π such that if $\mathcal{A}_n$ breaks Π then $\mathsf{M}^{\mathcal{A}}(1^n)$ wins game $\mathcal{G}$. Let us first define this machine $\mathsf{M}^{\mathcal{A}}$ producing the circuit that will be called a *reduction* in the following.

Definition 8 (oracle access circuit) A *quantum oracle access machine* $\mathsf{M}^{(\cdot)}$ for oracle $\mathcal{O} = \{\mathcal{O}_n\}_n$ is a polynomial-time Turing machine that, on input 1^n, outputs the description of a quantum circuit over a universal set of quantum gates along with a special quantum gate: $\mathcal{O}_n : \mathrm{D}(P) \to \mathrm{D}(R)$ with standard interface P for the input and R for the output. The circuit produced by $\mathsf{M}^{\mathcal{O}}(1^n)$ is called an *oracle access quantum circuit*. The oracle calls behave as a CPTP map as the internal register E is not part of the interface. The i–th call to $\mathcal{O}_n$ is denoted $\mathcal{O}^i_n : \mathrm{D}(P_i) \to \mathrm{D}(R_i)$. We say $\mathsf{M}^{\mathcal{O}}$ makes $q(n)$ oracle queries if

we can represent the action of circuit $\mathcal{M}_n^{\mathcal{O}}$ produced by $\mathsf{M}^{\mathcal{O}}(1^n)$ on initial state $|0\rangle$ on all registers of the circuit as the CPTP map

$$\mathcal{M}_n^{\mathcal{O}_n}(|0\rangle\langle 0|) := U_n^q \circ \mathcal{O}_n^q \circ \cdots \circ U_n^1 \circ \mathcal{O}_n^1 \circ U_n^0(|0\rangle\langle 0|)$$

where $U_n^i \in L(R_i \otimes Q_i, P_{i+1} \otimes Q_{i+1})$ are unitaries made out of the universal set of gates representing the action of circuit $\mathcal{M}_n^{\mathcal{O}_n}$ between the calls to $\mathcal{O}_n$. Register Q_i is the reduction's working register before the action of U_n^i. □

As defined above, an oracle access circuit performs each oracle call using exactly the same functionality. No information can be kept by the oracle between calls. We then say that the oracle is *stateless*. In general, an oracle could be allowed to store quantum information between calls. This information is unavailable through the oracle input-output interface but is passed from one call to the next. These oracle access circuits are said to be *stateful*.

5.3 Quantum Black-Box Reductions

It remains to define what we mean exactly by polynomial-time *black-box reductions*. This notion was introduced by Impagliazzo and Rudich in [IR89] after observing that most proofs establishing the security of a crypto primitive constructed from one-way functions consider only the input-output behaviour of the function. In other words, the one-way function is only used as a black-box to construct the primitive and to prove its security. In [IR89], Impagliazzo-Rudich show that if it is possible to establish the security of a secret-key agreement based solely on the input-output behaviour of a one-way function then this security proof also establishes that $\mathbf{P} \neq \mathbf{NP}$. Reingold, Trevisan, and Vadhan in [RTV04] introduce three variants of black-box reductions called fully-BB, semi-BB, and mildly-BB from the stronger to the weaker flavour. In [BBF13], fully black-box reductions are described informally as follows:

> A fully black-box reduction R is an efficient algorithm that transforms any (even inefficient) adversary $\mathcal{A}$, breaking any instance Π^f of primitive $\mathcal{P}$, into an algorithm $\mathsf{R}^{\mathcal{A},f}$ breaking the instance f of $\mathcal{Q}$. Here, the reduction treats both the adversary as well as the primitive as black-boxes, and Π^f denotes the (black-box) construction out of f.

In our setting and as in [Bit+13; BGW12], we consider proofs establishing the security of protocol Π (for WOTRO) by providing an efficient oracle access quantum circuit $\mathsf{R}^{(\cdot)}$ with the property that for any $\mathcal{A}$ breaking Π, $\mathsf{R}^{\mathcal{A}}$ wins game $\mathcal{G}$ (to be more precise, $\mathsf{R}^{\mathcal{A},\Gamma}$ wins game $\mathcal{G} = (\Gamma, c)$). The adversary $\mathcal{A}$ against Π is therefore only used through its standard intput-output interface in reduction $\mathsf{R}^{\mathcal{A}}$ (i.e. as a quantum channel). This is what we call a quantum fully black-box reduction (or f-BB reduction), the quantum version of a *fully black-box reduction*, also called a *BBB reduction* in [BBF13]. Our definition agrees with other works in which fully black-box reductions are defined in the quantum setting [AG22; HY20].

Definition 9 (Quantum fully black-box reduction to a crypto game) Consider Π a protocol and let $\mathcal{A} = \{\mathcal{A}_n\}_n$ be an adversary provided through its standard interface. We say that *the security of* Π *is established by quantum fully black-box reduction to crypto game* $\mathcal{G} = (\Gamma, c)$ if there exists an efficient oracle access circuit $\mathsf{R}^{(\cdot)}$, called *the reduction*, such that when $\mathcal{A}$ breaks Π then

$$\Pr\left[\langle \mathsf{R}^{\mathcal{A}}(1^n) \rightleftharpoons \Gamma(1^n)\rangle = 1\right] \geq c + \frac{1}{\mathsf{poly}(n)} \ .$$

□

5.4 Efficient Simulation of the Chernoff Attack

We show that no reduction $\mathsf{R}^{\mathcal{A}}$ can establish the security of a WOTRO protocol by quantum black-box reduction to game assumption $\mathcal{G}$. The reason for this state of affair is that the Chernoff attacks described in Theorem 8 is *efficiently simulatable*. This means that there is an efficient algorithm Sim_n such that no oracle machine can tell whether it is given oracle access to the inefficient Chernoff adversary $\mathcal{A}_n^f$ hitting a random function $f(\cdot)$ or to Sim_n that does not know anything about $f(\cdot)$.

The most general attack against a WOTRO protocol in the CRQS model (Definition 7) is a POVM on the prover's part of the CRQS that produces a classical message which makes the verifier accept the output $c = f(a)$ with high probability. The attack takes no input other than the prover's register P of the CRQS and produces its output in registers $A \otimes Y \otimes W$. Let $\mathcal{F}^*$ be the set of functions defined in Theorem 8, i.e. such that for $f \in \mathcal{F}^*$, $\sum_{a,w} P_{a,w}^f \leq \mathbb{1}$ so that $P_\perp^f = \mathbb{1} - \sum_{a,w} P_{a,w}^f \geq 0$ and the set of operators $\{P_{a,w}^f\}_{(a,w)} \cup \{P_\perp^f\}$ forms a POVM. For $f \in \mathcal{F}^*$, the adversary $\mathcal{A}_n^f$ defined in Theorem 8 can be implemented by the following isometry $\mathcal{A}_n^f \in L(P, R \otimes E)$, where $R = A \otimes Y \otimes W \approx \mathscr{H}_n \otimes \mathscr{H}_{m(n)} \otimes \mathscr{H}_{\ell(n)}$ and where $E = E' \otimes P \approx \mathscr{H}_{p(n)} \otimes \mathscr{H}_n$, for $\ell(n), p(n)$ polynomials:

$$\mathcal{A}_n^f : |\psi\rangle_P \mapsto \sum_{\substack{a \in \{0,1\}^n \\ w \in \{0,1\}^{\ell(n)}}} |a, f(a), w\rangle_{AYW} \otimes |a, f(a), w\rangle_{E'} \otimes \sqrt{P_{a,w}^f}|\psi\rangle_{E''} + |\perp, \perp, \perp\rangle \otimes |\perp, \perp, \perp\rangle \otimes \sqrt{P_\perp^f}|\psi\rangle_{E''} \ , \quad (4)$$

Let $\mathcal{A}_n^{\mathcal{F}^*} = \{\mathcal{A}_n^f\}_{f \in \mathcal{F}^*}$ be the *family of all Chernoff adversaries* against protocol Π implementing $\mathsf{WOTRO}^{n,m}$. The standard output interface of any adversary $\mathcal{A}_n^f$ is made out of registers $A \otimes Y \otimes W$ while register $E = E' \otimes P$ is the working register of the adversary. It is easy to verify that any quantum black-box reduction $\mathsf{R}^{(\cdot)}$ establishing the security of Π by quantum black-box reduction to game $\mathcal{G}$ is such that $\mathsf{R}^{\mathcal{A}_n^f}$ wins $\mathcal{G}$ even when $f \in_R \mathcal{F}^*$. Next, we define what it means for the family of all Chernoff adversaries to be simulatable.

Definition 10 (Simulatable Attack for WOTRO) Let $n \in \mathbb{N}$, $m(n) \leq n$, Π a $\mathsf{WOTRO}^{n,m}$ protocol, and $\mathcal{A}_n^{\mathcal{F}^*}$ be the family of adversaries defined above. We say that $\mathcal{A}_n^{\mathcal{F}^*}$ is *efficiently $\epsilon(n)$–simulatable* if there exists a family of polynomial-time quantum algorithms $\mathsf{Sim} = \{\mathsf{Sim}_n\}_n$, called the simulator, such that

- The success probability of $\mathcal{A}_n^f$ is at least $1 - \mathsf{negl}(n - m)$ on average over $f \in_R \mathcal{F}^*$.
- For every (possibly inefficient) oracle acess machine $\mathsf{M}^{(\cdot)}$ making $q(n) = \mathsf{poly}(n)$ queries to its oracle, the CPTP map $\mathcal{M}_n^{(\cdot)}$ describing the action of circuit $\mathsf{M}^{(\cdot)}(1^n)$ satisfies
$$\|\mathbb{E}_{f \in \mathcal{F}^*}[\mathcal{M}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)] - \mathcal{M}_n^{\mathsf{Sim}_n}(|0\rangle\langle 0|_R)\|_1 \leq \epsilon(n) \ . \quad (5)$$

□

Next theorem shows that the family adversaries $\mathcal{A}_n^{\mathcal{F}}$ is efficiently simulatable. Unlike the simulator used in [Bit+13] for their family of inefficient adversaries, our simulator is **not** stateful. The full proof is in Appendix A.

Theorem 9 *Let n, m and $\Pi^{n,m}_{\mathsf{WOTRO}}$ be as in the statement of Theorem 8. The family of adversaries $\mathcal{A}_n^{\mathcal{F}^*}$ is efficiently $\mathsf{negl}(n-m)$–simulatable.*

Theorem 8 and Theorem 9 give a simulatable attack against any $\mathsf{WOTRO}^{n,m}$ protocol where $n-m$ is linear in the security parameter n. We conclude,

Corollary 1 *Let $\mathcal{G}$ be a cryptographic game assumption and let $\Pi^{n,m}$ be a $\mathsf{WOTRO}^{n,m}$ protocol with $n-m \in \omega(\lg n)$. For $\delta \geq 1/\mathsf{poly}(n)$, if there is a quantum black-box reduction showing that $\Pi^{n,m}$ δ–avoids all functions assuming game $\mathcal{G}$ then assumption $\mathcal{G}$ is false.*

5.4.1 Proof of Theorem 9

Consider a protocol Π for $\mathsf{WOTRO}^{n,m}$ using a CRQS $|\Psi\rangle_{PV}$ with $\dim(P) = 2^k$. Let $\mathcal{F}$ be the set of all functions $f : \{0,1\}^n \to \{0,1\}^m$. We consider the following isometric implementation of the Chernoff adversary $\mathcal{A}^f$ described in Section 5.4 using internal working quantum register $E = E' \otimes E''$,

$$\mathcal{A}_n^f : |\psi\rangle_P \mapsto \sum_{\substack{a\in\{0,1\}^n \\ w\in\{0,1\}^{\ell(n)}}} |a, f(a), w\rangle_{AYW} \otimes |a, f(a), w\rangle_{E'} \otimes \sqrt{P^f_{a,w}}|\psi\rangle_{E''} + |\perp,\perp,\perp\rangle \otimes |\perp,\perp,\perp\rangle \otimes \sqrt{P^f_\perp}|\psi\rangle_{E''} \ , \quad (6)$$

with input register P and output register $A \otimes Y \otimes W$. Remember however that Theorem 8 does not guarantee that for all $f \in \mathcal{F}$, $\mathcal{A}_n^f$ is a POVM (in which case (6) is not an isometry). It only tells us that there exists $\mathcal{F}^* \subseteq \mathcal{F}$ such that $\forall f \in \mathcal{F}^*$, $\mathcal{A}_n^f$ implements a valid POVM (and therefore, (6) is indeed an isometry) and $\Pr[f \in \mathcal{F}^*] \geq 1 - \mathsf{negl}(n)$. When $f \notin \mathcal{F}^*$, the implementation of $\mathcal{A}_n^f$ defined in (6) is not an isometry but is still a linear map, though not a physically realizable one. Whenever $f \notin \mathcal{F}^*$, we set $P^f_\perp = 0$ such that it is always positive semidefinite for all $f \in \mathcal{F}$. When $f \in \mathcal{F}^*$, we have $\sum_{a,w} P^f_{a,w} + P^f_\perp = \mathbb{1}_{E''}$ as $\{P^f_{a,w}\}_{a,w} \cup \{P^f_\perp\}$ is a valid POVM. Otherwise, when $f \notin \mathcal{F}^*$, $\sum_{a,w} P^f_{a,w} + P^f_\perp \geq \mathbb{1}_{E''}$.

Now, consider the simulator $\mathsf{Sim} = \{\mathsf{Sim}_n\}_n$ where Sim_n is defined as follows:

Sim_n**:**

1. Pick $a \in_R \{0,1\}^n$,
2. Apply the honest POVM N^a to register P to get outcome (c, w),
3. Output (a, c, w).

This simulator corresponds to the isometry

$$\mathsf{Sim}_n : |\psi\rangle_P \mapsto 2^{-n/2} \sum_{\substack{a\in\{0,1\}^n \\ y\in\{0,1\}^m \\ w\in\{0,1\}^{\ell(n)}}} |a, y, w\rangle_{ACW} \otimes |a, y, w\rangle_{E'} \otimes \sqrt{N^a_{y,w}}|\psi\rangle_{E''} \ , \quad (7)$$

with the same input-output interface than any adversary against Π. The above simulator is efficiently implementable since it only purifies the honest prover's measurement. It is not too difficult to show that if POVM $\mathcal{N}^a = \{N^a_{y,w}\}_{a,y,w}$ can be implemented efficiently for all $a \in \{0,1\}^n$ then the isometry (7) is efficient. Notice that the simulator never produces an error as we assume that the honest strategy in Π never produces an error. We could

have allowed a protocol for $\mathsf{WOTRO}^{n,m}$ to produce an error with negligible probability in n. This would not cause any problem with what we establish in the following.

Before going further, we use the operator Chernoff bound of lemma 1 to establish a few useful properties of the Chernoff adversaries. The following is a direct consequence of the operator bound.

Lemma 2 *Let Π be a protocol for $\mathsf{WOTRO}^{n,m}$ with POVM $\mathcal{N}^a = \{N^a_{y,w}\}_{y,w}$ for prover $\mathsf{P}_{\mathsf{WOTRO}}$. Consider the Chernoff adversaries $\{\mathcal{A}^f_n\}_{f\in\mathcal{F}}$ against Π as defined in Theorem 8 with $\eta = \sqrt{2\ln 2(n+k)\cdot 2^{m-n}}$. Let $P^f := \{P^f_{a,w}\}_{a,w}$ be the POVM applied by $\mathcal{A}^f_n$ where $P^f_{a,w} = \frac{N^a_{f(a),w}}{2^{n-m}(1+\eta)}$. Then, for any $\frac{1}{2\eta} \geq t > 1$,*

$$\Pr_{f\in\mathcal{F}}\left[\sum_{a,w} P^f_{a,w} \notin \left[\frac{(1-t\eta)}{1+\eta}\mathbb{1}_P, \frac{(1+t\eta)}{1+\eta}\mathbb{1}_P\right]\right] \leq 2^{-n\cdot t^2} \quad .$$

The proof is rather direct and can be found in Appendix A.

Suppose for a contradiction that $\mathsf{R}^{\mathcal{A}}$ is a reduction that, for any successful adversary $\mathcal{A}$ against protocol Π, produces a circuit that wins game $\mathcal{G} = (\Gamma, c)$ using $q(n) = \mathsf{poly}(n)$ queries to $\mathcal{A}$. We show that oracle access circuits $\mathsf{R}^{\mathcal{A}^f}(1^n)$ and $\mathsf{R}^{\mathsf{Sim}}(1^n)$ produce states at negligible trace-norm distance when both are evaluated on $|0\rangle\langle 0|$ and when $\mathcal{A}^f$ is picked with $f \in_R \mathcal{F}^*$. Let $\mathcal{R}^{\mathcal{A}_n}_n := \mathsf{R}^{\mathcal{A}}(1^n)$ be the oracle-access circuit produced by the reduction with security parameter n upon oracle $\mathcal{A}_n$.

The first thing to observe is that picking $f \in_R \mathcal{F}^*$ in reduction $\mathcal{R}^{\mathcal{A}^f_n}_n$ is essentially the same as running the reduction with $f \in_R \mathcal{F}$, even though in this case $\mathcal{R}^{\mathcal{A}^f_n}_n$ may not be physically realizable.

Lemma 3 *Let $\mathcal{F}$ and $\mathcal{F}^*$ be defined as above for $n, m \in \mathbb{N}$. For $f \in \mathcal{F}$, consider adversary $\mathcal{A}^f_n$ defined in (6). Then,*

$$\left\| \mathop{\mathbb{E}}_{f\in\mathcal{F}^*}\left[\mathcal{R}^{\mathcal{A}^f_n}_n(|0\rangle\langle 0|)\right] - \mathop{\mathbb{E}}_{f\in\mathcal{F}}\left[\mathcal{R}^{\mathcal{A}^f_n}_n(|0\rangle\langle 0|)\right]\right\|_1 \leq \mathsf{negl}(n) \quad .$$

The proof of this lemma can also be found in Appendix A.

As a direct consequence of lemma 3, we get

$$\begin{aligned}\left\|\mathcal{R}^{\mathsf{Sim}}_n(|0\rangle\langle 0|) - \mathop{\mathbb{E}}_{f\in\mathcal{F}^*}\left[\mathcal{R}^{\mathcal{A}^f}_n(|0\rangle\langle 0|)\right]\right\|_1 \leq \\ \mathsf{negl}(n) + \left\|\mathcal{R}^{\mathsf{Sim}}_n(|0\rangle\langle 0|) - \mathop{\mathbb{E}}_{f\in\mathcal{F}}\left[\mathcal{R}^{\mathcal{A}^f}_n(|0\rangle\langle 0|)\right]\right\|_1 \quad .\end{aligned} \tag{8}$$

To bound the trace-distance between $\mathcal{R}^{\mathcal{A}^f}_n(|0\rangle\langle 0|)$ for $f \in_R \mathcal{F}$ and $\mathcal{R}^{\mathsf{Sim}_n}_n(|0\rangle\langle 0|)$ when $\mathcal{R}^{(\cdot)}_n$ is an oracle access circuit with $q := q(n) \in \mathsf{poly}(n)$ queries, we use $q+1$ hybrid reductions where hybrid i acts as Sim_n on the first i queries and acts as $\mathcal{A}^f_n$ on the remaining $q-i$ queries. In the following, we denote by $\mathcal{R}^{\mathsf{Sim},\mathcal{A}^f}_{n,j}$ the oracle-access circuit $\mathsf{R}^{\mathcal{A}}(1^n)$ where the first j calls are made to oracle Sim_n and the last $q-j$ calls are made to $\mathcal{A}^f_n$. We therefore have that $\mathcal{R}^{\mathsf{Sim},\mathcal{A}^f}_{n,q}$ corresponds to $\mathcal{R}^{\mathsf{Sim}}_n$ and $\mathcal{R}^{\mathsf{Sim},\mathcal{A}^f}_{n,0}$ corresponds to $\mathcal{R}^{\mathcal{A}^f}_n$, and

$$\begin{aligned}\left\|\mathcal{R}^{\mathsf{Sim}}_n(|0\rangle\langle 0|_R) - \mathop{\mathbb{E}}_{f\in\mathcal{F}}\left[\mathcal{R}^{\mathcal{A}^f}_n(|0\rangle\langle 0|)\right]\right\|_1 = \\ \left\|\mathop{\mathbb{E}}_{f\in\mathcal{F}}\left[\mathcal{R}^{\mathsf{Sim},\mathcal{A}^f}_{n,q}(|0\rangle\langle 0|) - \mathcal{R}^{\mathsf{Sim},\mathcal{A}^f}_{n,0}(|0\rangle\langle 0|)\right]\right\|_1 \quad .\end{aligned} \tag{9}$$

By a standard hybrid argument, the right-hand side of (9) is upper bounded as follows:

$$\left\| \mathop{\mathbb{E}}_{f\in\mathcal{F}} \left[\mathcal{R}_{n,q}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) - \ \mathcal{R}_{n,0}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) \right] \right\|_1 \leq \sum_{j=1}^{q} \left\| \mathop{\mathbb{E}}_{f\in\mathcal{F}} \left[\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) \right] \right\|_1 . \tag{10}$$

We now upper bound $\left\| \mathop{\mathbb{E}}_{f\in\mathcal{F}} \left[\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) \right] \right\|_1$ for $j \in \{1,\ldots,q\}$. Notice that the working registers (register E' in (6)) of any oracle call can be measured without modifying the behaviour of the reduction as these registers are not under its control, these measurements all commute with the operations of the reduction. For any $j \in \{1,\ldots,n\}$, circuits $\mathcal{R}_{n,j}^{\mathcal{A}^f,\mathsf{Sim}}$ and $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}$ differ only in the j-th query, which is made to Sim_n in $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}$ and to $\mathcal{A}_n^f$ in $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}$. Otherwise, both $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}$ and $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}$ query Sim_n for all queries prior to the j-th and both query $\mathcal{A}_n^f$ for all queries following the j-th. Let $\mathcal{S} = \{0,1\}^n \times \{0,1\}^m \times \{0,1\}^{\ell(n)}$ be the set of possible announcements (a,y,w) for a prover in Π except when an error occured (i.e. when $a = \perp$ is obtained). For $j \geq 1$, let $S^{j-1} = (S_1^{j-1},\ldots,S_{j-1}^{j-1}) \in \mathcal{S}^{j-1}$ be the random variable for the outcomes of the $j-1$ first queries to Sim_n in $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$ and $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$, where S_h^{j-1}, for $h \in \{1,\ldots,j-1\}$, represents the *result* of the h–th call. Remember that the portion of the adversary's circuit up to but not including the j–th call is an isometry as it is independent of $f \in \mathcal{F}$. This independence of all $j-1$ first outcomes is important in applying the hybrid argument. Only querying $\mathcal{A}_n^f$ can produce the special error outcome $(\perp,\perp,\perp)$ and only querying $\mathcal{A}_n^f$ with $f \notin \mathcal{F}^*$ for the j–th query can transform the state of the reduction before the j–th query into a non-physical one, as its trace-norm could exceed 1. Remember that outcome $S_h^{j-1} = (a,y,w)$ corresponds to the outcome when *measuring* in the computational basis the internal register E' of the h–th call to $\mathcal{A}_n^f$. We say that S^{j-1} *is confused about* a if $S_h^{j-1} = (a,y,w)$ and $S_{h'}^{j-1} = (a,y',w')$ for some $h \neq h'$ and $y \neq y'$. For $s \in \mathcal{S}^{j-1}$, we denote by $Q_{S^{j-1}}(s)$ the probability of results s for the $j-1$ first calls (to Sim_n) in $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$. By construction, this also corresponds to the probability of s for the $j-1$ first calls in $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$. For $s \in \mathcal{S}$, we let $|\psi_j(s)\rangle$ be the state obtained just prior the j-th query in both $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$ and $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$ given that registers $E_1'',\ldots,E_{j-1}''$ have each been measured in the computational basis to get s. In the following, we abuse the notation and write $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|\psi_j(s)\rangle\langle\psi_j(s)|)$ and $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|\psi_j(s)\rangle\langle\psi_j(s)|)$ to denote the result of the each hybrid reductions when $|\psi_j(s)\rangle$ is used for the j–th query onward. $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|\psi_j(s)\rangle\langle\psi_j(s)|)$ will make all its remaining queries to $\mathcal{A}_n^f$ while $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|\psi_j(s)\rangle\langle\psi_j(s)|)$ will query Sim_n one last time before querying $\mathcal{A}_n^f$.

Let F be a random variable uniformly distributed in $\mathcal{F}$. For $s \in \mathcal{S}^{j-1}$, let $\mathbb{S}_s$ be the projector on the subspace producing outcomes s when registers $E_1'',\ldots,E_{j-1}''$ of the $(j-1)$–th first calls to Sim_n are each measured in the computational basis. By construction of the simulator, $\{\mathbb{S}_s\}_{s\in\mathcal{S}^{j-1}}$ defines a complete Von Neumann measurement of register $\bigotimes_{i=1}^{j-1} E_i''$ provided by the $j-1$ first calls to Sim_n. We have,

$$\begin{aligned} &\left\| \mathop{\mathbb{E}}_{f} \left[\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) \right] \right\|_1 \\ &= \left\| \sum_{f\in\mathcal{F}} \Pr[F=f] \left(\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|) \right) \right\|_1 \end{aligned}$$

$$
\begin{aligned}
&= \left\| \sum_{\substack{s\in\mathcal{S}^{j-1}\\ f\in\mathcal{F}}} \Pr[F=f] \left(\mathbb{S}_s \mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)\mathbb{S}_s - \mathbb{S}_s \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)\mathbb{S}_s \right) \right\|_1 \\
&= \left\| \sum_{\substack{s\in\mathcal{S}^{j-1}\\ f\in\mathcal{F}}} \Pr[F=f] Q_{S^{j-1}}(s) \left(\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s)) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s)) \right) \right\|_1 \\
&\leq \underbrace{\sum_{\substack{s\in\mathcal{S}^{j-1}\\ f\in\mathcal{F}}} \Pr[F=f] Q_{S^{j-1}}(s) \left\| \mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s)) - \mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s)) \right\|_1}_{\text{(D)}} . \qquad (11)
\end{aligned}
$$

We now find a negligible upper bound for (D) in (11). This is where the work to apply the hybrid argument to the j–th query is done. Given $s \in \mathcal{S}^{j-1}$, remember that $|\psi_j(s)\rangle$ is the state obtained (a pure state) just prior to j–th query in both $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$ and $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(|0\rangle\langle 0|)$. Remember also that we denote by $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s))$ and $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}(\psi_j(s))$ the action of the circuit from the j–th query onward with initial state $|\psi_j(s)\rangle$. The j–th query is made to Sim_n in $\mathcal{R}_{n,j}^{\mathsf{Sim},\mathcal{A}^f}$ and to $\mathcal{A}_n^f$ in $\mathcal{R}_{n,j-1}^{\mathsf{Sim},\mathcal{A}^f}$.

Let $q_{a,w}^{\mathcal{A}_n^f}(\psi_j(s)) := \mathrm{tr}((P_{a,w}^f \otimes \mathbb{1}_{Z_j})|\psi_j(s)\rangle\langle\psi_j(s)|_{P_j Z_j})$ be the *likelihood* of outcome $(a, f(a), w)$ for the j–th query made to $\mathcal{A}_n^f$ upon $|\psi_j(s)\rangle$. Let $q_{\perp}^{\mathcal{A}_n^f}(\psi_j(s)) = \mathrm{tr}(P_{\perp}^f|\psi_j(s)\rangle\langle\psi_j(s)|)$ be the likelihood that the j–th query to $\mathcal{A}_n^f$ produces an error and let $|\psi_j^{f,\perp}(s)\rangle$ the normalized vector obtained after the j–th query has produced an error (notice that if the j–th query is made to Sim_n then no error can be produced). The set $\{q_{a,w}^{\mathcal{A}_n^f}(\psi_j^f(s))\}_{a,w} \cup \{q_{\perp}^{\mathcal{A}_n^f}(\psi_j^f(s))\}$ is not guaranteed to be a probability distribution when $f \notin \mathcal{F}^*$ (this is the reason why we call these values *likelihoods* instead of probabilities).

Likewise, $q_{a,y,w}^{\mathsf{Sim}_n}(\psi_j(s)) := 2^{-n}\,\mathrm{tr}((N_{y,w}^a \otimes \mathbb{1}_{Z_j})|\psi_j(s)\rangle\langle\psi_j(s)|)$ be the probability that Sim_n picks a uniformly at random and observes (y, w) when applying Π's honest measurement $\mathcal{N}^a$ on vector $|\psi_j(s)\rangle$. Notice that by definition of $\{P_{a,w}^f\}_{a,w}$, for all $(a, f(a), w) \in \mathcal{S}$, states

$$
\left(\sqrt{P_{a,w}^f} \otimes \mathbb{1}_{Z_j}\right)|\psi_j(s)\rangle \text{ and } \left(\sqrt{N_{f(a),w}^a} \otimes \mathbb{1}_{Z_j}\right)|\psi_j(s)\rangle
$$

are identical once normalized. Let $|\psi_j^{a,f(a),w}(s)\rangle$ be that state. In the following, we write $(a, z) \in s$ if there exists $w \in \{0,1\}^{\ell(n)}$ such that $(a, z, w) \in s$. We also write $a \in s$ if there exist z, w such that $(a, z, w) \in s$. Let $\delta(s) := \{a \in \{0,1\}^n \mid a \in s\}$. The sum over f in (D) can now be written as

$$
\begin{aligned}
&\frac{1}{\#\mathcal{F}} \sum_{f\in\mathcal{F}} \mathrm{tr}_{E_j} \Big((\mathsf{Sim}_n \otimes \mathbb{1}_{Z_j})|\psi_j(s)\rangle\langle\psi_j(s)|(\mathsf{Sim}_n \otimes \mathbb{1}_{Z_j})^* \\
&\qquad\qquad - (\mathcal{A}_n^f \otimes \mathbb{1}_{Z_j})|\psi_j(s)\rangle\langle\psi_j(s)|(\mathcal{A}_n^f \otimes \mathbb{1}_{Z_j})^* \Big) \\
&= \sum_{\substack{a\in\{0,1\}^n\setminus\delta(s)\\ y\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} \underbrace{\frac{2^m}{\#\mathcal{F}} \sum_{\substack{f\in\mathcal{F}\\ f(a)=y}} \left| q_{a,y,w}^{\mathsf{Sim}_n}(\psi_j(s)) - 2^{-m} q_{a,w}^{\mathcal{A}_n^f}(\psi_j(s)) \right|}_{\text{(M)}} \qquad (12) \\
&\qquad\qquad |a,y,w\rangle\langle a,y,w| \otimes \left|\psi_j^{a,f(a),w}(s)\right\rangle\!\left\langle\psi_j^{a,f(a),w}(s)\right|
\end{aligned}
$$

$$+\underbrace{\frac{1}{\#\mathcal{F}}\sum_{\substack{f\in\mathcal{F}\\ a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}}\left|q^{\mathsf{Sim}_n}_{a,z,w}(\psi_j(s))-q^{\mathcal{A}^f_n}_{a,w}(\psi_j(s))\right||a,z,w\rangle\langle a,z,w|\otimes\left|\psi_j^{a,f(a),w}(s)\right\rangle\left\langle\psi_j^{a,f(a),w}(s)\right|}_{(\mathrm{A})} \tag{13}$$

$$+\underbrace{\frac{1}{\#\mathcal{F}}\sum_{f\in\mathcal{F}}q^{\mathcal{A}^f_n}_{\perp}(\psi_j(s))\,|\perp,\perp,\perp\rangle\langle\perp,\perp,\perp|\otimes\left|\psi_j^{f,\perp}(s)\right\rangle\left\langle\psi_j^{f,\perp}(s)\right|}_{(\perp)}\ . \tag{14}$$

In the equation above, (M) is the main difference between $\mathcal{A}^f_n$ and Sim_n, (A) represents the outcomes in which s is confused about a and ($\perp$) represents the adversary's inconclusive outcome. We refer to Appendix A for the proofs that (A), ($\perp$), and the main part (M) are all negligible in $n-m$.

Putting things together using the bounds on (38),(39), and (40), we conclude that

$$(D)\leq\mathsf{negl}(n-m)\ , \tag{15}$$

and this negligible upper bound on (D) is independent of $s\in\mathcal{S}^{j-1}$ and $1\leq j\leq q(n)$. We conclude from (11) that for all $1\leq j\leq q(n)$,

$$\left\|\mathop{\mathbb{E}}_f\left[\mathcal{R}^{\mathcal{A}^f,\mathsf{Sim}}_{n,j}(|0\rangle\langle 0|)-\mathcal{R}^{\mathcal{A}^f,\mathsf{Sim}}_{n,j-1}(|0\rangle\langle 0|)\right]\right\|_1\leq\mathsf{negl}(n-m)\ . \tag{16}$$

Finally, plugging (16) into (10) completes the proof of Theorem 9.

6 Black-Box Impossibility of Fiat-Shamir in the CRQS Model

We assume the reader is familiar with Σ–protocols and the Fiat-Shamir transform. For more information, we refer to Section 3.1.

In this section, we consider the natural extension of the Fiat-Shamir transform in the CRQS model where the prover and verifier share an arbitrary entangled state $|\varphi_{n,m}\rangle$, the prover performs some measurement specified by a on its part of the CRQS, sends the result to the verifier who performs its own measurement based on the prover's message. Since a universal instantiation of the Fiat-Shamir is required to transform any Σ–protocol into a sound argument, the CRQS $|\varphi_{n,m}\rangle$, as well as the measurement operators of the prover and verifier must be independent of the actual Σ–protocol and of the statement x. The quantum Fiat-Shamir transform proceeds as follows:

1. P^{FS} computes $a=\mathsf{P}(x,w)$ and performs some measurement $\mathcal{N}^a$ on its part of $|\varphi_{n,m}\rangle$ that yield classical outcomes (c,v). It computes $z=\mathsf{P}_2(a,x,w,c)$, and sends (a,c,v,z) to V^{FS}.

2. V^{FS} performs a binary-outcome measurement $\mathcal{V}^{a,c,v}$ on its part of $|\varphi_{n,m}\rangle$ and rejects if the outcome is 0, and otherwise outputs $\mathsf{V}(x,a,c,z)$.

We consider without loss of generality that all communication remains classical, since the CRQS could contain polynomially many EPR pairs allowing for the teleportation of quantum states from the prover to the verifier.

An *abstract* Fiat-Shamir transform that captures all of the above would look like the following. Since we are proving a negative result, we only ask that a universal instantiation of the Fiat-Shamir transform has constant soundness error (instead of $\mathsf{negl}(n)$).

Definition 11 The Fiat-Shamir transform is given by $\Pi_{\mathsf{FS}}^{n,m} = (\mathsf{P}_{\mathsf{FS}}, \mathsf{V}_{\mathsf{FS}})$ where P_{FS} takes as input the commitment $a \in \{0,1\}^n$ and outputs a challenge $c \in \{0,1\}^m$ and a auxiliary verification information v. V_{FS} takes input (a, c, v) and outputs `accept` or `reject`. For a Σ–protocol $\Sigma = (\mathsf{P}_\Sigma, \mathsf{V}_\Sigma)$, the Fiat-Shamir transform applied to Σ is the non-interactive protocol $\Pi_{\mathsf{FS}}^{n,m}[\Sigma] = (\mathsf{P}, \mathsf{V})$ defined as

1. P computes $a = \mathsf{P}_\Sigma^1(x, w)$ and runs $(c, v) \leftarrow \mathsf{P}_{\mathsf{FS}}(a)$. It computes $z = \mathsf{P}_\Sigma^2(a, x, w, c)$, and sends (a, c, v, z) to V.

2. V runs $\mathsf{V}_{\mathsf{FS}}(a, c, v)$ and rejects if V_{FS} rejects, and otherwise outputs $\mathsf{V}_\Sigma(x, a, c, z)$.

The Fiat-Shamir transform $\Pi_{\mathsf{FS}}^{n,m}$ is (n, m)–universal if for any Σ–protocol Σ, $\Pi_{\mathsf{FS}}^{n,m}[\Sigma]$ is an argument with soundness error bounded above by some constant greater than zero. □

Note that an instantiation of the Fiat-Shamir transform is also one for WOTRO (and vice-versa). More precisely, the WOTRO protocol implied by Fiat-Shamir is the protocol where $\mathsf{P}_{\mathsf{WOTRO}}$ invokes P_{FS}, sends (a, c, v) to $\mathsf{V}_{\mathsf{WOTRO}}$ that outputs (a, c) if $\mathsf{V}_{\mathsf{FS}}(a, c, v)$ accepts. The main distinction between the two is that a secure protocol for WOTRO needs to avoid all functions, whereas a universal instantiation of Fiat-Shamir only needs to avoid functions that are "bad challenges" functions for some Σ–protocol for language membership to L upon some public parameter $x \notin L$.

6.1 Black-Box Impossibility of Universal Fiat-Shamir

We begin by defining what is a black-box reduction from FS to a cryptographic game assumption similarly to how it is done in [Bit+13].

Definition 12 (Black-Box Reduction for Quantum Fiat-Shamir) Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game assumption and let $\Pi_{\mathsf{FS}}^{n,m}$ be an instantiation of the Fiat-Shamir transform in the CRQS model. A *black-box reduction showing the* (n, m)*–QFS–universality of* $\Pi_{\mathsf{FS}}^{n,m}$ *under the assumption* $\mathcal{G}$ *in the CRQS model* is an oracle-access machine $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ such that the following holds. Let

1. $\Sigma = (\mathsf{P}, \mathsf{V})$ be a Σ–protocol for a language L with commitment length n and challenge length m that has perfect completeness and special soundness, and

2. $\mathcal{A}$ be a (possibly inefficient) attacker that breaks the computational soundness of the non-interactive proof system $\Pi_{\mathsf{FS}}^{n,m}[\Sigma]$ with advantage $1 - \mathsf{negl}(n)$.

The reduction $\mathcal{B}$ has black-box access to P, V and $\mathcal{A}$, runs in time polynomial in the running times of P, V and $\mathcal{A}$, and $\mathcal{B}^{\mathsf{P},\mathsf{V},\mathcal{A}}$ has advantage at least $1/\mathsf{poly}(n)$ in game $\mathcal{G}$. □

As mentioned previously, a FS protocol is essentially a WOTRO protocol, albeit satisfying a weaker notion of security. In particular, a WOTRO protocol avoiding *only* the "bad challenge" functions of Σ–protocols would be enough for FS. The impossibility to black-box reduce the security of WOTRO to a cryptographic game, as expressed in Corollary 1, does not apply directly to Fiat-Shamir.

To show black-box impossibility of FS in the CRQS model, we construct a family of Σ–protocols $\{\Sigma^f\}_{f:\{0,1\}^n \to \{0,1\}^m}$ such that Σ^f has bad challenge function $f(\cdot)$ for any f. The verifier V^f in Σ^f is not necessarily efficient, but we again exploit the simulation paradigm, where the inefficient adversary is replaced by an efficient indistinguishable simulator, to simulate this verifier in a way that is consistent with the adversarial prover. By definition of the reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$, if an adversary $\mathcal{A}^f$ breaks the soundness of $\Pi_{\mathsf{FS}}[\Sigma^f]$,

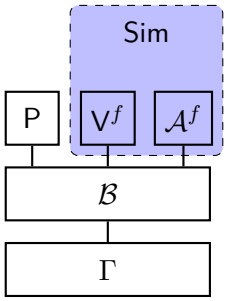

Figure 4: Visualization of the proof of Theorem 10. The black-box reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ wins the game $\mathcal{G} = (\Gamma, c)$ if $(\mathsf{P}, \mathsf{V}^f)$ forms a Σ–protocol Σ^f and $\mathcal{A}^f$ breaks the soundness of $\Pi_{\mathsf{FS}}[\Sigma^f]$. Since $\mathsf{Sim} = (\mathsf{Sim}_{\mathsf{V}}, \mathsf{Sim}_{\mathcal{A}})$ jointly simulates V^f and $\mathcal{A}^f$, neither $\mathcal{B}$ nor Γ can distinguish if Sim or $(\mathsf{V}^f, \mathcal{A}^f)$ is being used. Since Sim is efficient, this means $\mathcal{B}^{(\mathsf{P},\mathsf{Sim}_{\mathsf{V}},\mathsf{Sim}_{\mathcal{A}})}$ is an efficient machine that wins game $\mathcal{G}$.

$\mathcal{B}^{\mathsf{P},\mathsf{V}^f,\mathcal{A}^f}$ wins game $\mathcal{G}$. By replacing $(\mathsf{V}^f, \mathcal{A}^f)$ with a pair of simulators $(\mathsf{Sim}_{\mathsf{V}}, \mathsf{Sim}_{\mathcal{A}})$ such that no $\mathsf{poly}(n)$–query machine can distinguish between the two pairs, we obtain an efficient algorithm $\mathcal{B}^{\mathsf{P},\mathsf{Sim}_{\mathsf{V}},\mathsf{Sim}_{\mathcal{A}}}$ breaking the security of $\mathcal{G}$. We formalize this joint simulation below and then prove the black-box impossibility result using the strategy outlined above and pictured in Fig. 4.

Definition 13 (Joint Simulatability) A family of (possibly inefficient) algorithms $\{(\mathcal{A}^f, V^f)\}_f$ that have access to the same (possibly inefficient) resource $f : \{0,1\}^n \to \{0,1\}^m$ are *jointly simulatable* if there exist two QPT stateful algorithms Sim_1 and Sim_2 that share a *common state* and such that for any $\mathsf{poly}(n)$–query oracle access machine $M^{(\cdot,\cdot)}$,

$$\left| \Pr_f[M^{(\mathcal{A}^f, V^f)} = 1] - \Pr[M^{(\mathsf{Sim}_1,\mathsf{Sim}_2)} = 1] \right| \leq \mathsf{negl}(n) \ .$$

□

Theorem 10 *Let $\mathcal{G} = (\Gamma, c)$ be a cryptographic game assumption, let n, m be such that $n - m \in \omega(\lg n)$ and let $\Pi_{\mathsf{FS}}^{n,m}$ be a Fiat-Shamir instantiation in the CRQS model. There does not exist a black-box reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ showing the $\Sigma_{n,m}$–universality of $\Pi_{\mathsf{FS}}^{n,m}$ from the security of game $\mathcal{G}$, unless assumption $\mathcal{G}$ is false.*

Proof Assume there exists a black-box reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ showing the (n,m)–universality of $\Pi_{\mathsf{FS}}^{n,m}$ from the security of game $\mathcal{G}$. We will show that game $\mathcal{G}$ is insecure.

We begin by constructing a family of Σ–protocols that has bad challenge function f for any function $f \in \mathcal{F}^*$ where $\mathcal{F}^*$ is the set of functions for which the operators $P_{a,w}^f$ defined in Theorem 8 form a POVM with $P_\perp^f$. The Σ–protocol Σ^f defined below is an interactive proof of language membership for the empty language. On public input x,

1. P: does nothing.
2. V^f: interact with a potentially malicious prover in the following way.
   (a) On first message $a \in \{0,1\}^n$, pick $c \in_R \{0,1\}^m$ uniformly at random and send c to the prover.
   (b) On response z from the prover, accept iff $c = f(a)$.

This is indeed a Σ–protocol as it satisfies perfect correctness and special soundness.

Next, we build a dishonest prover that breaks the soundness of the QFS transform $\Pi_{\mathsf{FS}}^{n,m}[\Sigma^f]$ of this Σ–protocol. Recall that an implementation of Fiat-Shamir is also an implementation of WOTRO, but that might not be secure. Let $\{\mathcal{A}_{\mathsf{WOTRO}}^f\}_f$ be the attack from Theorem 9 against Π_{FS} (seen as an implementation of WOTRO), such that there exists a $\mathsf{negl}(n-m)$–simulatable such that $\mathcal{A}_{\mathsf{WOTRO}}^f$ produces $(a, f(a), v)$ that V_{FS} accepts

with probability $1 - \mathsf{negl}(n-m)$. Let $\mathsf{Sim}_{\mathsf{WOTRO}}$ be the simulator for $\{\mathcal{A}^f_{\mathsf{WOTRO}}\}_f$. For a function $f \in \mathcal{F}^*$, define the adversarial prover $\mathcal{P}^f$ that attacks protocol $\Pi^{n,m}_{\mathsf{FS}}[\Sigma^f]$ as follows:

1. Invoke $\mathcal{A}^f_{\mathsf{WOTRO}}$ on register P of the CRQS to obtain (a, c, v).
2. Send a, c, v and $z = \perp$ to the verifier.
3. Recall that the verifier for $\Pi^{n,m}_{\mathsf{FS}}[\Sigma^f]$ runs V_{FS} of Π_{FS} with message (a, c, v) on register V of the CRQS and then runs V^f of Σ^f on input (a, c, v, z).

The probability that the verifier accepts in protocol $\Pi^{n,m}_{\mathsf{FS}}[\Sigma^f]$ is equal to the probability that V_{FS} accepts and that $c = f(a)$, which by construction of $\mathcal{A}^f_{\mathsf{WOTRO}}$ happens with probability at least $1 - \mathsf{negl}(n-m)$.

Plugging P, V^f and $\mathcal{P}^f$ into the reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ gives an algorithm $\mathcal{B}^{\mathsf{P},\mathsf{V}^f,\mathcal{P}^f}$ that breaks the security of game $\mathcal{G}$, and yet that is not efficient. Using the simulator $\mathsf{Sim}_{\mathsf{WOTRO}}$ for the adversary $\mathcal{A}^f_{\mathsf{WOTRO}}$ allows us to replace the inefficient malicious prover $\mathcal{P}^f$ against the QFS transform with an indistinguishable efficient simulator, but V^f is still not efficiently computable.

We now show how $\mathcal{P}^f$ and V^f can be jointly simulated (Definition 13) using the stateless simulator $\mathsf{Sim}_{\mathsf{WOTRO}}$ for $\{\mathcal{A}^f_{\mathsf{WOTRO}}\}_f$. The two stateful algorithms $\mathsf{Sim}_{\mathcal{P}}$ and $\mathsf{Sim}_{\mathsf{V}}$ are defined as follows

1. Common State: a partial function $f_A : \{0,1\}^n \to \{0,1\}^m$ defined on an initially empty set $A = \emptyset$.
2. $\mathsf{Sim}_{\mathcal{P}}$: when invoked on a quantum register P, call the simulator $\mathsf{Sim}_{\mathsf{WOTRO}}$ for the family of adversaries $\{\mathcal{A}^f_{\mathsf{WOTRO}}\}_{f\in\mathcal{F}^*}$. Let $(a, c, v) \leftarrow \mathsf{Sim}_{\mathsf{WOTRO}}$, set $A \leftarrow A \cup \{a\}$ and $f_A(a) = c$, and return $(a, c, v, \perp)$. If $\mathsf{Sim}_{\mathsf{WOTRO}}$ produces an a that is already in A, the simulation fails.
3. $\mathsf{Sim}_{\mathsf{V}}$: when invoked on classical message (a, c, v, z) and quantum register V, run V_{FS} on register V of the CRQS with input (a, c, v). If $a \notin A$, pick $x \in_R \{0,1\}^m$ uniformly at random, set $A \leftarrow A \cup \{a\}$ and $f_A(a) = x$. Output `reject` if V_{FS} rejects or if $c \neq f_A(a)$, otherwise output `accept`.

Claim 1 *The pair of stateful (with common state) algorithms* $(\mathsf{Sim}_{\mathcal{P}}, \mathsf{Sim}_{\mathsf{V}})$ *jointly simulates* $\{(\mathcal{P}^f, \mathsf{V}^f)\}_{f\in\mathcal{F}^*}$.

Proof Let $M^{(\cdot,\cdot)}$ be an oracle-access machine and let $q = \mathsf{poly}(n)$ be an upper-bound on the number of queries made by M to either of its oracles. We first bound the probability that the simulation fails and then condition on the simulation succeeding. Let α denote the random variable of the value a produced by $\mathsf{Sim}_{\mathsf{WOTRO}}$. Since α is uniformly distributed (by the definition of $\mathsf{Sim}_{\mathsf{WOTRO}}$ in the proof of Theorem 9), on any given query, the probability that $\mathsf{Sim}_{\mathsf{WOTRO}}$ produces a that is already in the set A is upper-bounded by

$$\Pr[\alpha \in A] = \sum_{a\in A} \Pr[\alpha = a] \leq q \cdot 2^{-n} \ .$$

A union bound over the q queries allows us to upper-bound the probability that any of the queries returns an a that was already in A by $q^2 \cdot 2^{-n}$ which is $\mathsf{negl}(n)$.

Conditionned on the event that $\mathsf{Sim}_{\mathsf{WOTRO}}$ never produces $a \in A$, we show that black-box query access to $(\mathsf{Sim}_{\mathcal{P}}, \mathsf{Sim}_{\mathsf{V}})$ is indistinguishable on average over $f \in \mathcal{F}^*$ from black-box query access to $(\mathcal{P}^f, \mathsf{V}^f)$. First, observe that $\mathsf{Sim}_{\mathcal{P}}$ behaves exactly as $\mathcal{P}^f$, except that

it invokes $\mathsf{Sim}_{\mathsf{WOTRO}}$ instead of $\mathcal{A}^f_{\mathsf{WOTRO}}$. Therefore the BB-indistinguishability of $\mathsf{Sim}_{\mathcal{P}}$ and $\mathcal{P}^f$ follows from that of $\mathsf{Sim}_{\mathsf{WOTRO}}$ and $\mathcal{A}^f_{\mathsf{WOTRO}}$. Second, we note that $\mathsf{Sim}_{\mathsf{V}}$ picks each new point of the partial function f_A uniformly at random, so that f_A is identically distributed to a random function f restricted to A. Since a uniformly random function f is in $\mathcal{F}^*$ with probability at least $1 - \mathsf{negl}(n)$, we have that $\mathsf{Sim}_{\mathsf{V}}$ is indistinguishable from V^f on average over $f \in \mathcal{F}^*$. Finally, since we condition on the event $\alpha \notin A$ at every call of $\mathsf{Sim}_{\mathsf{WOTRO}}$, the answers of $\mathsf{Sim}_{\mathcal{P}}$ and $\mathsf{Sim}_{\mathsf{V}}$ are always consistent with the same function f (i.e. the simulation doesn't fail).

Therefore, the probability that $M^{(\cdot,\cdot)}$ distinguishes $(\mathcal{P}^f, \mathsf{V}^f)$ from $(\mathsf{Sim}_{\mathcal{P}}, \mathsf{Sim}_{\mathsf{V}})$ is at most the probability that $\mathsf{Sim}_{\mathsf{WOTRO}}$ and $\mathcal{A}^f_{\mathsf{WOTRO}}$ can be distinguished plus the probability that the simulation fails, which sum to at most $\mathsf{negl}(n-m)$. ■

We are now ready to conclude the proof. Given the reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ we construct an efficient algorithm for winning game $\mathcal{G}$ as follows. The machine $\mathcal{B}^{(\mathsf{P},\mathsf{Sim}_{\mathsf{V}},\mathsf{Sim}_{\mathcal{P}})}$ either:

1. wins game $\mathcal{G}$, or
2. if it does not, allows to distinguish $(\mathsf{Sim}_{\mathsf{V}}, \mathsf{Sim}_{\mathcal{P}})$ from $\{(\mathsf{V}^f, \mathcal{P}^f)\}_f$.

Since we have established the black-box indistinguishability of $(\mathsf{Sim}_{\mathsf{V}}, \mathsf{Sim}_{\mathcal{P}})$ and $\{(\mathsf{V}^f, \mathcal{P}^f)\}_{f\in\mathcal{F}^*}$, we conclude that a BB-reduction $\mathcal{B}^{(\cdot,\cdot,\cdot)}$ from the QFS-universality of $\Pi^{n,m}_{\mathsf{WOTRO}}$ to game $\mathcal{G}$ would allow to win the game. ■

7 A Quantum Assumption Allowing for $\mathsf{WOTRO}^{n,m}$

In [BLV06], Barak, Lindell, and Vadhan introduce a computational assumption allowing for Σ–universal Fiat-Shamir in the CRS model. It assumes the existence of a family of *entropy preserving* hash functions. In [DRV12], Dodis, Ristenpart, and Vadhan showed that a family of entropy preserving hash functions is necessary for a Σ–universal implementation of Fiat-Shamir in the CRS model. Of course, it follows from [BLV06; Bit+13] that this assumption cannot be black-box reduced to any cryptographic game. In this section, we define a different computational assumption allowing for $\mathsf{WOTRO}^{n,m}$ in the CRQ$ model (and therefore allowing for Σ–universal Fiat-Shamir). Our assumption is a quantum assumption on hash functions called a *collision-shelter*. We first show in Section 7.1 how to construct $\mathsf{WOTRO}^{n,n}$ with unconditional security in the CRQ$ model. In Section 7.2, we define the collision-shelter assumption and we show how to use it to convert $\mathsf{WOTRO}^{n,n}$ into a computationally secure $\mathsf{WOTRO}^{n,m}$ as long as $m \in \Omega(n)$. We conclude in Section 7.3 by a short discussion about some relations and distinctions between collision-shelters and collision resistant families of hash functions.

7.1 Unconditionally Secure $\mathsf{WOTRO}^{n,n}$ in the CRQ$ Model

Let us get back to the implementation of $\mathsf{WOTRO}^{n,n}$ roughly described in the introduction. The result stated in Theorem 11 requires the set of MUB to be the one introduced by Wootters and Fields in [WF89]. These bases are for the tensor product of n Hilbert spaces, each of odd prime dimension p. Let $\Gamma = \{0, \ldots, p-1\}$ denote the elements of the finite field $\mathbb{F}_p$ for $p \geq 3$ prime. We refer to the Wootters and Fields MUB for Γ^n as $\Theta^{p,n}_{\mathsf{WF}} = \{\theta_a\}_{a\in\Gamma^n}$ where $\theta_a = \{|x_a\rangle\}_{x\in\Gamma^n}$ is an orthonormal basis for Γ^n that, by virtue of mutual unbiasedness, satisfies $|\langle x_a|x'_{a'}\rangle| = p^{-\frac{n}{2}}$ when $a \neq a'$. The formal definition of $\Theta^{p,n}_{\mathsf{WF}}$ can be seen in Appendix B. The CRQ$ we use to implement $\mathsf{WOTRO}^{n,n}$ is composed of $3n$ p–dimensional EPR pairs, each denoted by $|\mathsf{EPR}_\Gamma\rangle_{PV} := \frac{1}{\sqrt{p}}\sum_{j\in\Gamma}|jj\rangle_{PV}$. The CRQ$ is

then set to $|\mathsf{EPR}_\Gamma^{3n}\rangle_{PV} := |\mathsf{EPR}_\Gamma\rangle^{\otimes 3n}$. Henceforth, we denote by $\mathsf{WOTRO}_\Gamma^{n,n}$ the primitive $\mathsf{WOTRO}^{n,n}$ where both the input and the output are in Γ^n.

Before giving our protocol $\Pi_{\mathsf{WRO}}^{n,n} = (\mathsf{P}', \mathsf{V}')$ for $\mathsf{WOTRO}_\Gamma^{n,n}$, we first consider a simpler (but insecure) version of it where the CRQ$ is $|\mathsf{EPR}_\Gamma^n\rangle_{PV}$ rather than $|\mathsf{EPR}_\Gamma^{3n}\rangle_{PV}$. Upon input $a \in \Gamma^n$, the simpler scheme asks the prover to measure register P of the CRQ$ in basis $\theta_a \in \Theta_{\mathsf{WF}}^{p,n}$ to obtain outcome $c \in \Gamma^n$. The prover then announces (a, c) to the verifier who verifies that when measuring register V of the CRQ$ the outcome c is obtained. If the test is perform with success then the output of the primitive is set to c.

This simple protocol cannot be proven secure as it stands. Instead, $\Pi_{\mathsf{WRO}}^{n,n}$ asks P' to measure 3 batches of EPR pairs $|\mathsf{EPR}_\Gamma^n\rangle_{PV}$ in the same basis θ_a to get outcomes $x_1, x_2, x_3 \in \Gamma^n$. The challenge produced by the primitive is then $c = x_3(x_1 + x_2)^{-1}$ (where the operations are done in $\mathbb{F}_{p^n}$). This choice for determining c follows from our proof technique. P' announces (a, x_1, x_2, x_3) that is checked by V' after measuring register V for each of the three instances of $|\mathsf{EPR}_\Gamma^n\rangle_{PV}$ in basis θ_a. If the test is successful then the output of the primitive is set to c.

Protocol $\Pi_{\mathsf{WRO}}^{n,n}$ for $\mathsf{WOTRO}_\Gamma^{n,n}$
Setup: A CRQ$ $|\mathsf{EPR}_\Gamma^{3n}\rangle_{PV}$.

Prover: On input $a \in \Gamma^n$,

1. Measures its part of $|\mathsf{EPR}_\Gamma^{3n}\rangle$ in basis $\theta_a^{\otimes 3}$, let $x = (x_1, x_2, x_3) \in \Gamma^{3n}$ be the result.
2. If $x_1 + x_2 = 0$, set $c = 0$. Otherwise, output $c := x_3(x_1 + x_2)^{-1}$ and sends (a, x) to verifier.

Verifier: Upon reception of (a, x),

1. Measure its part of $|\mathsf{EPR}_\Gamma^{3n}\rangle$ in basis $\theta_a^{\otimes 3}$, let $x' = (x'_1, x'_2, x'_3) \in \Gamma^{3n}$ be the result.
2. Output `reject` if $x \neq x'$ and output (a, c') where $c' = x'_3(x'_1 + x'_2)^{-1}$ otherwise.

Next theorem establishes that $\Pi_{\mathsf{WRO}}^{n,n}$ is $\frac{1}{4}$–secure against all adversaries. The proof is given in Appendix B and may be of independent interest. It consists in showing that the best measurement to distinguish the state transmitted by a quantum source that selects a basis $a \in_R \Gamma^n$ at random and sends $|x(a)_a\rangle$ for any set $\{(a, x(a))\}_{a\in\Gamma^n}$ cannot be recognized with probability better than $\frac{3}{4}$. Wootters and Fields' MUBs are useful here as this probability is given by a Weil sum that can be bounded by Deligne's resolution of Weil third conjecture[10] [Del74].

Theorem 11 *Let $\Gamma = \{0, \ldots, p-1\}$ be the set of elements in finite field $\mathbb{F}_p$ for $p \geq 3$ a prime number. Protocol $\Pi_{\mathsf{WRO}}^{n,n}$, presented above, is a statistically correct and statistically $(\frac{1}{4} - \mathsf{negl}(n))$–secure implementation of $\mathsf{WOTRO}_\Gamma^{n,n}$.*

We use a set of mutually unbiased bases (MUBs) introduced by Wootters and Fields in [WF89]. These bases of dimension p^n are for n instances of p–level quantum mechanical

[10] Weil's third conjecture is analogue to the Riemann hypothesis over finite fields and is called as such.

systems with $p \geq 3$ prime. The construction is as follows:

Definition 14 (Mutually Unbiased Bases of [WF89]) Let $p \geq 3$ be prime. Define the set of mutually unbiased bases $\Theta[\mathbb{F}_{p^n}] = \{\theta_a\}_{a\in\mathbb{F}_{p^n}}$ for a Hilbert space of dimension p^n where $\theta_a = \{|u\rangle_a\}_{u\in\mathbb{F}_{p^n}}$ is composed of vectors $|u\rangle_a$ expressed in the computational basis as

$$|u\rangle_a = p^{-\frac{n}{2}} \sum_{x\in\mathbb{F}_{p^n}} \exp\left(\frac{2\pi i}{p} \cdot \mathrm{tr}\left(ax^2 + ux\right)\right)|x\rangle \ , \tag{17}$$

where $\mathrm{tr} : \mathbb{F}_{p^n} \to \mathbb{F}_p$ denotes the *field trace* $\mathrm{tr}(x) := x + x^p + x^{p^2} + \cdots + x^{p^{n-1}}$. □

Notice that Klappenecker and Rötteler in [KR04] have shown a very similar construction for the case $p = 2$ (mutually unbiased bases of qubits). Unfortunately, our results do not apply to this construction as Weil sums need a field of odd characteristics.

Proof (of Theorem 11) For correctness, observe that if both parties are honest, their measurement triplets X and X' will be uniformly distributed and perfectly correlated unless $X_1 + X_2 = 0$. Since $X_1 + X_2$ is a random element of Γ^n due to it being the result of the measurement of EPR pairs, it holds that this event occurs with probability at most $|\Gamma|^{-n}$, which is negligible in n.

Now onto security. Let $\mathsf{c} : \Gamma^n \to \Gamma^n$ be an arbitrary target function. In order to cheat, i.e. to bias the output challenge towards $\mathsf{c}(a)$, a dishonest prover must produce a basis selected by a (the commitment) and measurement outcome x_1, x_2, x_3 such that

1. $x_3(x_1 + x_2)^{-1} = \mathsf{c}(a)$ and
2. V obtains the same outcomes x_1, x_2, x_3 when he measures his part of $|\mathsf{EPR}_\Gamma^{3n}\rangle$ in basis $\theta_a^{\otimes 3}$.

We say that x is a *bad* outcome if $x_3(x_1 + x_2)^{-1} = \mathsf{c}(a)$. Let $\mathcal{B}(a) \subseteq \Gamma^{3n}$ denote the set of bad outcomes for commitment a. Note that $|\mathcal{B}(a)| = p^{2n}$ for any $a \in \Gamma^n$.

The most general strategy for the prover is to apply a POVM $\{M_{a,x}\}_{a\in\Gamma^n, x\in\Gamma^{3n}}$ to its part of the EPR pairs to determine its message to V. The probability that P can bias the output towards $\mathsf{c}(a)$ when V accepts is then the probability that it can produce a commitment (i.e. a basis) such that a bad outcome will be observed by V in that basis.

$$\begin{aligned} P_w &= \Pr[X \in \mathcal{B}(A)] && (18)\\ &= \sum_{a\in\Gamma^n, x\in\mathcal{B}(a)} \mathrm{tr}\left((|x\rangle\langle x|_a \otimes M_{a,x}) \cdot \left|\mathsf{EPR}_\Gamma^{3n}\right\rangle\left\langle\mathsf{EPR}_\Gamma^{3n}\right|\right) && (19)\\ &= \frac{1}{p^{3n}} \sum_{a\in\Gamma^n, x\in\mathcal{B}(a)} \mathrm{tr}\left(M_{a,x}|x\rangle\langle x|_a\right) \ . && (20) \end{aligned}$$

To simplify our computations, we have slightly abused notation by writing $|x\rangle_a := |x_1\rangle_a \otimes |x_2\rangle_a \otimes |x_3\rangle_a$ when $x \in \Gamma^{3n}$ and $x_1, x_2, x_3 \in \Gamma^n$. Using this notation, for $x, y \in \Gamma^{3n}$ we have $|\langle x|_a|y\rangle_b|^2 = p^{-3n}$ whenever $a \neq b$.

The optimal cheating strategy for P can be framed as the solution to the following semidefinite program (SDP):

$$\begin{aligned} \max_{\{M_{a,x}\}} \quad & \frac{1}{p^{3n}} \sum_{a\in\Gamma^n} \sum_{x\in\mathcal{B}(a)} \mathrm{tr}\left(M_{a,x}|x\rangle\langle x|_a\right) \\ \text{s.t.} \quad & \sum_{a\in\Gamma^n} \sum_{x\in\mathcal{B}(a)} M_{a,x} \leq \mathbb{1} \ . \end{aligned} \tag{21}$$

The dual of this SDP is:

$$\begin{aligned} \min_{Z \geq 0} \quad & \frac{1}{p^{3n}} \operatorname{tr}(Z) \\ \text{s.t.} \quad & \forall a \in \Gamma^n, x \in \mathcal{B}(a) \quad |x\rangle\langle x|_a \leq Z \ . \end{aligned} \tag{22}$$

By the duality of semidefinite programming, a feasible solution to the dual will yield an upper-bound on the optimal solution of the primal. We now show how to construct a feasible solution that has constant value for $p^{-3n} \operatorname{tr}(Z)$.

Let $S = \sum_{a \in \Gamma^n, x \in \mathcal{B}(a)} |x\rangle\langle x|_a$ and define $f_\alpha(x) = \frac{x}{\alpha + x}$ for $\alpha \in \mathbb{R}$. Since f_α is an operator monotone function (meaning that $A \leq B \Rightarrow f_\alpha(A) \leq f_\alpha(B)$ for A, B positive semidefinite)[11], we have that $\frac{1}{\alpha+1}|x\rangle\langle x|_a \leq f_\alpha(S)$ for any $0 < \alpha \leq 1$. The operator $Z = (\alpha+1) f_\alpha(S)$ is thus a feasible solution to the dual with associated value $\frac{\alpha+1}{p^{3n}} \operatorname{tr}(f_\alpha(S))$.

We now proceed to upper-bound this probability. Since f_α is difficult to deal with directly, we will bound it using Taylor's theorem, yielding powers of Z that will then be easier to compute. To get a good bound, we will use a third degree Taylor bound for f_α centered around $\lambda \in \mathbb{R}$:

$$f_\alpha(x) \leq \frac{\lambda}{\alpha + \lambda} + \frac{\alpha}{(\alpha+\lambda)^2}(x - \lambda) - \frac{\alpha}{(\alpha+\lambda)^3}(x-\lambda)^2 + \frac{\alpha}{(\alpha+\lambda)^4}(x-\lambda)^3 \ .$$

Using the Taylor approximation defined above,

$$\begin{aligned} \frac{1}{p^{3n}} \operatorname{tr}(Z) &\leq \frac{\alpha+1}{p^{3n}} \operatorname{tr}(f_\alpha(S)) \\ &\leq \frac{\alpha+1}{p^{3n}} \Bigg(\frac{\lambda}{\alpha+\lambda} \operatorname{tr}(\mathbb{1}) + \frac{\alpha}{(\alpha+\lambda)^2} \operatorname{tr}(S - \lambda \mathbb{1}) \\ &\qquad - \frac{\alpha}{(\alpha+\lambda)^3} \operatorname{tr}\left((S-\lambda\mathbb{1})^2\right) + \frac{\alpha}{(\alpha+\lambda)^4} \operatorname{tr}\left((S-\lambda\mathbb{1})^3\right) \Bigg) \ . \end{aligned} \tag{23}$$

We can rewrite the above traces in the powers of $S - \lambda\mathbb{1}$ in the following way.

$$\left.\begin{aligned} \operatorname{tr}(\mathbb{1}) &= p^{3n} \ , \\ \operatorname{tr}(S - \lambda\mathbb{1}) &= \operatorname{tr}(S) - \lambda p^{3n} \ , \\ \operatorname{tr}\left((S-\lambda\mathbb{1})^2\right) &= \operatorname{tr}\left(S^2\right) - 2\lambda \operatorname{tr}(S) + \lambda^2 p^{3n} \ , \\ \operatorname{tr}\left((S-\lambda\mathbb{1})^3\right) &= \operatorname{tr}\left(S^3\right) - 3\lambda \operatorname{tr}\left(S^2\right) + 3\lambda^2 \operatorname{tr}(S) - \lambda^3 p^{3n} \ . \end{aligned}\right\} \tag{24}$$

We refer to Lemmas 7, 8 and 9 in Appendix B for the proofs that the following relations hold:

$$\operatorname{tr}(S) = p^{3n}, \ \operatorname{tr}\left(S^2\right) = 2 \cdot p^{3n} - p^{2n} \text{ and } \operatorname{tr}(S^3) \leq 4p^{3n} + p^{2n} \ .$$

Choosing to center the Taylor approximation around $\lambda = 1$ gives the following bounds for (24):

$$\begin{aligned} \operatorname{tr}(\mathbb{1}) &= p^{3n} \ , \\ \operatorname{tr}(S - \lambda\mathbb{1}) &= 0 \ , \\ \operatorname{tr}\left((S-\lambda\mathbb{1})^2\right) &= 2p^{3n} - p^{2n} - 2p^{3n} + p^{3n} \end{aligned}$$

[11] We can see this by noting that $\frac{x}{\alpha+a} = \alpha\left(\frac{1}{\alpha} - \frac{1}{\alpha+x}\right)$ and that $x \mapsto -x^{-1}$ is operator monotone, see e.g. [Car10, Lemma 2.7]

$$= p^{3n} - p^{2n} \ , \text{ and}$$
$$\operatorname{tr}\left((S-\lambda\mathbb{1})^3\right) \leq 4p^{3n} + p^{2n} - 3(2p^{3n} - p^{2n}) + 3p^{3n} - p^{3n}$$
$$= 4p^{2n} \ .$$

Substituting these values into (23), we get

$$P_w \leq \frac{1}{p^{3n}} \operatorname{tr}(Z) \leq \frac{\alpha+1}{p^{3n}} \left(\frac{p^{3n}}{\alpha+1} - \frac{\alpha(p^{3n}-p^{2n})}{(\alpha+1)^3} + \frac{\alpha \cdot 4p^{2n}}{(\alpha+1)^4} \right) \ .$$

Looking only at the non-negligible terms, we have

$$P_w \leq 1 - \frac{\alpha}{(\alpha+1)^2} + \mathsf{negl}(n)$$

which is minimized at $\alpha = 1$ with value $P_w \leq \frac{3}{4} + \mathsf{negl}(n)$. Since this probability is the same for all functions $\mathsf{c}(\cdot)$, it follows that the protocol $(\frac{1}{4} - \mathsf{negl}(n))$–avoids all functions. ■

7.2 Collision-Shelters

We are now ready to define a quantum computational assumption that allows for a secure implementation of $\mathsf{WOTRO}^{n,m}$ for $m < n$. A collision-shelter for security parameter n, is is a family $\mathbb{G}_\Gamma^{n,m} = \{G_s^n : \Gamma^n \times \Gamma^n \to \Gamma^m\}_{s \in \{0,1\}^{\ell(n)}}$ of hash functions that exhibits a strong quantum flavour of collision resistance. Intuitively, $\mathbb{G}_\Gamma^{n,m}$ is a *collision-shelter* if, for any function $\mathsf{c} : \Gamma^n \to \Gamma^m$, no QPT adversary can produce a state *close* to

$$|\psi_s\rangle = \sum_a \alpha_a |a\rangle_A \otimes \sum_{c : G_s^n(a,c) = \mathsf{c}(a)} \gamma_c^a |c\rangle_C \otimes |\varphi(a,c)\rangle_{W'} \ , \tag{25}$$

for $s \in_R \{0,1\}^{\ell(n)}$ and in average over outcome a when register A is measured in the computational basis, $\sum_{c : G_s^n(a,c) = \mathsf{c}(a)} \gamma_c^a |c\rangle_C \otimes |\varphi(a,c)\rangle_{W'}$ *contains* collisions in superposition. Notice that no such state can be produced efficiently when the number of possible a is in $O(\lg n)$ and G_s^n is collision resistant, as the generation of 2 such states would provide a collision for G_s^n with good probability.

Definition 15 (δ–Colliding States) Let $\mathsf{c} : \Gamma^n \to \Gamma^m$ be arbitrary and let $G_s^n \in \Gamma^n \times \Gamma^n \to \Gamma^m$. Let

$$|\psi\rangle = \sum_a \alpha_a |a\rangle_A \otimes \sum_{c : G_s^n(a,c) = \mathsf{c}(a)} \gamma_c^a |c\rangle_C \otimes |\varphi(a,c)\rangle_{W'}$$

be a state hitting function $\mathsf{c}(\cdot)$. Let $c^*(a)$ be such that $|\gamma_{c^*(a)}^a|^2 = \max_c \{|\gamma_c^a|^2\}$ for every $a \in \{0,1\}^n$ and let $|\tilde{\psi}^*\rangle = \sum_a \alpha_a |a\rangle_A \otimes \gamma_{c^*(a)}^a |c^*(a)\rangle_X \otimes |\varphi(a, c^*(a))\rangle_W$ be the corresponding sub-normalized state obtained from $|\psi\rangle$. If $\||\tilde{\psi}^*\rangle\|^2 < 1 - \delta$ then $|\psi\rangle$ is said to be *δ–colliding to $\mathsf{c}(\cdot)$ under G_s^n*. □

A collision-shelter is a family of hash functions (efficiently samplable and efficiently evaluable) that prevents any QPT adversary from generating a δ–colliding state hitting any function $\mathsf{c}(\cdot)$.

Definition 16 (Collision Shelter) The efficiently samplable and efficiently evaluable hash function family $\mathbb{G}_\Gamma^{n,m} = \{G_s^n : \Gamma^n \times \Gamma^n \to \Gamma^m\}_{s \in \{0,1\}^{\ell(n)}}$ is a *collision-shelter* if, for all $\delta > 0$, all functions $\mathsf{c} : \Gamma^n \to \Gamma^m$, and all QPT adversaries $\mathcal{A} = \{\mathcal{A}_n\}$, the probability over $s \in_R \{0,1\}^{\ell(n)}$ that $|\psi\rangle_{ACW'} \leftarrow \mathcal{A}_n(s)$ is δ–colliding to $\mathsf{c}(\cdot)$ under $\mathbb{G}_\Gamma^{n,m}$ is negligible in n. The *collision-shelter assumption* simply posits the existence of a collision-shelter $\mathbb{G}_\Gamma^{n,m}$ for $m \leq (1-\alpha)n$ with $0 < \alpha < 1$. □

We now consider the obvious implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$ using $\Pi_{\mathsf{WRO}}^{n,n}$ and a function-shelter $\mathbb{G}_\Gamma^{n,m}$ that simply sets the challenge $\hat{c} \in \Gamma^m$ as $\hat{c} = G_s^n(a,c)$ where $c \in \Gamma^n$ is the challenge produced in $\Pi_{\mathsf{WRO}}^{n,n}$ and s is the CR\$. Let us denote this implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$ by $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$. The following theorem is an easy consequence of Definition 16 and Theorem 11.

Theorem 12 *Assuming that $\mathbb{G}_\Gamma^{n,m} = \{G_s^n\}_s$ is a collision-shelter with Γ a set of elements in finite field $\mathbb{F}_p$ with $p \geq 3$ prime. Then, $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$ is a computationally $\left(\frac{1}{4} - o(1)\right)$–secure implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$.*

Proof Let $\mathcal{A}_n$ be a purified adversary against $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$, i.e. whose actions are described by a unitary transform up to the point where it needs to send a classical message to the verifier. Let $s \in \{0,1\}^{\ell(n)}$ be the CR\$ of an execution of $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$ where the adversary $\mathcal{A}_n$ hits function $\mathsf{c} : \{0,1\}^n \to \{0,1\}^m$ with probability p_{hit}. It follows that $|\psi_s^{\mathcal{A}_n}\rangle$ – the joint state of $\mathcal{A}_n$ and V generated from the CRQ\$ by $\mathcal{A}_n$ after running $\Pi_{\mathsf{WRO}}^{n,n}$ in $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$ (but before measuring) – can be assumed w.l.g. to be of the following form

$$\left|\psi_s^{\mathcal{A}_n}\right\rangle_{ACXWV} = \sqrt{p_{\text{hit}}}|\psi_s\rangle_{ACXWV} + \sqrt{1-p_{\text{hit}}}|\boxtimes\rangle_{ACXWV} \ ,$$

where $|\boxtimes\rangle_{ACXWV}$ results either in V's rejection or acceptance without hitting target $\mathsf{c}(\cdot)$ and $|\psi_s\rangle_{AXWV}$ hits function $\mathsf{c}(\cdot)$,

$$\begin{aligned}
|\psi_s\rangle_{ACXWV} &= \sum_{a\in\{0,1\}^n} \alpha_a|a\rangle_A \sum_{x:G_s(a,x_3(x_1+x_2)^{-1})=\mathsf{c}(a)} \beta_x^a \left|x_3(x_1+x_2)^{-1}\right\rangle_C |x\rangle_X |\varphi(a,x)\rangle_W |\mathrm{v}(a,x)\rangle_V \\
&= \sum_{a\in\{0,1\}^n} \alpha_a|a\rangle_A \sum_{\substack{c:G_s(a,c)=\mathsf{c}(a)\\ x:x_3(x_1+x_2)^{-1}=c}} \beta_x^a|c\rangle_C|x\rangle_X|\varphi(a,x)\rangle_W|\mathrm{v}(a,x)\rangle_V \\
&= \sum_{a\in\{0,1\}^n} \alpha_a|a\rangle_A \sum_{c:G_s(a,c)=\mathsf{c}(a)} \gamma_c^a|c\rangle_C \sum_{x:x_3(x_1+x_2)^{-1}=c} \hat{\beta}_{c,x}^a|\hat{\varphi}(a,x)\rangle_{XWV} \ .
\end{aligned}$$

Registers A and X contain the final results $a \in \Gamma^n$ and $x \in \Gamma^{3n}$ for the execution of $\Pi_{\mathsf{WRO}}^{n,n}$ in $\Pi_{\mathsf{WRO}}^{n,m}[\mathbb{G}_\Gamma^{n,m}]$ and C contains the final challenge $c = x_3(x_1+x_2)^{-1}$, W is $\mathcal{A}_n$'s working register, and V is the register for the state of the verifier $|\mathrm{v}(a,x)\rangle$ when (a,x) is announced. We assume that when V measures $|\mathrm{v}(a,x)\rangle$, it always accepts (a,x) and therefore sets the final challenge as $c = x_3(x_1+x_2)^{-1}$. Notice that if $\mathcal{A}_n$ produces state $|\psi_s^{\mathcal{A}_n}\rangle$ efficiently then it can also produce $|\psi_s\rangle$ efficiently as long as p_{hit} is polynomial since projecting $|\psi_s^{\mathcal{A}_n}\rangle$ into the subspace of states hitting $\mathsf{c}(\cdot)$ can be implemented efficiently. Let

$$\left|\tilde{\psi}_s^*\right\rangle_{ACXWV} = \sum_a \alpha_a|a\rangle_A \otimes \gamma_{\mathsf{c}^*(a)}^a|\mathsf{c}^*(a)\rangle_C \otimes \sum_{x:x_3(x_1+x_2)^{-1}=\mathsf{c}^*(a)} \beta_{\mathsf{c}^*(a),x}^a|\hat{\varphi}(a,x)\rangle_{XWV}$$

be defined so that $\mathsf{c}^*(a) \in \Gamma^n$ is such that $|\gamma_{\mathsf{c}^*(a)}^a|^2$ is maximum for each $a \in \Gamma^n$. By definition 16, $\mathbb{G}_\Gamma^{n,m} = \{G_s^n\}_s$ being a collision-shelter implies that

$$\left|\left\langle\psi_s\middle|\tilde{\psi}_s^*\right\rangle\right|^2 \geq 1 - o(1) \ . \tag{26}$$

$\mathcal{A}_n$'s strategy is almost the same as the one where $|\psi_s\rangle$ is replaced by $|\tilde{\psi}_s^*\rangle$. By theorem 11, the (subnormalized) state

$$\left|\hat{\psi}_s^{\mathcal{A}_n}\right\rangle = \sqrt{p_{\text{hit}}}\left|\tilde{\psi}_s^*\right\rangle + \sqrt{1-p_{\text{hit}}}|\boxtimes\rangle$$

allows to hit function $\mathsf{c}^*(\cdot)$ in $\Pi_{\mathsf{WRO}}^{n,n}$ with probability $p_{\text{hit}} \leq \frac{3}{4} + \mathsf{negl}(n)$. From (26), we therefore have that $|\psi_s^{\mathcal{A}_n}\rangle$ hits target function $\mathsf{c}(\cdot)$ with probability no larger than $\frac{3}{4} + o(1) + \mathsf{negl}(n)$. The result follows. ■

7.3 Is the Collision-Shelter Assumption Realistic?

While $h : \Gamma^{\ell(n)} \times \Gamma^n \to \Gamma^m$ is *entropy-preserving* if no efficient adversary can, given the first argument $\mathbf{s} \in_R \Gamma^{\ell(n)}$ picked uniformly at random, find $x \in \Gamma^n$ such that $h(\mathbf{s}, x)$ has almost no entropy (when $\mathbf{s}$ has been forgotten), collision-shelters prevent efficient quantum adversaries from preparing a state with entropy in the second argument when the output of the hash function applied to both arguments is fixed to a function of its first argument. Why would it be possible for collision-shelters to exist?

Suppose that for all $a \in \Gamma^n$, the hash function $G_s^n(a, \cdot)$ is collision-resistant against quantum adversaries. Let $\mathsf{c}(a) \in \Gamma^m$ be arbitrary. It follows that for any $a \in \Gamma^n$, no efficient quantum adversary can produce a state of the form $|\psi_a\rangle = \sum_{x:G_s^n(a,x)=\mathsf{c}(a)} \beta_x^a |x\rangle_X \otimes |\varphi(a,x)\rangle$ where $|\langle\psi_a|\tilde{\psi}_a^*\rangle|^2 < 1 - \delta$ since two states of that form would allow to find a collision with non-negligible probability. This, of course, does not imply that $\{G_s^n(\cdot,\cdot)\}_s$ is a collision-shelter as $G_s^n(a,\cdot) = G_s^n(a',\cdot) =: h_s(\cdot)$ for all $a, a' \in \Gamma^n$ is such that $G_s^n(a,\cdot)$ is collision-resistant when $h_s(\cdot)$ is collision-resistant but the following easy-to-generate state is $o(1)$–colliding to $\mathsf{c}(a) = a_1 \dots a_m$ when $m < n$. We start with a uniform superposition over x over $|x\rangle|h_s(x)\rangle$, which "fixes" the first m values of a, and introduce a superposition over the $n - m$ remaining values $a_{m+1} \dots a_n$:

$$\begin{aligned} p^{-n/2} \sum_x |x\rangle_X \otimes |h_s(x)\rangle &\mapsto p^{-n+\frac{m}{2}} \sum_a \sum_{x:h_s(x)=a_1\dots a_m} |x\rangle_X \otimes |a\rangle_A \\ &= p^{-n/2} \sum_a |a\rangle_A \otimes p^{\frac{-n+m}{2}} \sum_{x:G_s^n(a,x)=\mathsf{c}(a)} |x\rangle_X \ . \end{aligned}$$

Such an attack seems difficult to conduct when $\{G_s^n(a,\cdot)\}_a$ is a set of collision resistant hash functions that *appear independent* of each other as far as collisions are concerned. What it means exactly for hash functions in $\{G_s^n(a,\cdot)\}_a$ to *appear independent* is unclear. It is easy to see that a random oracle $\mathcal{O}_\Gamma^{n,m} : \Gamma^n \times \Gamma^n \to \Gamma^m$ acts as a collision-shelter, in fact, it is a much stronger assumption as $\Pi_{\mathsf{WRO}}^{n,m}[\mathcal{O}_\Gamma^{n,m}]$ is already statistically secure.

Theorem 13 *Let $\mathcal{O}_\Gamma^{n,m} : \Gamma^n \times \Gamma^n \to \Gamma^m$ be a random oracle accepting quantum queries. Then, $\mathcal{O}_\Gamma^{n,m}$ is a collision-shelter and moreover, $\Pi_{\mathsf{WRO}}^{n,m}[\mathcal{O}_\Gamma^{n,m}]$ is $(1 - \mathsf{negl}(n))$–secure.*

Proof As before, we set $p := |\Gamma|$. Let $\mathsf{c} : \Gamma^n \to \Gamma^m$ be an arbitrary target function and let $L_\mathsf{c} := \{(a,c) \in \Gamma^n \times \Gamma^n \,|\, \mathcal{O}_\Gamma^{n,m}(a,c) = \mathsf{c}(a)\}$. In order to succeed, an adversary must announce $(a,c) \in L_\mathsf{c}$. Suppose for a contradiction that $\mathcal{A}_n$ hits $\mathsf{c}(\cdot)$ with non-negligible probability $\delta(n) := \frac{1}{q(n)}$ for a positive polynomial $q(n)$. Remember that $\mathcal{A}_n$ wins in $\Pi_{\mathsf{WRO}}^{n,m}[\mathcal{O}_\Gamma^{n,m}]$ when it can produce $(a,c) \in L_\mathsf{c}$. It corresponds to searching one element of L_c in a random database containing $N = p^{2n}$ elements. Let $t = |L_\mathsf{c}|$ be the number of good elements for the adversary. It is straightforward to see that except with negligible probability, $t \leq N/p^{(1-\epsilon)m}$. In [BW98], it is shown that searching in a database of size N for a solution when there are no more than t solutions using T queries has a probability of error p_e that satisfies

$$p_e \geq \exp\left(-4bT^2/(N-t) - 8T\sqrt{tN/(N-t)^2}\right) \ , \tag{27}$$

for $b > 0$ some fixed constant. Since $e^{-x} \geq 1 - x$, we then get that for $T \in \mathsf{poly}(n)$, (27) satisfies

$$p_e \geq 1 - \mathsf{negl}(n) \ .$$

■

The random oracle makes it difficult to produce a single $(a, c) \in L_{\mathsf{c}}$ while a collisions-shelter posits that it is difficult to generate a collision on $\mathsf{c}(a)$ for many $a \in \Gamma^n$, which at least requires finding $(a, c) \in L_{\mathsf{c}}$.

Notice that while any secure universal Fiat-Shamir transform in the CRS model requires the existence of an entropy-preserving family of hash functions[DRV12], this does not seem to be the case for collisions-shelters with respect to WOTRO in the CRQS model.

8 Black-Box Impossibility of a Flavour of Quantum Lightning

In this section, we show that a secure WOTRO can be constructed from a quantum lightning scheme that satisfies a slightly stronger security notion. Quantum lightning was introduced by Zhandry in [Zha19] as a primitive allowing for publicly verifiable quantum money schemes and provable randomness among others.

8.1 Typed Quantum Lightning.

Quantum lightning provides some fresh randomness that even an adversarial procedure cannot bias towards a certain value. We present a strenghtened version of this property that requires that this randomness remains in the presence of an input to the lightning generation procedure. This notion is sufficiently strong to provide a secure WOTRO protocol.

Definition 17 A *typed quantum lightning scheme* is a tuple of QPT algorithms $(\mathsf{tQLSetup}, \mathsf{tQLGen}, \mathsf{tQLVer})$ where

- $\mathsf{tQLSetup}(1^n)$ produces a storm 🌩.
- $\mathsf{tQLGen}(🌩, a)$ takes an additional parameter $a \in \{0,1\}^n$, and produces a lightning state $|🗲_a\rangle$.
- $\mathsf{tQLVer}(🌩, |🗲\rangle)$ returns the type a, a serial number s or $\perp$ if the state is not valid, and a leftover quantum register.

Correctness is defined similarly to regular QL: serial numbers are deterministic for honestly generated bolts and verification does not noticeably affect the bolt. The security properties of a tQL scheme are as follows: For any QPT adversary $\mathcal{A}$ that on input 🌩 produces a type $A \in \{0,1\}^n$ and a state $|🗲\rangle$, if we let $\rho_{QSA'} = \mathsf{tQLVer}(🌩, |🗲\rangle)$, then

$$\Pr\left[H_\infty(S \mid A \wedge (S \neq \perp) \wedge (A = A')) \leq \lg p(n)\right] \leq \mathsf{negl}(n) \ .$$

□

Based on Definition 17, typed quantum lightning provides randomness in the serial number conditionned on the type. It is the ability of the adversary to choose the type a that makes this primitive stronger than regular QL. A natural WOTRO protocol in the CRS+CRQ$ model based on this new primitive is presented below.

Protocol $\Pi^{\mathsf{tQL}}_{\mathsf{WRO}}$ for $\mathsf{WOTRO}^{n,m}$

Setup: A CRS containing 🌩 $\leftarrow \mathsf{tQLSetup}(1^n)$ for a tQL scheme with n–bit types and m–bit serial numbers. A CRQ$ containing $|\mathsf{EPR}\rangle^{\otimes q}$ where q is the qubit size of a tQL state.

1. On input $a \in \{0,1\}^n$, P calls $|\lightning_a\rangle \leftarrow \mathsf{tQLGen}(\text{🌧}, a)$, sets $\rho_{QSA} = \mathsf{tQLVer}(\text{🌧}, |\lightning_a\rangle)$, teleports register Q to V using the EPR pairs and sends (A, S) to V.

2. Upon reception of (a, s, ρ_Q), V calls $\sigma_{Q'S'A'} \leftarrow \mathsf{tQLVer}(\text{🌧}, \rho_Q)$ and tests that $A' = a$ and $S' = s$. V aborts if the tests failed, otherwise V sets $c = s$ and outputs (a, c).

Theorem 14 *The above protocol is a secure instantiation of* $\mathit{WOTRO}^{n,m}$.

The proof is a direct consequence of the security of the tQL primitive.

Corollary 2 *There is no black-box reduction from the security of a* tQL *scheme with type length* n *and serial length* m *satisfying* $n - m \in \omega(\lg n)$ *to the security of a cryptographic game assumption, unless the assumption is false.*

8.2 Justification for the tQL assumption.

Why is typed quantum lightning a realistic assumption? It turns out that the tQL primitive can be built from "vanilla" QL for types of length $O(\lg n)$. We present a construction of a tQL scheme for $\lg p(n)$ bits types for any polynomial $p(\cdot)$ from an arbitrary (regular) QL scheme.

Prerequisite: A QL scheme $(\mathsf{QLSetup}, \mathsf{QLGen}, \mathsf{QLVer})$. A family of $n \cdot p(n)$–wise independent hash functions $\mathcal{H} \subset \{0,1\}^n \to \{0,1\}^{\lg p(n)}$.

- $\mathsf{tQLSetup}(1^n)$: Let $\text{🌧} \leftarrow \mathsf{QLSetup}(1^n)$ and $h \leftarrow\!\!\$\ \mathcal{H}$, output $\text{🌧}' = (\text{🌧}, h)$
- $\mathsf{tQLGen}(\text{🌧}', a)$: Parse $\text{🌧}'$ as $(\text{🌧}, h)$. Do $|\lightning\rangle \leftarrow \mathsf{QLGen}(\text{🌧})$ until $s = \mathsf{QLVer}(\text{🌧}, |\lightning\rangle)$ satisfies $h(s) = a$ and output $|\lightning\rangle$.
- $\mathsf{tQLVer}(\text{🌧}', |\lightning\rangle)$: Parse $\text{🌧}'$ as $(\text{🌧}, h)$. Compute $\rho_{SQ} \leftarrow \mathsf{QLVer}(\text{🌧}, |\lightning\rangle)$ and set $A = h(S)$. Output ρ_{ASQ}.

Theorem 15 $(\mathsf{tQLSetup}, \mathsf{tQLGen}, \mathsf{tQLVer})$ *is a* tQL *scheme of* $\lg p(n)$ *bits types.*

Proof Correctness follows from that of the underlying QL scheme: a state produced by tQLGen will be recognized as a valid state by tQLVer if QLGen produces valid states.

The expected running time of tQLGen is exponential in $\lg p(n)$ and thus polynomial in n. Since h is sampled from a family of $n \cdot p(n)$–pairwise independent hash functions, the probability that tQLGen does not produce an output after $n \cdot p(n)$ steps is at most

$$\Pr[h(s_1) \neq a \wedge \cdots \wedge h(s_{n \cdot p(n)}) \neq a] = \left(1 - \frac{1}{p(n)}\right)^{n \cdot p(n)} \leq e^{-n}$$

For security (Definition 17), let $\mathcal{A}$ be an attacker against the min-entropy of the tQL scheme, i.e. $\mathcal{A}$ produces with inverse polynomial probability a state $|\lightning\rangle$ such that $\rho_{ASQ} \leftarrow \mathsf{tQLVer}(\text{🌧}, |\lightning\rangle)$ has logarithmic min-entropy in S conditioned on A:

$$\Pr\left[H_\infty(S \mid A \wedge (S \neq \bot)) \leq \lg n^r\right] \geq \frac{1}{n^k} \tag{28}$$

for some $r, k > 0$. We construct an adversary $\mathcal{B}$ against the uniqueness of the original lightning scheme from this $\mathcal{A}$. The strategy of $\mathcal{B}$ is as follows: call $\mathcal{A}(1^n)$ twice to obtain $|\lightning_1\rangle$ and $|\lightning_2\rangle$, if $\mathsf{QLVer}(|\lightning_1\rangle) = \mathsf{QLVer}(|\lightning_2\rangle)$ halt and output $|\lightning_1\rangle$ and $|\lightning_2\rangle$, otherwise repeat. We now show that this strategy will produce a collision for the underlying QL scheme with an expected polynomial number of calls to $\mathcal{A}$.

Let $\bar{a}$ be such that $H_\infty(S \mid (A = \bar{a}) \wedge (S \neq \bot)) \leq \lg n^r$ and such that $\Pr[A = \bar{a}] \geq \frac{1}{q(n)}$ for some polynomial $q(\cdot)$ when ρ_{ASQ} is obtained from $\mathsf{tQLVer}(\mathcal{A}(\includegraphics[height=1em]{sheep.pdf}))$. Note that since a is $\lg p(n)$ in length, such an $\bar{a}$ must exist for (28) to hold (otherwise all a that have low conditional min-entropy have negligible probability of being produced by $\mathcal{A}$). Then for each pair of invocations of $\mathcal{A}$, the following holds with probability at least $\frac{1}{n^k}$:

$$
\begin{aligned}
&\Pr[\mathsf{QLVer}(|\lightning_1\rangle) = \mathsf{QLVer}(|\lightning_2\rangle)] \\
&\geq \frac{1}{q(n)^2} \Pr[\mathsf{QLVer}(|\lightning_1\rangle) = \mathsf{QLVer}(|\lightning_2\rangle) \mid A_1 = \bar{a} \wedge A_2 = \bar{a}] \\
&\geq \frac{1}{q(n)^2} 2^{-H_2(\mathsf{QLVer}(\rho)|A=\bar{a})} \\
&\geq \frac{1}{q(n)^2} 2^{-H_\infty(\mathsf{QLVer}(\rho)|A=\bar{a})} \\
&\geq \frac{1}{q(n)^2} \frac{1}{n^r}
\end{aligned}
$$

where H_2 denotes the collision entropy and is upper-bounded by the min-entropy H_∞. The probability that $\mathcal{B}$ halts and succeeds is therefore at least $(q(n)^2 n^{r \cdot k})^{-1}$. ■

References

[AG22] P. Ananth and A. B. Grilo. *Post-Quantum Zero-Knowledge with Space-Bounded Simulation*. Oct. 12, 2022. DOI: 10.48550/arXiv.2210.06093. arXiv: 2210.06093[quant-ph].

[Ara02] P. Aravind. "Bell's theorem without inequalities and only two distant observers". In: *Foundations of Physics Letters* 15 (2002), pp. 397–405. DOI: 10.1023/A:1021272729475.

[Ara03] P. Aravind. *A simple demonstration of Bell's theorem involving two observers and no probabilities or inequalities*. 2003. DOI: 10.48550/arXiv.quant-ph/0206070. arXiv: quant-ph/0206070.

[ARU14] A. Ambainis, A. Rosmanis, and D. Unruh. "Quantum Attacks on Classical Proof Systems: The Hardness of Quantum Rewinding". In: *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*. Oct. 2014, pp. 474–483. DOI: 10.1109/FOCS.2014.57. arXiv: 1404.6898.

[AW02] R. Ahlswede and A. Winter. "Strong converse for identification via quantum channels". In: *IEEE Transactions on Information Theory* 48.3 (2002), pp. 569–579. DOI: 10.1109/18.985947. arXiv: quant-ph/0012127.

[BBF13] P. Baecher, C. Brzuska, and M. Fischlin. "Notions of Black-Box Reductions, Revisited". In: *Advances in Cryptology - ASIACRYPT 2013*. Springer, 2013, pp. 296–315. DOI: 10.1007/978-3-642-42033-7˙16.

[BBT05] G. Brassard, A. Broadbent, and A. Tapp. "Quantum Pseudo-Telepathy". In: *Foundations of Physics* 35 (2005), pp. 1877–1907. DOI: 10.1007/s10701-005-7353-4. arXiv: quant-ph/0407221.

[BFM88] M. Blum, P. Feldman, and S. Micali. "Non-interactive Zero-knowledge and Its Applications". In: *Proc. Twentieth Annual ACM Symposium on Theory of Computing*. STOC '88. ACM, 1988, pp. 103–112. DOI: 10.1145/62212.62222.

[BGW12] N. Bitansky, S. Garg, and D. Wichs. *Why Fiat-Shamir for proofs lacks a proof*. 2012. IACR eprint: 2012/705.

[Bit+13] N. Bitansky, D. Dachman-Soled, S. Garg, A. Jain, Y. T. Kalai, A. López-Alt, and D. Wichs. "Why "fiat-shamir for proofs" lacks a proof". In: *Theory of Cryptography Conference*. Springer, 2013, pp. 182–201. DOI: 10.1007/978-3-642-36594-2˙11.

[Blu+91] M. Blum, A. De Santis, S. Micali, and G. Persiano. "Noninteractive Zero-Knowledge". In: *SIAM Journal on Computing* 20.6 (1991), pp. 1084–1118. DOI: 10.1137/0220068.

[BLV06] B. Barak, Y. Lindell, and S. Vadhan. "Lower bounds for non-black-box zero knowledge". In: *Journal of Computer and System Sciences*. JCSS FOCS 2003 Special Issue 72.2 (Mar. 1, 2006), pp. 321–391. DOI: 10.1016/j.jcss.2005.06.010.

[BR93] M. Bellare and P. Rogaway. "Random Oracles Are Practical: A Paradigm for Designing Efficient Protocols". In: *Proc. 1st ACM Conference on Computer and Communications Security*. CCS '93. ACM, 1993, pp. 62–73. DOI: 10.1145/168588.168596.

[BW98] H. Buhrman and R. de Wolf. "Lower Bounds for Quantum Search and Derandomization". In: (1998). DOI: 10.48550/arXiv.quant-ph/9811046. arXiv: quant-ph/9811046 [quant-ph].

[Can+18] R. Canetti, Y. Chen, J. Holmgren, A. Lombardi, G. N. Rothblum, and R. D. Rothblum. *Fiat-Shamir From Simpler Assumptions*. Cryptology ePrint Archive, Report 2018/1004. 2018. IACR eprint: 2018/1004.

[Can01] R. Canetti. "Universally Composable Security: A New Paradigm for Cryptographic Protocols". In: *Proc. 42Nd IEEE Symposium on Foundations of Computer Science*. FOCS '01. IEEE Computer Society, 2001, pp. 136–. DOI: 10.1109/SFCS.2001.959888.

[Car10] E. A. Carlen. *Trace inequalities and quantum entropy: An introductory course*. 2010.

[CF01] R. Canetti and M. Fischlin. "Universally Composable Commitments". In: *Proc. CRYPTO*. CRYPTO '01. 2001, pp. 19–40. DOI: 10.1007/3-540-44647-8˙2.

[CGH04] R. Canetti, O. Goldreich, and S. Halevi. "The Random Oracle Methodology, Revisited". In: *J. ACM* 51.4 (July 2004), pp. 557–594. DOI: 10.1145/1008731.1008734.

[Cra96] R. Cramer. "Modular Design of Secure yet Practical Cryptographic Protocols". PhD thesis. CWI and University of Amsterdam, 1996.

[CVZ20] A. W. Coladangelo, T. G. Vidick, and T. Zhang. "Non-Interactive Zero-Knowledge Arguments for QMA, with preprocessing". In: Springer-Verlag, 2020. DOI: 10.1007/978-3-030-56877-1˙28.

[CX22] S. Cao and R. Xue. "The Gap Is Sensitive to Size of Preimages: Collapsing Property Doesn't Go Beyond Quantum Collision-Resistance for Preimages Bounded Hash Functions". In: *Advances in Cryptology – CRYPTO 2022*. Lecture Notes in Computer Science. Springer Nature Switzerland, 2022, pp. 564–595. DOI: 10.1007/978-3-031-15982-4˙19.

[Dac+12] D. Dachman-Soled, A. Jain, Y. T. Kalai, and A. Lopez-Alt. *On the (In)security of the Fiat-Shamir Paradigm, Revisited*. Cryptology ePrint Archive, Report 2012/706. 2012. IACR eprint: 2012/706.

[Dam10] I. Damgård. *On Sigma-protocols*. 2010.

[Del74] P. Deligne. "La conjecture de Weil : I". In: *Publications Mathématiques de l'IHÉS* 43 (1974), pp. 273–307. DOI: 10.1007/BF02684373.

[DFS04] I. Damgård, S. Fehr, and L. Salvail. "Zero-Knowledge Proofs and String Commitments Withstanding Quantum Attacks". In: *Proc. CRYPTO*. 2004, pp. 254–272. DOI: 10.1007/978-3-540-28628-8˙16.

[DN02] I. Damgård and J. B. Nielsen. "Perfect Hiding and Perfect Binding Universally Composable Commitment Schemes with Constant Expansion Factor". In: *Proc. CRYPTO*. 2002, pp. 581–596. DOI: 10.1007/3-540-45708-9˙37.

[Don+19] J. Don, S. Fehr, C. Majenz, and C. Schaffner. "Security of the Fiat-Shamir Transformation in the Quantum Random-Oracle Model". In: *Proc. CRYPTO*. Lecture Notes in Computer Science. Springer International Publishing, 2019, pp. 356–383. DOI: 10.1007/978-3-030-26951-7˙13.

[DRV12] Y. Dodis, T. Ristenpart, and S. Vadhan. "Randomness Condensers for Efficiently Samplable, Seed-Dependent Sources". In: *Proc. TCC*. 2012, pp. 618–635. DOI: 10.1007/978-3-642-28914-9˙35.

[Fis05] M. Fischlin. "Communication-Efficient Non-interactive Proofs of Knowledge with Online Extractors". In: *Proc. CRYPTO*. 2005, pp. 152–168. DOI: 10.1007/11535218˙10.

[FS86] A. Fiat and A. Shamir. "How To Prove Yourself: Practical Solutions to Identification and Signature Problems". In: *Proc. CRYPTO*. 1986, pp. 186–194. DOI: 10.1007/3-540-47721-7˙12.

[GK03] S. Goldwasser and Y. T. Kalai. "On the (In)Security of the Fiat-Shamir Paradigm". In: *Proc. FOCS*. IEEE Computer Society, 2003, pp. 102–. DOI: 10.1109/SFCS.2003.1238185.

[GO94] O. Goldreich and Y. Oren. "Definitions and Properties of Zero-knowledge Proof Systems". In: *J. Cryptol.* 7.1 (Dec. 1994), pp. 1–32. DOI: 10.1007/BF00195207.

[GW11] C. Gentry and D. Wichs. "Separating Succinct Non-interactive Arguments from All Falsifiable Assumptions". In: *Proc. STOC*. 2011, pp. 99–108. DOI: 10.1145/1993636.1993651.

[HH09] I. Haitner and T. Holenstein. "On the (Im)Possibility of Key Dependent Encryption". In: *Proc. TCC*. 2009, pp. 202–219. DOI: 10.1007/978-3-642-00457-5˙13.

[HY20] A. Hosoyamada and T. Yamakawa. "Finding Collisions in a Quantum World: Quantum Black-Box Separation of Collision-Resistance and One-Wayness". In: *Advances in Cryptology – ASIACRYPT 2020*. Lecture Notes in Computer Science. Springer International Publishing, 2020, pp. 3–32. DOI: 10.1007/978-3-030-64837-4˙1.

[IR89] R. Impagliazzo and S. Rudich. "Limits on the Provable Consequences of One-Way Permutations". In: *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washington, USA*. 1989, pp. 44–61. DOI: 10.1145/73007.73012.

[KR04] A. Klappenecker and M. Rötteler. “Constructions of Mutually Unbiased Bases”. In: *Finite Fields and Applications*. 2004, pp. 137–144. DOI: 10.1007/978-3-540-24633-6˙10.

[KRR17] Y. T. Kalai, G. N. Rothblum, and R. D. Rothblum. “From Obfuscation to the Security of Fiat-Shamir for Proofs”. In: *CRYPTO*. 2017, pp. 224–251. DOI: 10.1007/978-3-319-63715-0˙8.

[LZ19] Q. Liu and M. Zhandry. “Revisiting Post-quantum Fiat-Shamir”. In: *Proc. CRYPTO*. 2019, pp. 326–355. DOI: 10.1007/978-3-030-26951-7˙12.

[MNY24] T. Morimae, B. Nehoran, and T. Yamakawa. “Unconditionally Secure Commitments with Quantum Auxiliary Inputs”. In: *Advances in Cryptology – CRYPTO 2024*. Springer Nature Switzerland, 2024, pp. 59–92. DOI: 10.1007/978-3-031-68394-7˙3.

[MY22] T. Morimae and T. Yamakawa. “Classically Verifiable NIZK for QMA with Preprocessing”. In: *Advances in Cryptology – ASIACRYPT 2022*. Springer Nature Switzerland, 2022, pp. 599–627. DOI: 10.1007/978-3-031-22972-5˙21.

[Nao03] M. Naor. “On Cryptographic Assumptions and Challenges”. In: *Proc. CRYPTO 2003*. Vol. 2729. 2003, pp. 96–109. DOI: 10.1007/978-3-540-45146-4˙6.

[PR97] S. Popescu and D. Rohrlich. “Thermodynamics and the measure of entanglement”. In: *Phys. Rev. A* 56 (5 Nov. 1997), R3319–R3321. DOI: 10.1103/PhysRevA.56.R3319.

[PS19] C. Peikert and S. Shiehian. “Noninteractive Zero Knowledge for NP from (Plain) Learning with Errors”. In: *Proc. CRYPTO*. 2019, pp. 89–114. DOI: 10.1007/978-3-030-26948-7˙4.

[PS96] D. Pointcheval and J. Stern. “Security Proofs for Signature Schemes”. In: *Proc. EUROCRYPT*. 1996, pp. 387–398. DOI: 10.1007/3-540-68339-9˙33.

[Qia24] L. Qian. “Unconditionally Secure Quantum Commitments with Preprocessing”. In: *Advances in Cryptology – CRYPTO 2024*. Springer Nature Switzerland, 2024, pp. 38–58. DOI: 10.1007/978-3-031-68394-7˙2.

[Rob21] B. Roberts. “Security Analysis of Quantum Lightning”. In: *Advances in Cryptology – EUROCRYPT 2021*. Vol. 12697. LNCS. Springer-Verlag, 2021, pp. 562–567. DOI: 10.1007/978-3-030-77886-6˙19.

[RTV04] O. Reingold, L. Trevisan, and S. Vadhan. “Notions of Reducibility between Cryptographic Primitives”. In: *Theory of Cryptography*. Springer Berlin Heidelberg, 2004, pp. 1–20. DOI: 10.1007/978-3-540-24638-1˙1.

[Sch60] J. Schwinger. “Unitary Operator Bases”. In: *Proc. National Academy of Sciences* 46.4 (1960), pp. 570–579. DOI: 10.1073/pnas.46.4.570.

[Sch89] C.-P. Schnorr. “Efficient Identification and Signatures for Smart Cards”. In: *Proc. CRYPTO*. Vol. 435. 1989, pp. 239–252. DOI: 10.1007/0-387-34805-0˙22.

[Seu12] Y. Seurin. “On the Exact Security of Schnorr-Type Signatures in the Random Oracle Model”. In: *EUROCRYPT*. Vol. 7237. 2012, pp. 554–571. DOI: 10.1007/978-3-642-29011-4˙33.

[Wat03] J. Watrous. “PSPACE has constant-round quantum interactive proof systems”. In: *Theoretical Computer Science*. Algorithms in Quantum Information Processing 292.3 (Jan. 31, 2003), pp. 575–588. DOI: 10.1016/S0304-3975(01)00375-9.

[WF89] W. K. Wootters and B. D. Fields. “Optimal state-determination by mutually unbiased measurements”. In: *Annals of Physics* 191.2 (1989), pp. 363–381. DOI: 10.1016/0003-4916(89)90322-9.

[Wic13] D. Wichs. “Barriers in cryptography with weak, correlated and leaky sources”. In: *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*. ITCS ’13. Association for Computing Machinery, Jan. 9, 2013, pp. 111–126. DOI: 10.1145/2422436.2422451.

[Zha19] M. Zhandry. “Quantum Lightning Never Strikes the Same State Twice”. In: *Advances in Cryptology – EUROCRYPT 2019*. Lecture Notes in Computer Science. Springer International Publishing, 2019, pp. 408–438. DOI: 10.1007/978-3-030-17659-4˙14.

A Technical Lemmas for Theorem 9

Proof (Proof of Lemma 2) Applying the Chernoff bound with $D = 2^k$, $M = 2^n$, $\alpha = 2^{-m}$, and $X_a^f = \sum_w N_{f(a),w}^a$ results in

$$\begin{aligned}\Pr_{f\in\mathcal{F}}\left[2^{-n}\sum_a X_a^f \notin \left[(1-t\eta)\frac{\mathbb{1}_P}{2^m},(1+t\eta)\frac{\mathbb{1}_P}{2^m}\right]\right] &\leq 2^{k+1}\exp\left(-\frac{1}{2\ln 2}2^{n-m}t^2\eta^2\right)\\ &= 2^{k+1}\exp\left(-(n+k)t^2\right)\\ &\leq 2^{-n\cdot t^2}\ .\end{aligned}$$

The result then follows easily assuming $\frac{1}{2^n}\sum_a X_a^f \leq (1+t\eta)\frac{\mathbb{1}_P}{2^m}$,

$$\begin{aligned}\sum_{a,w} P_{a,w}^f &= \frac{\sum_{a,w} N_{f(a),w}^a}{2^{n-m}(1+\eta)}\\ &= \frac{2^{-n}\sum_a X_a^f}{2^{-m}(1+\eta)}\\ &\leq \frac{(1+t\eta)\frac{\mathbb{1}_P}{2^m}}{2^{-m}(1+\eta)}\\ &= \frac{(1+t\eta)\mathbb{1}_P}{1+\eta}\ .\end{aligned}$$

On the other hand, assuming $\frac{1}{2^n}\sum_a X_a^f \geq (1-t\eta)\frac{\mathbb{1}_P}{2^m}$,

$$\begin{aligned}\sum_{a,w} P_{a,w}^f &= \frac{\sum_{a,w} N_{f(a),w}^a}{2^{n-m}(1+\eta)}\\ &= \frac{2^{-n}\sum_a X_a^f}{2^{-m}(1+\eta)}\\ &\geq \frac{(1-t\eta)\frac{\mathbb{1}_P}{2^m}}{2^{-m}(1+\eta)}\\ &= \frac{(1-t\eta)\mathbb{1}_P}{1+\eta}\ .\end{aligned}$$

■

Proof (Proof of Lemma 3) As we did before, we set the dimension of the CRQS on P's side to be 2^k (which we should write $k(n)$ rather than k). Let

$$\Delta := \left\| \mathop{\mathbb{E}}_{f\in\mathcal{F}^*} \left[\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right] - \mathop{\mathbb{E}}_{f\in\mathcal{F}} \left[\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right] \right\|_1 .$$

Let $\rho_{\mathcal{F}^*} = \frac{1}{\#\mathcal{F}^*}\sum_{f\in\mathcal{F}^*} |f\rangle\langle f|$ and $\rho_{\mathcal{F}} = \frac{1}{\#\mathcal{F}}\sum_{f\in\mathcal{F}} |f\rangle\langle f|$. It is easy to see that

$$\begin{aligned}\|\rho_{\mathcal{F}^*} - \rho_{\mathcal{F}}\|_1 &= \sum_{f\in\mathcal{F}^*}\left(\frac{1}{|\mathcal{F}^*|} - \frac{1}{|\mathcal{F}|}\right) + \sum_{f\in\mathcal{F}-\mathcal{F}^*}\frac{1}{|\mathcal{F}|} \\ &= 1 - \Pr\left[f\in\mathcal{F}^*\right] + \Pr\left[f\notin\mathcal{F}^*\right] \\ &\leq \mathsf{negl}(n) \ ,\end{aligned}$$

after applying lemma 8. Then, we have

$$\begin{aligned}\Delta &\leq \|\rho_{\mathcal{F}^*} - \rho_{\mathcal{F}}\|_1 + \frac{1}{\#\mathcal{F}}\sum_{f\in\mathcal{F}-\mathcal{F}^*}\left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 \\ &\leq \mathsf{negl}(n) + \frac{1}{\#\mathcal{F}}\sum_{f\in\mathcal{F}-\mathcal{F}^*}\left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 \ . \end{aligned} \tag{29}$$

Bounding this sum is not as straightforward as it might look at first glance: while we know that the sum has very few terms, we have no guarantee that $\mathcal{A}_n^f$ is a physically realizable map when $f\in\mathcal{F}-\mathcal{F}^*$, and hence we cannot trivially bound these norms by 1. Instead, let us consider reduction $\mathcal{R}_n^{\mathcal{A}_n^f} := U_q(\mathcal{A}_n^f\otimes\mathbb{1})U_{q-1}\ldots(\mathcal{A}_n^f\otimes\mathbb{1})U_1(\mathcal{A}_n^f\otimes\mathbb{1})U_0$ where all $U_j, j\in\{0,\ldots,q\}$ are unitaries and the $\mathbb{1}$'s are acting on the wires that are not part of $\mathcal{A}_n^f$'s standard interface. Let $|\varphi_0^f\rangle := U_0|0\rangle$ be a normalized state and let $|\varphi_j^f\rangle := U_j(\mathcal{A}_n^f\otimes\mathbb{1})|\varphi_{j-1}^f\rangle$ for $1<j\leq q$, not necessarily of norm 1 when $f\notin\mathcal{F}^*$. Using lemma 2 with $t := 2^{\frac{n-m}{4}}$ (and $\eta = \sqrt{2\ln 2(n+k)\cdot 2^{m-n}}$ as in the statement of lemma 2), we have that

$$\Pr_{f\in_R\mathcal{F}}\left[\sum_{a,w} P_{a,w}^f \notin \left[\frac{(1-t\eta)\mathbb{1}_P}{1+\eta}, \frac{(1+t\eta)}{1+\eta}\mathbb{1}_P\right]\right] \leq 2^{-n\sqrt{2^{n-m}}} \ . \tag{30}$$

For $|\varphi\rangle$ a state of norm 1 and for $f\in\mathcal{F}$ such that $\sum_{a,w}P_{a,w}^f \leq \left(\frac{1+t\eta}{1+\eta}\right)\mathbb{1}_P \leq (1+t\eta)\mathbb{1}_P$, $\|(\mathcal{A}_n^f\otimes\mathbb{1})|\varphi\rangle\langle\varphi|(\mathcal{A}_n^f\otimes\mathbb{1})^*\|_1 \leq 1+t\eta$. Starting with a normalized state $|\varphi\rangle$, after $q(n)$ queries to $\mathcal{A}_n^f$, the square of the norm of the resulting vector is upper bounded by $(1+t\eta)^{q(n)}$. Notice that when $t\eta = p(n)2^{-\beta n}$ for $p(n)$ a polynomial and $\beta>0$,

$$(1+t\eta)^{q(n)} = \left(1+\frac{p(n)}{2^{\beta n}}\right)^{q(n)} = \left(1+\frac{p(n)}{2^{\beta n}}\right)^{\frac{2^{\beta n}q(n)p(n)}{p(n)2^{\beta n}}} \approx \exp\left(\frac{q(n)p(n)}{2^{\beta n}}\right) \ ,$$

since $\lim_{N\to\infty}(1+\frac{1}{N})^N = e$. In other words, when $t\leq 2^{\frac{n-m}{4}}$, we have $t\eta = p(n)2^{-\beta n}$ and[12],

$$\lim_{n\to\infty}\left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 = \exp\left(\frac{q(n)p(n)}{2^{\beta n}}\right) \leq 1+\mathsf{negl}(n) \ . \tag{31}$$

[12]Using the fact that $1+2^{-x+1} > \exp(2^{-x})$ for all $x\geq 0$.

Therefore, for $t \leq 2^{\frac{n-m}{4}}$ and f such that $\sum_{a,w} P^f_{a,w} \leq (1+t\eta)\,\mathbb{1}_P$, we have that $\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)$ *essentially preserves* norms like an isometry: $\left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 = 1$. For handling the other case ($t > 2^{\frac{n-m}{4}}$), we first define

$$\mathcal{F}_t := \left\{ f \in \mathcal{F} \;\middle|\; \frac{(1-t\eta)\mathbb{1}_P}{1+\eta} \leq \sum_{a,w} P^f_{a,w} \leq \frac{(1+t\eta)\mathbb{1}_P}{1+\eta} \right\} \;,$$

and notice that by construction, for every $f \in \mathcal{F}$,

$$\sum_{a,w} P^f_{a,w} = \frac{\sum_{a,w} N^a_{f(a),w}}{2^{n-m}(1+\eta)} \leq \frac{\sum_a \mathbb{1}_P}{2^{n-m}(1+\eta)} = \frac{2^n\mathbb{1}_P}{2^{n-m}(1+\eta)} \leq \frac{2^m\mathbb{1}_P}{1+\eta} \;. \tag{32}$$

Let $t := 2^{\frac{n-m}{4}}$ and note that $t < \frac{1}{2\eta}$ so that the Chernoff bound expressed in lemma 2 can be used. We consider two cases for $f \in \mathcal{F} - \mathcal{F}^*$: either f is in $\mathcal{F}_t - \mathcal{F}^*$, or f is outside $\mathcal{F}_t \cup \mathcal{F}^*$. We have,

$$\begin{aligned}
\frac{1}{\#\mathcal{F}} \sum_{f\in\mathcal{F}-\mathcal{F}^*} \left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 &= \frac{1}{\#\mathcal{F}} \left(\sum_{f\in\mathcal{F}_t-\mathcal{F}^*} \left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 \right. \\
&\qquad \left. + \sum_{f\notin\mathcal{F}_t\cup\mathcal{F}^*} \left\|\mathcal{R}_n^{\mathcal{A}_n^f}(|0\rangle\langle 0|)\right\|_1 \right) \\
&\leq \Pr\left[F \in \mathcal{F}_t - \mathcal{F}^*\right] \cdot (1+t\eta)^{q(n)} \\
&\qquad + \Pr\left[F \notin \mathcal{F}_t \cup \mathcal{F}^*\right] \cdot 2^{mq(n)} && (33) \\
&\leq \mathsf{negl}(n)(1+\mathsf{negl}(n)) + 2^{-n\sqrt{2^{n-m}}+mq(n)} && (34) \\
&\leq \mathsf{negl}(n) \;, && (35)
\end{aligned}$$

as long as $n > m$, where (33) follows from (32) and (34) follows from (31) and the Chernoff bound, as stated in (30).

Finally, using (35) in (29) proves the statement. ■

Lemma 4 *(A) is negligible.*

Proof Since Sim_n picks $a \in \{0,1\}^n$ uniformly at random, we get

$$\begin{aligned}
\sum_{\substack{a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} q^{\mathsf{Sim}_n}_{a,z,w}(\psi_j(s)) &= \sum_{\substack{a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} 2^{-n}\,\mathrm{tr}\left(N^a_{z,w}|\psi_j(s)\rangle\langle\psi_j(s)|\right) \\
&= \sum_{a\in\delta(s)} 2^{-n}\,\mathrm{tr}\left(\sum_{z,w} N^a_{z,w}|\psi_j(s)\rangle\langle\psi_j(s)|\right) \\
&\leq \sum_{a\in\delta(s)} 2^{-n} \\
&\leq q(n)2^{-n} \\
&\leq \mathsf{negl}(n) \;. && (36)
\end{aligned}$$

A similar argument can be applied to $q^{\mathcal{A}_n^f}_{a,w}(\psi_j(s))$ although $\{P^f_{a,w}\}_{a,w}$ is a collection of positive operators that do not form a valid POVM when $f \notin \mathcal{F}^*$, as $\sum_{a,w} P^f_{a,w} \not\leq \mathbb{1}_P$ in

this case. We have,

$$\begin{aligned}
\frac{1}{\#\mathcal{F}} \sum_{\substack{f\in\mathcal{F}\\ a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} q_{a,w}^{\mathcal{A}_n^f}(\psi_j(s)) &= \frac{1}{\#\mathcal{F}} \sum_{\substack{a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} \sum_{\substack{f\in\mathcal{F}\\ f(a)=z}} \mathrm{tr}\left(P_{a,w}^f |\psi_j(s)\rangle\langle\psi_j(s)|\right) \\
&= \frac{1}{\#\mathcal{F}} \sum_{\substack{a\in\delta(s)\\ z\in\{0,1\}^m\\ w\in\{0,1\}^{\ell(n)}}} \sum_{\substack{f\in\mathcal{F}\\ f(a)=z}} \mathrm{tr}\left(\frac{N_{z,w}^a |\psi_j(s)\rangle\langle\psi_j(s)|}{2^{n-m}+\sqrt{2\ln 2(n+k)2^{n-m}}}\right) \\
&= \frac{1}{\#\mathcal{F}} \sum_{a\in\delta(s)} \sum_{\substack{f\in\mathcal{F}\\ f(a)=z}} \mathrm{tr}\left(\frac{\sum_{z,w} N_{z,w}^a |\psi_j(s)\rangle\langle\psi_j(s)|}{2^{n-m}+\sqrt{2\ln 2(n+k)2^{n-m}}}\right) \\
&\leq \frac{1}{\#\mathcal{F}} \sum_{a\in\delta(s)} \sum_{\substack{f\in\mathcal{F}\\ f(a)=z}} \frac{1}{2^{n-m}+\sqrt{2\ln 2(n-m+k)2^{n-m}}} \\
&\leq \left(\sum_{\substack{f\in\mathcal{F}\\ f(a)=z}} \frac{1}{\#\mathcal{F}}\right) \frac{q(n)}{2^{n-m}+\sqrt{2\ln 2(n-m+k)2^{n-m}}} \\
&\leq \mathsf{negl}(n-m) \ . \qquad (37)
\end{aligned}$$

We now conclude,

$$(A) \leq (36) + (37) \leq \mathsf{negl}(n-m) \ . \qquad (38)$$

■

Lemma 5 *(⊥) is negligible.*

Proof This corresponds to the likelihood (and not *the probability*, as $\{P_{a,w}^f\}_{a,w}$ is not a valid POVM when $f \notin \mathcal{F}^*$) of an error when $\mathcal{A}_n^f$ is queried once on $|\psi_j(s)\rangle$, a normalized state vector. Observe that when $f \in \mathcal{F}_t$, $\sum_{a,w} P_{a,w}^f \geq (1-t\eta)\mathbb{1}_P$ and therefore $0 \leq P_\perp^f \leq t\eta\,\mathbb{1}_P$. In other words, the probability to get an error when $f \in \mathcal{F}_t$ is upper bounded by $t\eta$. On the other hand, when $f \notin \mathcal{F}_t$, the only thing we can say from our construction is that $0 \leq P_\perp^f \leq \mathbb{1}_P$. In the following and as before, we set $t := 2^{\frac{n-m}{4}}$ and $\eta := \sqrt{2\ln 2(n+k)2^{m-n}}$. Using the operator Chernoff bound expressed in lemma 2, we have

$$\begin{aligned}
(\perp) &\leq \frac{1}{\#\mathcal{F}} \left(\sum_{f\in\mathcal{F}_t} q_\perp^{\mathcal{A}_n^f}(\psi_j(s)) + \sum_{f\notin\mathcal{F}_t} q_\perp^{\mathcal{A}_n^f}(\psi_j(s))\right) \\
&= \frac{1}{\#\mathcal{F}} \left(\sum_{f\in\mathcal{F}_t} \mathrm{tr}\left(P_\perp^f |\psi_j(s)\rangle\langle\psi_j(s)|\right) + \sum_{f\notin\mathcal{F}_t} \mathrm{tr}\left(P_\perp^f |\psi_j(s)\rangle\langle\psi_j(s)|\right)\right) \\
&\leq t\eta + \Pr\left[f \notin \mathcal{F}_t\right] \\
&\leq \sqrt{2\ln 2(n+k)2^{\frac{-n+m}{2}}} + 2^{-n\sqrt{2^{n-m}}} \\
&\leq \mathsf{negl}(n-m) \ . \qquad (39)
\end{aligned}$$

■

Lemma 6 *(M) is negligible.*

PROOF This the main part to show that the Chernoff adversary $\mathcal{A}_n^{\mathcal{F}}$ is simulatable. This is where we use the fact that Sim_n *simulates* the adversary $\mathcal{A}_n^f$ whenever $f \in_R \mathcal{F}$ (in real life this is for $f \in_R \mathcal{F}^*$). It follows essentially the same steps as in [Bit+13; BGW12] adapted to deal with our Chernoff adversary. For all $f \in \mathcal{F}$ such that $f(a) = y$, we have

$$
\begin{aligned}
(M) &= \sum_{\substack{a\in\{0,1\}^n\setminus\delta(s)\\ w\in\{0,1\}^{\ell(n)}\\ y\in\{0,1\}^m}} \frac{2^m}{\#\mathcal{F}} \sum_{\substack{f\in\mathcal{F}\\ f(a)=y}} \left| q^{\mathsf{Sim}_n}_{a,y,w}(\psi_j(s)) - 2^{-m} q^{\mathcal{A}^f_n}_{a,w}(\psi_j(s)) \right| \\
&= \sum_{\substack{a\in\{0,1\}^n\setminus\delta(s)\\ w\in\{0,1\}^{\ell(n)}\\ y\in\{0,1\}^m}} \left(\sum_{\substack{f\in\mathcal{F}\\ f(a)=y}} \frac{2^m}{\#\mathcal{F}} \right) \left| \operatorname{tr}\left(\frac{N^a_{y,w}}{2^n} - \frac{2^{-m} N^a_{y,w}}{\frac{2^n}{2^m} + \sqrt{2\ln 2(n+k)\frac{2^n}{2^m}}} |\psi_j(s)\rangle\langle\psi_j(s)| \right) \right| \\
&= \sum_{a\in\{0,1\}^n\setminus\delta(s)} \left| \frac{1}{2^n} - \frac{2^{-m}}{\frac{2^n}{2^m} + \sqrt{2\ln 2(n+k)\frac{2^n}{2^m}}} \right| \sum_{y,w} \operatorname{tr}\left(N^a_{y,w} |\psi_j(s)\rangle\langle\psi_j(s)| \right) \\
&\leq \sum_{a\in\{0,1\}^n\setminus\delta(s)} \left| \frac{1}{2^n} - \frac{2^{-m}}{\frac{2^n}{2^m} + \sqrt{2\ln 2(n+k)\frac{2^n}{2^m}}} \right| \\
&\leq \left| 1 - \frac{2^{-m+n}}{\frac{2^n}{2^m} + \sqrt{2\ln 2(n+k)\frac{2^n}{2^m}}} \right| \\
&\leq \sqrt{2\ln 2(n+k)2^{m-n}} \\
&\leq \mathsf{negl}(n-m) \ . \qquad (40)
\end{aligned}
$$

■

B Technical Lemmas for Theorem 11

We now proceed to compute the $\operatorname{tr}(S)$, $\operatorname{tr}(S^2)$ and $\operatorname{tr}(S^3)$ values used in the proof of Theorem 11.

Lemma 7 $\operatorname{tr}(S) = p^{3n}$.

PROOF Since $|\mathcal{B}(a)| = p^{2n}$,

$$\operatorname{tr}(S) = \sum_{a\in\Gamma^n} \sum_{x\in\mathcal{B}(a)} \operatorname{tr}\left(|x\rangle\langle x|_a\right) = p^{3n} \ .$$

■

Lemma 8 $\operatorname{tr}(S^2) = 2p^{3n} - p^{2n}$.

PROOF

$$\operatorname{tr}\left(S^2\right) = \sum_{a,b\in\Gamma^n} \sum_{x\in\mathcal{B}(a), y\in\mathcal{B}(b)} \operatorname{tr}(|x\rangle\langle x|_a |y\rangle\langle y|_b)$$

$$
\begin{aligned}
&= \sum_{a\in\Gamma^n}\left(\sum_{x\in\mathcal{B}(a)} 1 + \sum_{b\neq a}\sum_{x\in\mathcal{B}(a),y\in\mathcal{B}(b)} |\langle x|_a|y\rangle_b|^2\right)\\
&= \sum_{a\in\Gamma^n}\left(|\mathcal{B}(a)| + p^{-3n}\sum_{b\neq a}|\mathcal{B}(a)|\cdot|\mathcal{B}(b)|\right)\\
&= \sum_{a\in\Gamma^n}\left(p^{2n} + p^{-3n}\sum_{b\neq a}p^{4n}\right)\\
&= \sum_{a\in\Gamma^n}\left(p^{2n} + p^{n}\sum_{b\neq a}1\right)\\
&= p^n\left(p^{2n} + p^n(p^n-1)\right)\\
&= 2p^{3n} - p^{2n}
\end{aligned}
$$

■

Upper-bounding $\mathrm{tr}(S^3)$ will require a little more machinery. We introduce a theorem of Deligne [Del74] and some of its corollaries before proceeding with the proof.

Theorem 16 ([Del74], Theorem 8.4) *Let Q be a polynomial of n variables $x_1,\ldots,x_n$ and of degree d on $\mathbb{F}_q$, let Q_d be the homogeneous part of degree d of Q and let $\psi:\mathbb{F}_q\to\mathbb{C}^*$ be an additive non-trivial character on $\mathbb{F}_q$. Assume that*

1. *d is coprime with p, the characteristic of $\mathbb{F}_q$, and*
2. *the hypersurface H_0 of $\mathbb{P}^{n-1}_{\mathbb{F}_q}$ defined by Q_d is smooth,*

then

$$
\left|\sum_{x_1,\ldots,x_n\in\mathbb{F}_q}\psi\left(Q(x_1,\ldots,x_n)\right)\right| \leqslant (d-1)^n q^{n/2} \ .
$$

In the above, the second condition boils down to ensuring that there is no point at which the $\frac{\partial Q}{\partial x_i}$ all vanish simultaneously. Here is a version that is closer to what we will need:

Corollary 3 *Let $m\leq k$, A a $k\times m$ matrix with rank m in $\mathbb{F}_q$, and let C be a $k\times k$ matrix in $\mathbb{F}_q$. Then, if $A^\intercal CA$ is non-singular,*

$$
\left|\sum_{\vec{v},\vec{x}=A\vec{v}}\psi\left(\vec{x}^\intercal C\vec{x}\right)\right| \leqslant q^{m/2} \ .
$$

In other words, we take the sum over all $(x_1,\ldots,x_k)$ that satisfy a system of $k-m$ independent linear equations.

Proof Let $Q=\vec{x}^\intercal C\vec{x}=\vec{v}^\intercal A^\intercal CA\vec{v}$, and observe that

$$
\frac{\partial Q}{\partial v_i} = e_i^\intercal A^\intercal CA\vec{v} + \vec{v}^\intercal A^\intercal CAe_i = 2e_i^\intercal A^\intercal CA\vec{v} \ .
$$

Condition 2 of Theorem 16 is thus equivalent to

$$
A^\intercal CA\vec{v}=0 \Leftrightarrow \vec{v}=0 \ ,
$$

which amounts to saying that $A^\intercal CA$ is non-singular. ■

Here is now a version that is more directly relevant to our case.

Corollary 4 *Let $m \leq k$ and let $B \in \mathbb{F}_q^{(k-m)\times k}$ and $C \in \mathbb{F}_q^{k\times k}$ be full rank matrices. Then,*

$$\left| \sum_{\vec{x}: B\vec{x}=0} \psi\left(\vec{x}^\intercal C \vec{x}\right) \right| \leqslant q^{m/2} \ .$$

Proof Let $B^c \in \mathbb{F}_q^{m\times k}$ such that $M := \begin{bmatrix} B \\ B^c \end{bmatrix} \in \mathbb{F}_q^{k\times k}$ has full rank. Then condition $B\vec{x} = 0$ is equivalent to $\vec{x} = M^{-1}\begin{bmatrix} 0 \\ \vec{v} \end{bmatrix}$ for some $\vec{v} \in \mathbb{F}_q^m$. We can thus define $P := \begin{bmatrix} 0 \\ \mathbb{1} \end{bmatrix}$ and apply corollary 3 with $A = M^{-1}P$, while observing that $P^\intercal M^{-1\intercal} C M^{-1} P$ has full rank, since $M^{-1\intercal} C M^{-1}$ also has full rank. ■

Lemma 9 $\mathrm{tr}(S^3) \leq 4p^{3n} + p^{2n}$

Proof Let's first write out the expression of interest:

$$\begin{aligned} \mathrm{tr}\left(S^3\right) &= \sum_{a,b,c\in\Gamma^n} \sum_{x\in\mathcal{B}(a)} \sum_{y\in\mathcal{B}(b)} \sum_{z\in\mathcal{B}(c)} \mathrm{tr}\left(|x\rangle\langle x|_a |y\rangle\langle y|_b |z\rangle\langle z|_c\right) \\ &= \sum_{a=b=c} \sum_{x} 1 + 3 \sum_{a\neq b} \sum_{x,y} |\langle x|_a |y\rangle_b|^2 + \sum_{a\neq b\neq c} \sum_{x,y,z} \langle x|_a |y\rangle_b \langle y|_b |z\rangle_c \langle z|_c |x\rangle_a \end{aligned} \tag{41}$$

where the middle term groups the three cases $a \neq b$, $a \neq c$ and $b \neq c$ that all have the same value. We know how to upper-bound the first two sums using the same techniques as Lemma 8. Most of the proof is dedicated to finding an upper-bound to the third term.

Recall our construction of mutually unbiased bases θ_a presented in Definition 14. For $r \in \mathbb{F}_{p^n}$ and $a \in \mathbb{F}_{p^n}$:

$$|r\rangle_a = p^{-\frac{n}{2}} \sum_{u\in\mathbb{F}_{p^n}} \exp\left(\frac{2\pi i}{p} \cdot \mathrm{tr}(au^2 + ru)\right) |u\rangle \ .$$

Extending this basis to 3 systems through $\theta_a^{\otimes 3}$ yields vectors of the form

$$|x\rangle_a = p^{-3n/2} \sum_{u\in\mathbb{F}_{p^n}^3} \exp\left(\frac{2\pi i}{p} \mathrm{tr}\left(au^\intercal u + x^\intercal u\right)\right) |u\rangle \ ,$$

where $x^\intercal$ denotes the transpose of $x \in \mathbb{F}_{p^n}^3 \simeq \Gamma^{3n}$. Here, we slightly abuse notation by writing $|x\rangle_a$ for a vector in basis $\theta_a^{\otimes 3}$.

The inner product of two such vectors is given by the expression

$$\langle y|_b |x\rangle_a = p^{-3n} \sum_{u\in\mathbb{F}_{p^n}^3} \exp\left(\frac{2\pi i}{p} \mathrm{tr}\left((a-b)u^\intercal u + (x-y)^\intercal u\right)\right) \ .$$

Combining the three inner products in the expression of interest (41), we have

$$\langle x|_a |y\rangle_b \langle y|_b |z\rangle_c \langle z|_c |x\rangle_a = p^{-9n} \sum_{u,v,w\in\mathbb{F}_{p^n}^3} \exp\left(\frac{2\pi i}{p} \mathrm{tr}\begin{pmatrix} (a-b)u^\intercal u + (x-y)^\intercal u \\ +(b-c)v^\intercal v + (y-z)^\intercal v \\ +(c-a)w^\intercal w + (z-x)^\intercal w \end{pmatrix}\right)$$

We introduce some notation that will allow us to present the above expression in a more compact, albeit more complicated form. Let $\mathsf{c} : \mathbb{F}_{p^n} \to \mathbb{F}_{p^m}$ and for $a \in \mathbb{F}_{p^n}$, define

$$B_a = \begin{bmatrix} 1 & 0 & \mathsf{c}(a) \\ 0 & 1 & \mathsf{c}(a) \end{bmatrix} \in \mathbb{F}_{p^n}^{2\times 3} \tag{42}$$

such that for $x_1, x_2 \in \mathbb{F}_{p^n}$, the expression

$$[x_1, x_2] \cdot B_a{}^{\mathsf{T}} = [x_1, x_2, \mathsf{c}(a)(x_1 + x_2)]^{\mathsf{T}} \in \mathbb{F}_{p^n}^3 \tag{43}$$

is a sequence of measurement outcomes that leads to the bad outcome $\mathsf{c}(a)$ in the protocol.

For $a, b, c \in \mathbb{F}_{p^n}$, write

$$B_{a,b,c} := \begin{bmatrix} -B_a & 0 & B_a \\ 0 & B_b & -B_b \\ B_c & -B_c & 0 \end{bmatrix} \in \mathbb{F}_{p^n}^{6\times 9} \ ,$$

and

$$C_{a,b,c} := \begin{bmatrix} (c-a)\mathbb{1}_{\mathbb{F}_{p^n}^{3\times 3}} & 0 & 0 \\ 0 & (b-c)\mathbb{1}_{\mathbb{F}_{p^n}^{3\times 3}} & 0 \\ 0 & 0 & (a-b)\mathbb{1}_{\mathbb{F}_{p^n}^{3\times 3}} \end{bmatrix} \in \mathbb{F}_{p^n}^{9\times 9} \ . \tag{44}$$

The previous operators are defined such that

$$\begin{aligned} &\sum_{\substack{x\in\mathcal{B}(a)\\ y\in\mathcal{B}(b)\\ z\in\mathcal{B}(c)}} \langle x|_a |y\rangle_b \langle y|_b |z\rangle_c \langle z|_c |x\rangle_a \\ &\qquad = p^{-9n} \sum_{\varrho\in\mathbb{F}_{p^n}^6} \sum_{\xi\in\mathbb{F}_{p^n}^9} \exp\left(\frac{2\pi i}{p} \operatorname{tr}\left(\xi^{\mathsf{T}} C_{a,b,c} \xi + \varrho^{\mathsf{T}} B_{a,b,c} \xi\right)\right) \end{aligned}$$

with the goal of bounding above the right-hand side using Corollary 4. The construction of $B_{a,b,c}$ appears more complex than necessary because we want it to have a large rank.

Equipped with the above, we are now ready to upper-bound the third term in (41) with Corollary 4.

$$\begin{aligned} &\sum_{a\neq b\neq c} \sum_{\substack{x\in\mathcal{B}(a)\\ y\in\mathcal{B}(b)\\ z\in\mathcal{B}(c)}} \langle x|_a |y\rangle_b \langle y|_b |z\rangle_c \langle z|_c |x\rangle_a \\ &= p^{-9n} \sum_{a\neq b\neq c} \sum_{\varrho\in\mathbb{F}_{p^n}^6} \sum_{\xi\in\mathbb{F}_{p^n}^{9n}} \exp\left(\frac{2\pi i}{p} \operatorname{tr}\left(\xi^{\mathsf{T}} \cdot C_{a,b,c} \cdot \xi + \varrho^{\mathsf{T}} \cdot B_{a,b,c} \cdot \xi\right)\right) \\ &= p^{-9n} \sum_{a\neq b\neq c} \sum_{\varrho\in\mathbb{F}_{p^n}^6} \sum_{\substack{\xi\in\mathbb{F}_{p^n}^{9n}\\ B_{a,b,c}\cdot\xi=0}} \exp\left(\frac{2\pi i}{p} \operatorname{tr}\left(\xi^{\mathsf{T}} \cdot C_{a,b,c} \cdot \xi\right)\right) \qquad (45) \\ &\leq p^{-9n} \sum_{a\neq b\neq c} \sum_{\varrho\in\mathbb{F}_{p^n}^6} p^{2n} \qquad (46) \\ &= p^{-9n} \sum_{a\neq b\neq c} p^{6n} p^{2n} \\ &= p^{-n}(p^n)(p^n-1)(p^n-2) \ . \end{aligned}$$

Equality (45) above follows from the observation that once ξ is fixed, if $B_{a,b,c} \cdot \xi$ is non-zero then the sum over ϱ will span all pth roots of unity in equal proportions which sums to 0. In more details, letting $\alpha = \xi^\intercal \cdot C_{a,b,c} \cdot \xi \in \mathbb{F}_{p^n}$ and $0 \neq v = B_{a,b,c} \cdot \xi \in \mathbb{F}_{p^n}^6$,

$$\begin{aligned}\sum_{\varrho \in \mathbb{F}_{p^n}^6} \exp\left(\frac{2\pi i}{p} \operatorname{tr}\left(\alpha + \varrho^\intercal \cdot v\right)\right) &= p^{5n} \sum_{\beta \in \mathbb{F}_{p^n}} \exp\left(\frac{2\pi i}{p} \operatorname{tr}\left(\alpha + \beta\right)\right) \\ &= p^{6n-1} \sum_{\gamma \in \mathbb{F}_p} \exp\left(\frac{2\pi i}{p}\gamma\right) = 0 \ .\end{aligned}$$

Inequality (46) follows from Corollary 4 by observing that $\operatorname{rank}(B_{a,b,c}) \geq 4$. To see this, note that by removing columns 3, 6 and 9 from $B_{a,b,c}$ (those corresponding to $\mathsf{c}(a), \mathsf{c}(b)$ or $\mathsf{c}(c)$), we are left with the matrix

$$\begin{bmatrix} \mathbb{1} & 0 & -\mathbb{1} \\ 0 & \mathbb{1} & -\mathbb{1} \\ \mathbb{1} & -\mathbb{1} & 0 \end{bmatrix} \ .$$

Taking linear combinations of the above we can obtain

$$\begin{bmatrix} \mathbb{1} & 0 & -\mathbb{1} \\ 0 & \mathbb{1} & -\mathbb{1} \\ 0 & 0 & 0 \end{bmatrix}$$

and hence $B_{a,b,c}$ has rank at least that of the above matrix, which is equal to 4 since each of the identities act on $\mathbb{F}_{p^n}^2$.

We can now complete the proof by taking the expected value over g. Continuing from (41),

$$\begin{aligned}\operatorname{tr}\left(S^3\right) &= \sum_{a=b=c} \sum_{x \in \mathcal{B}(a)} 1 + 3 \sum_{a \neq b} \sum_{\substack{x \in \mathcal{B}(a) \\ y \in \mathcal{B}(b)}} |\langle x|_a |y\rangle_b|^2 + \sum_{a \neq b \neq c} \sum_{\substack{x \in \mathcal{B}(a) \\ y \in \mathcal{B}(b) \\ z \in \mathcal{B}(c)}} \langle x|_a |y\rangle_b \langle y|_b |z\rangle_c \langle z|_c |x\rangle_a \\ &\leq p^{3n} + 3p^{2n}(p^n - 1) + (p^n - 1)(p^n - 2) \leq 4p^{3n} + p^{2n} \ .\end{aligned}$$

■

C Basic Properties of WOTRO

Proposition 1 *The 2-message protocol in which* P *sends* $a \in \Gamma^n$ *directly to* V*, and* V *then chooses* $c \in_R \Gamma^m$ *at random, sends it to* P *and always accepts is a correct and* δ*–secure implementation of* $\mathsf{WOTRO}_\Gamma^{n,m}$ *for* $\delta = 1 - \frac{1}{|\Gamma|^m}$ *and for any alphabet* Γ *and* $n, m \geqslant 1$*.*

Proof Let $\Pi(c|a)$ denote the conditional distribution of the protocol output. Indeed, correctness is obvious as a and c are correctly distributed with $\Pi(c|a) = \frac{1}{\#\Gamma^m}$. For security, let A be the random variable produced by $\tilde{\mathsf{P}}$ and C be the random variable produced by V, and let $\mathsf{c} : \Gamma^n \to \Gamma^m$ be some function. Then,

$$\begin{aligned}\Pr\left[V = 1 \wedge C = \mathsf{c}(A)\right] &= \Pr\left[C = \mathsf{c}(A)\right] \\ &= \frac{1}{|\Gamma|^m} \ .\end{aligned}$$

■

Proposition 2 *Let Γ be an arbitrary finite alphabet, let $m, n \geq 1$ and let $0 < \delta \leq 1$. There is no correct and δ–secure 1-message implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$ in the bare model. Moreover, there is no such δ–secure non-interactive implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$ common random string (resp. random oracle) model if the function c from Definition 5 can depend on the CR\$ r (resp. the random oracle $\mathcal{O}$).*

Proof Consider the message sent from the prover P to the verifier V. Without loss of generality, it is of the form (a, c, w) where a is P's input, c is the joint output and w is additional information for V to decide whether to accept or reject. Let $\Pi = (\mathsf{P}, \mathsf{V})$ (resp. $\Pi^r = (\mathsf{P}^r, \mathsf{V}^r)$ and $\Pi^{\mathcal{O}} = (\mathsf{P}^{\mathcal{O}}, \mathsf{V}^{\mathcal{O}})$) be a correct implementation of WOTRO in the bare model (resp. CR\$ model and ROM). Define the first message of the prover in each model by

$$P(a, s) := (a, c(a, s), w(a, s)) \qquad \text{(bare)}$$
$$P^r(a, s) := (a, c^r(a, s, r), w^r(a, s, r)) \qquad \text{(CR\$)}$$
$$P^{\mathcal{O}}(a, s) := (a, c^{\mathcal{O}}(a, s, v), w^{\mathcal{O}}(a, s, v)) \qquad \text{(ROM)}$$

where s is the random tape of the prover, r is the value of the CR\$ and $v = (\mathcal{O}(a_1), \mathcal{O}(a_2), \ldots, \mathcal{O}(a_{\kappa(n)}))$ where $a_1, \ldots, a_{\kappa(n)} \in \Gamma^n$ are chosen using s for some upper bound $\kappa(n)$ on the number of oracle queries performed by P in $\Pi^{\mathcal{O}}$.

Since the protocol is correct, it must hold that

$$\Pr[V(P(A, S)) = 1] = \Pr[V^r(P^r(A, S)) = 1] = \Pr[V^{\mathcal{O}}(P^{\mathcal{O}}(A, S)) = 1] = 1 \qquad (47)$$

where the probability is taken over the values of A and S. Then for each a with non zero probability, there exist a value $s(a)$, $s^r(a)$ and $s^{\mathcal{O}}(a)$ such that

$$V(P(a, s(a))) = V^r(P^r(a, s^r(a))) = V^{\mathcal{O}}(P^{\mathcal{O}}(a, s^{\mathcal{O}}(a))) = 1 \ . \qquad (48)$$

Define malicious prover $\tilde{\mathsf{P}}$ (resp. $\tilde{\mathsf{P}}^r$ and $\tilde{\mathsf{P}}^{\mathcal{O}}$) that on input a uses random tape value $s(a)$ (resp. $s^r(a)$ and $s^{\mathcal{O}}(a)$). Then the protocol Π (resp. Π^r and $\Pi^{\mathcal{O}}$) does not avoid the functions $\mathsf{c}(a) := c(a, s(a))$ (resp. $\mathsf{c}^r(a) := c^r(a, s^r(a), r)$ and $\mathsf{c}^{\mathcal{O}}(a) := c^{\mathcal{O}}(a, s^{\mathcal{O}}(a), v)$). ■

Proposition 3 *Let $m > n$. The protocol for $\mathsf{WOTRO}_\Gamma^{n,m}$ in the CR\$ model where both parties output the CR\$ $r \in \Gamma^m$ for any $a \in \Gamma^n$ and V always accepts is correct and δ–secure, for $\delta = 1 - |\Gamma|^{n-m}$.*

Proof Correctness is obvious, and security is easy to prove as well: suppose that $\tilde{\mathsf{P}}$ wants to steer the output of the protocol towards some function c. He must then look at the CR\$ r, and announce an a such that $\mathsf{c}(a) = r$. Hence, r must happen to be in the image of c. However, since c is a function from Γ^n to Γ^m and $m > n$, there are at most $|\Gamma|^n$ strings in the image of c, and the probability that a uniformly chosen r falls into that set is at most $|\Gamma|^{n-m}$. ■

Proposition 4 *Let Γ be an arbitrary finite alphabet of size $q \geq 2$. Then, for any m, n with $m \leq n$, there exists no $\exp(-q^{n-m})$–secure implementation of $\mathsf{WOTRO}_\Gamma^{n,m}$ in the ROM.*

Proof We will show that a cheating prover that is unbounded in time can search for an a that will satisfy V. Consider a dishonest prover $\tilde{\mathsf{P}}$ who uses the following strategy: run the honest prover P on all possible inputs a in lexicographic order, and declare victory

if it ever outputs $\mathsf{c}(a)$. We will also assume that function $\mathsf{c}(\cdot)$ is chosen uniformly at random, and show that the expected winning probability of the cheating prover is at least $1 - \exp(-q^{n-m})$. We have the following:

$$\begin{aligned}
\Pr_{\mathcal{O},\mathsf{c}}[\tilde{\mathsf{P}} \text{ loses}] &= \Pr_{\mathcal{O},\mathsf{c}}\left[\tilde{\mathsf{P}} \text{ loses at step } a = 0 \wedge \tilde{\mathsf{P}} \text{ loses at } a = 1 \wedge \ldots\right] \\
&= \prod_{a \in \Gamma^n} \Pr_{\mathcal{O},\mathsf{c}}\left[\tilde{\mathsf{P}} \text{ loses at step } a \middle| \tilde{\mathsf{P}} \text{ loses at all steps before } a\right] \\
&= \prod_{a \in \Gamma^n} \Pr_{\mathcal{O},\mathsf{c}}\left[\mathsf{P} \text{ does not output } c(a) \text{ on input } a \middle| \tilde{\mathsf{P}} \text{ loses at all steps before } a\right] \\
&= \prod_{a \in \Gamma^n} \frac{q^m - 1}{q^m} \\
&= \left(1 - \frac{1}{q^m}\right)^{q^n} \\
&= \left[\left(1 - \frac{1}{q^m}\right)^{q^m}\right]^{q^{n-m}} \\
&< \exp(-q^{n-m}).
\end{aligned}$$

since $\mathsf{c}(a)$ is chosen uniformly at random for each a. Hence, $\tilde{\mathsf{P}}$'s winning probability is at least $1 - \exp(-q^{n-m})$ as advertised, and there must exist a choice of function $\mathsf{c}(\cdot)$ that achieves this bound. ■

Proposition 5 *The protocol for $\mathsf{WOTRO}_\Gamma^{n,m}$ in the ROM model where both parties output the $\mathcal{O}(a)$ for any $a \in \Gamma^n$ and V always accepts is correct and statistically δ–secure, for $\delta = 1 - |\Gamma|^{n-m}$.*

The proof is identical to that of Proposition 3 by considering $r = \mathcal{O}(a)$.

Proposition 6 *The protocol described in Proposition 5 is $1 - \mathsf{negl}(n)$–secure in the ROM against polynomial-time provers as long as m is at least linear in n.*

Proof Let $\ell(n)$ be a polynomial which bounds the number of oracle queries that $\tilde{\mathsf{P}}$ can make. Furthermore, without loss of generality we will assume that $\tilde{\mathsf{P}}$ never makes the same oracle call twice. Then, given any function $\mathsf{c} : \Gamma^n \to \Gamma^m$, in order to cheat successfully, $\tilde{\mathsf{P}}$ must be able to find an a such that $\mathcal{O}(a) = \mathsf{c}(a)$.

Now, let $A_1, \cdots, A_{\ell(n)}$ be random variables taking values in Γ^n where A_i represents the ith query to the oracle (if $\tilde{\mathsf{P}}$ makes fewer than $\ell(n)$ queries, let A_i be any string that was not queried so far). These random variables are functions of the oracle $\mathcal{O}$, in that they can depend on the results of previous queries. We then have by the union bound that

$$\begin{aligned}
\Pr_{\mathcal{O}}\left[\tilde{\mathsf{P}} \text{ wins}\right] &\leqslant \Pr\left[\mathcal{O}(A_1) = \mathsf{c}(A_1) \vee \mathcal{O}(A_2) = \mathsf{c}(A_2) \vee \ldots \vee \mathcal{O}(A_{\ell(n)}) = \mathsf{c}(A_{\ell(n)})\right] \\
&\leqslant \sum_{i=1}^{\ell(n)} \Pr\left[\mathcal{O}(A_i) = \mathsf{c}(A_i)\right] \\
&= \ell(n) q^{-m} \\
&\leqslant \mathsf{negl}(n).
\end{aligned}$$

■

Proposition 7 *There are one-message implementations of* $\mathsf{WOTRO}_{\Gamma}^{n,n}$ *arbitrarily close to be* $\frac{1}{\mathrm{e}}$*–avoiding against unbounded provers in the CR$ model.*

PROOF Let $\ell(n)$ be the length of the CR$ (i.e. $r \in_R \Gamma^{\ell(n)}$) upper bounded by some polynomial. Let $\mathsf{P}^r : \Gamma^n \to \Gamma^m \times \Gamma^*$ denote P's message to V upon CR$ r and input $a \in \Gamma^n$. For $a \in \Gamma^n$ and CR$ $r \in \Gamma^{\ell(n)}$, we have $\mathsf{P}^r(a) = (c(r,a), v(r,a))$ which defines announcement $(a, c(r,a), v(r,a))$ to V. The verifier's algorithm $\mathsf{V}^r : \Gamma^n \times \Gamma^n \times \Gamma^* \to \{0,1\}$ upon CR$ r accepts (α, β, γ) when $\mathsf{V}^r(\alpha, \beta, \gamma) = 1$. The prover's algorithm can be considered deterministic given r, all randomness being provided by r. For $\{1, \ldots, p^n\} = \Gamma^n$ an enumeration of all elements in Γ^n, let

$$C^r := c(r,1) \| c(r,2) \| c(r,3) \| \ldots \| c(r,p^n)$$

be the sequence of all challenges announced by P upon CR$ r, one for each possible input $a \in \Gamma^n$. Let $\mathcal{C} := \{C^r\}_{r \in \Gamma^{\ell(n)}}$. For $\omega \in (\Gamma^n)^{p^n}$, we define

$$H_\omega := \{C \in \mathcal{C} \mid (\exists j \in [p^n])\, [C_j = \omega_j]\}$$

as the set of sequences containing challenges *hitting* ω somewhere. If Π is δ–avoiding then for all $\omega \in (\Gamma^n)^{p^n}$, $|H_\omega| \leq \delta \cdot p^{\ell(n)}$.

We define Π and then show it is $\frac{3}{4}$–avoiding using a CR$ $r \in (\Gamma^n)^2$. Π is simply defined from $r = r_1 \| r_2 \in (\Gamma^n)^2$ as

$$C^r = \underbrace{r_1, r_1, \ldots, r_1}_{\frac{p^n}{2} \text{ times}}, \underbrace{r_2, r_2, \ldots, r_2}_{\frac{p^n}{2} \text{ times}} \ .$$

We denote the elements of Γ^n by $\{1, 2, \ldots, p^n\}$. Let $\omega^* \in (\Gamma^n)^{p^n}$ be defined as

$$\omega^* := 1, 2, 3, \ldots, \frac{p^n}{2}, 1, 2, 3, \ldots, \frac{p^n}{2} \ .$$

It is not difficult to see that ω^* maximizes the probability to be hit by C^R. We have,

$$\begin{aligned} \Pr\left[C^R \in H_{\omega^*}\right] &= \Pr\left[\left(R_1 \leq \frac{p^n}{2}\right) \vee \left(R_2 \leq \frac{p^n}{2}\right)\right] \\ &= 1 - \Pr\left[\left(R_1 > \frac{p^n}{2}\right) \wedge \left(R_2 > \frac{p^n}{2}\right)\right] \\ &= 1 - \frac{1}{4} = \frac{3}{4} \ . \end{aligned}$$

By considering longer CR$ $r = r_1, r_2, \ldots, r_{\ell(n)}$ where $r_i \in \Gamma^n$, it is possible to get arbitrarily close to a correct $\frac{1}{\mathrm{e}}$–avoiding scheme with

$$C^r = \underbrace{r_1, r_1, \ldots, r_1}_{\frac{p^n}{\ell(n)} \text{ times}}, \underbrace{r_2, r_2, \ldots, r_2}_{\frac{p^n}{\ell(n)} \text{ times}}, \ldots, \underbrace{r_{\ell(n)}, r_{\ell(n)}, \ldots, r_{\ell(n)}}_{\frac{p^n}{\ell(n)} \text{ times}} \ .$$

■