Security of the Fiat-Shamir Transformation in the Quantum Random-Oracle Model

Jelle Don[1,2], Serge Fehr[1,3,4], Christian Majenz[2,4], and Christian Schaffner[2,4]

[1] Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands
[2] Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, Netherlands
[3] Mathematical Institute, Leiden University, Netherlands
[4] QuSoft, Amsterdam, Netherlands
donjelle@gmail.com, serge.fehr@cwi.nl, c.majenz@uva.nl, c.schaffner@uva.nl

Abstract. The famous Fiat-Shamir transformation turns any public-coin three-round interactive proof, i.e., any so-called Σ-protocol, into a non-interactive proof in the random-oracle model. We study this transformation in the setting of a *quantum adversary* that in particular may query the random oracle in quantum superposition.

Our main result is a generic reduction that transforms any quantum dishonest prover attacking the Fiat-Shamir transformation in the quantum random-oracle model into a similarly successful quantum dishonest prover attacking the underlying Σ-protocol (in the standard model). Applied to the standard soundness and proof-of-knowledge definitions, our reduction implies that both these security properties, in both the computational and the statistical variant, are preserved under the Fiat-Shamir transformation even when allowing quantum attacks. Our result improves and completes the partial results that have been known so far, but it also proves wrong certain claims made in the literature.

In the context of post-quantum secure signature schemes, our results imply that for any Σ-protocol that is a proof-of-knowledge against quantum dishonest provers (and that satisfies some additional natural properties), the corresponding Fiat-Shamir signature scheme is secure in the quantum random-oracle model. For example, we can conclude that the non-optimized version of Fish, which is the bare Fiat-Shamir variant of the NIST candidate Picnic, is secure in the quantum random-oracle model.

1 Introduction

The (quantum) random-oracle model. The *random-oracle model* (ROM) is a means to treat a cryptographic hash function H as an ideal primitive. In the ROM, the only way to "compute" the hash $H(x)$ of any value x is by making a *query* to an imaginary entity, the *random oracle* (RO), which has chosen H uniformly at random from the set of *all* functions with the considered domain and range.

The hope is that if a cryptographic scheme is secure in the ROM then it is also secure in the standard model, as long as H is instantiated with a "good

enough" cryptographic hash function. Even though in general we cannot hope to obtain provable security in the standard model in this way, (since there exist artificial counter examples [CGH04]), this approach works extremely well in practice, leading to very efficient schemes that tend to resist all known attacks.

What makes the ROM particularly convenient is that in the security proof of a cryptographic scheme, we can control the RO. For instance, simply by recording the queries that the adversary makes to the RO, we know exactly which hash values he knows, and the hash value $H(x)$ is random to him for any x that he has not queried. Furthermore, we can *reprogram* the RO, meaning that we can let $H(x)$ be some particular value y for some specific x, as long as it is random from the adversary's perspective.

When considering a *quantum* adversary, the picture changes a bit. In order to model that such an adversary can evaluate any hash function *in superposition* on different inputs, we must allow such a quantum adversary in the ROM to make superposition queries to the RO: for any superposition $\sum_x \alpha_x |x\rangle$ it may learn $\sum_x \alpha_x |x\rangle |H(x)\rangle$ by making a single query to the RO. This is referred to as the *quantum random-oracle model* (QROM) [BDF+11].

Unfortunately, these superposition queries obstruct the above mentioned advantages of the ROM. By basic properties of quantum mechanics one cannot observe or locally copy such superposition queries made by the adversary without disturbing them. Also, reprogramming is usually done for an x that is queried by the adversary at a certain point, so also here we are stuck with the problem that we cannot look at the queries without disturbing them.

As a consequence, security proofs in the ROM almost always do not carry over to the QROM. This lack of proof does not mean that the schemes become insecure; on the contrary, unless there is some failure because of some other reason[5], we actually expect typical schemes to remain secure. However, it is often not obvious how to find a security proof in the QROM. Some examples where security in the QROM has been established are [Unr14, Zha15, ES15, Unr15, KLS18, ABB+17, Zha18, SXY18, BDK+18].

Main technical result. Our main technical result (Theorem 2) can be understood as a particular way to overcome — to some extent — the above described limitation in the QROM of not being able to "read out" any query to the RO and to then reprogram the corresponding hash value. Concretely, we achieve the following.

We consider an arbitrary quantum algorithm $\mathcal{A}$ that makes queries to the RO and in the end outputs a pair (x, z), where z is supposed to satisfy some relation with respect to $H(x)$, e.g., $z = H(x)$. We then show how to *extract* early on, by measuring one of the queries that $\mathcal{A}$ makes, the very x that $\mathcal{A}$ will output, and to *reprogram* the RO at the point x with a fresh random value Θ, with the effect that the pair (x, z) that $\mathcal{A}$ then outputs now satisfies the given relation with respect to Θ, with a not too large loss in probability.

---

[5] E.g., the underlying computational hardness assumption does not hold anymore in the context of a quantum adversary.

The way this works is surprisingly simple. We choose the query that we measure uniformly at random among all the queries that $\mathcal{A}$ makes (also counting $\mathcal{A}$'s output), in order to (hopefully) obtain x. Subsequently we reprogram the RO, so as to answer x with Θ, *either* from this point on *or* from the following query on, where this binary choice is made at random. This last random decision seems counter-intuitive, but it makes our proof work. Indeed, we prove that the probability that (x, z) satisfies the required relation drops by no more than a factor $O(q^2)$, where q is the number of oracle queries $\mathcal{A}$ makes.

Application to the Fiat-Shamir transformation. The Fiat-Shamir transformation [FS87] turns any public-coin three-round interactive proof, i.e., any so-called Σ-protocol, into a non-interactive proof in the (Q)ROM. In the classical case it is well known that the security properties of the Σ-protocol are inherited by the Fiat-Shamir transformation [BR93, FKMV12]. In the quantum setting, when considering the security of the Fiat-Shamir transformation against quantum dishonest provers in the QROM, mainly negative results are known — see below for a more detailed exposition of previous results and how they compare to ours.

It is quite easy to see that the above result on the reprogrammability of the RO is exactly what is needed to turn a quantum prover that attacks the Fiat-Shamir transformation into a quantum prover that attacks the underlying Σ protocol. Indeed, from any Fiat-Shamir dishonest prover $\mathcal{A}$ that tries to produce a proof $\pi = (a, z)$ for a statement x, we obtain an interactive dishonest prover for the Σ protocol that extracts a from $\mathcal{A}$ and sends it to the verifier, and then uses the received challenge c to reprogram the RO, so that the z output by $\mathcal{A}$ will be a correct reply with respect to c with a probability not much smaller than the probability that $\mathcal{A}$ succeeds in forging π in the QROM.

This gives us a very generic transformation (stated in Theorem 8 below) from a Fiat-Shamir dishonest prover to a Σ-protocol dishonest prover that is similarly successful, up to a loss in probability of order $O(q^2)$. Applied to the standard notions of soundness and proof-of-knowledge, we prove that both these security properties, in both the computational and the statistical variant, are preserved under the Fiat-Shamir transformation in the QROM (Corollaries 13 and 16).

Comparison with prior results. Mainly negative results are known about the security of the Fiat-Shamir transformation against quantum attacks. Figure 1 shows a table copied from [ARU14], which outlines the different negative results on the security of Σ-protocols against quantum attackers that carry over to the Fiat-Shamir transformation. All the potential positive claims on the security of the Fiat-Shamir transformation were left unanswered (see Figure 1).

Currently, the only known positive result on the security of the Fiat-Shamir transformation against quantum attacks is the result by Unruh [Unr17], which shows that *statistical* soundness of the Σ-protocol implies *computational* soundness of the Fiat-Shamir transformation. This means, the lower right **stat** in the Fiat-Shamir column in Figure 1 was known to be 'comp'.

Properties of Σ-protocol		Σ-protocol directly		Fiat-Shamir transf.	
special soundness	strict soundness	PoK	proof	PoK	proof
perf	comp	attack	stat	attack	**stat**
comp	comp	attack	attack	attack	attack
perf	perf	stat	stat	**stat**	**stat**

Fig. 1: Table adapted from [ARU14], showing which versions of *special soundness* and *strict soundness*, which we call *unique responses*, imply that the Σ-protocol is a *proof of knowledge (PoK)* or a *proof* (in the sense of ordinary soundness). The values comp, stat and perf mean that the considered property holds respectively computationally, statistically and perfectly, and attack means that there exist example schemes that allow an attack. Gray values are copied from [ARU14]. The last column shows that the negative results carry over to the Fiat-Shamir transformation, while our results (in **bold face**) complete the table by showing that also the positive results carry over. Previously, the lower right corner entry was only known to be comp, and the other two entries were unknown. If the computational version of the unique responses, or strict soundness, property is replaced by our quantum strengthening (Definition 24), all instances of attack can be replaced by **comp**.

Our generic transformation from a Fiat-Shamir dishonest prover to a Σ-protocol dishonest prover implies that *all* the (considered) security properties of the Σ-protocol carry over *unchanged* to the Fiat-Shamir transformation, i.e., without degradation from 'stat' to 'comp'. Hence, we show that all the three open settings from [ARU14] are statistically secure, as shown in Figure 1. [6]

We point out that [DFG13] claims an impossibility result about the soundness of the Fiat-Shamir transformation as a quantum proof of knowledge, which contradicts one of our implications above. However, their result only applies to a restricted notion of proof of knowledge where the extractor is not allowed to measure any of the adversary's queries to the random oracle. The rational for this restriction was that such a measurement would disturb the adversary's quantum state beyond control; however, our technical result shows that it actually is possible to measure one of the adversary's queries and still have sufficient control over the adversary's behavior.

Relativizing prior negative results. At first glance, the negative results from [ARU14] together with our new positive results, as shown in the Fiat-Shamir column in Figure 1, seem to give a complete answer to the question of the security of the Fiat-Shamir transformation against quantum attacks. However, there is actually more to it.

We consider a *stronger* but still meaningful notion of *computational unique responses*, which is in the spirit of the *collapsing property* as introduced by Unruh [Unr16]. We call the new notion *quantum computationally unique responses*

---

[6] In the (quantum) random-oracle model, *statistical* security considers a computationally unbounded attacker with a polynomially bounded number of oracle queries.

and define it in Definition 24. Adapting a proof from [Unr12], it is not hard to see that a Σ-protocol with (perfect or computational) special soundness and quantum computational unique responses is a computational proof of knowledge. Therefore, our result then implies that its Fiat-Shamir transformation is a computational proof of knowledge as well.

Finally, our result also implies that if the Σ-protocol is *computationally sound* (as a 'proof'), then its Fiat-Shamir transformation is computationally sound as well. Interestingly, Unruh seems to suggest in [Unr17] (right after Theorem 21) that this is not true in general, due to a counterexample from [ARU14]. The counter example is, however, a Σ-protocol that is computationally *special* sound but not computationally sound (the issue being that in the quantum setting, special soundness does not imply ordinary soundness).

Thus, with the right adjustments of the considered *computational* soundness properties, the three negative answers in the Fiat-Shamir column in Figure 1 may actually be turned into positive answers. One caveat here is that we expect proving quantum computationally unique responses to be much harder than computational unique responses.

Application to signatures. Our positive results on the Fiat-Shamir transformation have direct applications to the security of Fiat-Shamir signatures. From the proof-of-knowledge property of the Fiat-Shamir transformation we immediately obtain the security of the Fiat-Shamir signature scheme under a *no-message attack*, assuming that the public key is a hard instance (Theorem 21). Furthermore, [Unr17] and [KLS18] have shown that for Fiat-Shamir signatures, up to some loss in the security parameter and under some additional mild assumptions on the underlying Σ-protocol, one can also derive security under *chosen-message attack*.

In conclusion, Fiat-Shamir signatures offer security against quantum attacks (in the QROM) if the underlying Σ-protocol is a proof of knowledge against quantum attacks and satisfies a few additional natural assumptions (Theorem 22).

As a concrete application, using Unruh's result on the collapsing property of the RO [Unr16] to argue the collapsing version of computational unique responses (which we call *quantum* computational unique responses) for the underlying Σ-protocol, we can conclude that the non-optimized version of Fish, which is the Fiat-Shamir variant of Picnic, is secure in the QROM.

Comparison with concurrent results. In concurrent and independent work [LZ19][7], Liu and Zhandry show results that are very similar to ours: they also show the security of the Fiat-Shamir transformation in the QROM, and they introduce a similar stronger version of the computational unique responses property in order to argue that a Σ-protocol is a (computational) proof of knowledge against a quantum adversary. In short, [LZ19] differs from the work here in the

---

[7] The paper [LZ19] was put on eprint (`ia.cr/2019/262`) a few days after our eprint version (`ia.cr/2019/190`).

following aspects. In [LZ19], the result on the Fiat-Shamir transformation is obtained using a very different approach, resulting in a greater loss in the reduction: $O(q^9)$ compared to the $O(q^2)$ loss that we obtain. On the other hand, on the quantum proof of knowledge front, Liu and Zhandry introduce some additional techniques that, for instance, allow them to prove that the Σ-protocol underlying Dilithium satisfies (their variant) of the newly introduced strong version of the computational unique responses property, while we phrase this as a conjecture in order to conclude the security of (some variant of) the Dilithium signature scheme.

2 Reprogramming the Quantum Random Oracle

We show and analyze a particular way to reprogram a random oracle in the quantum setting, where the oracle can be queried in superposition.

2.1 Notation

We consider a quantum oracle algorithm $\mathcal{A}$ that makes q queries to an *oracle*, i.e., an unspecified function $H : \mathcal{X} \to \mathcal{Y}$ with finite non-empty sets $\mathcal{X}, \mathcal{Y}$. We may assume without loss of generality that $\mathcal{A}$ makes no intermediary measurements. Formally, $\mathcal{A}$ is then described by a sequence of unitaries $A_1, \ldots, A_q$ and an initial state $|\phi_0\rangle$.[8] The unitaries A_i act on registers X, Y, Z, E, where X and Y have respective $|\mathcal{X}|$- and $|\mathcal{Y}|$-dimensional state spaces, while Z and E is arbitrary. As will become clear, X and Y are the quantum registers for the queries to H as well as for the final output x, Z is for the output z, and E is internal memory. For any concrete choice of $H : \mathcal{X} \to \mathcal{Y}$, we can write

$$\mathcal{A}^H |\phi_0\rangle := A_q \mathcal{O}^H \cdots A_1 \mathcal{O}^H |\phi_0\rangle \,,$$

for the execution of $\mathcal{A}$ with the oracle instantiated by H, where $\mathcal{O}^H$ is the unitary $\mathcal{O}^H : |x\rangle|y\rangle \mapsto |x\rangle|y \oplus H(x)\rangle$ that acts on registers X and Y.

It will be convenient to introduce the following notation. For $0 \leq i, j \leq q$ we set

$$\mathcal{A}^H_{i \to j} := A_j \mathcal{O}^H \cdots A_{i+1} \mathcal{O}^H$$

with the convention that $\mathcal{A}^H_{i \to j} := \mathbb{1}$ for $j \leq i$. Furthermore, we set

$$|\phi^H_i\rangle := \left(\mathcal{A}^H_{0 \to i}\right)|\phi_0\rangle$$

to be the state of $\mathcal{A}$ after the i-th step but right before the $(i+1)$-st query, and so that $|\phi^H_q\rangle$ equals $\left(\mathcal{A}^H_{0 \to q}\right)|\phi_0\rangle = \mathcal{A}^H |\phi_0\rangle$, the output state produced by $\mathcal{A}$.

Finally, for a given function $H : \mathcal{X} \to \mathcal{Y}$ and for fixed $x \in \mathcal{X}$ and $\Theta \in \mathcal{Y}$, we define the *reprogrammed* function $H * \Theta x : \mathcal{X} \to \mathcal{Y}$ that coincides with H on $\mathcal{X} \setminus \{x\}$ but maps x to Θ. With this notation at hand, we can then write

$$\left(\mathcal{A}^{H * \Theta x}_{i \to q}\right) \left(\mathcal{A}^H_{0 \to i}\right) |\phi_0\rangle = \left(\mathcal{A}^{H * \Theta x}_{i \to q}\right)|\phi^H_i\rangle$$

---

[8] Alternatively, we may understand $|\phi_0\rangle$ as an auxiliary input given to $\mathcal{A}$.

for an execution of $\mathcal{A}$ where the oracle is reprogrammed at a given point x after the i-th query.

We are interested in the probability that after the execution of $\mathcal{A}^H$ and upon measuring register X in the computational basis to obtain $x \in \mathcal{X}$, the state of register Z is of a certain form dependent on x and $H(x)$. This relation is captured by a projection G_x^H, where, more generally, for $x, x' \in \mathcal{X}$ and $\Theta \in \mathcal{Y}$ we set

$$G_{x,x'}^\Theta = |x'\rangle\langle x'| \otimes \mathbb{1} \otimes \Pi_{x,\Theta} \otimes \mathbb{1}\,,$$

where $\{\Pi_{x,\Theta}\}_{x \in \mathcal{X}, \Theta \in \mathcal{Y}}$ is a family of projections acting on Z, which we refer to as a *quantum predicate*. We use the short hands G_x^Θ for $G_{x,x}^\Theta$ and G_x^H for $G_x^{H(x)}$, i.e.,

$$G_x^H = |x\rangle\langle x| \otimes \mathbb{1} \otimes \Pi_{x,H(x)} \otimes \mathbb{1}\,.$$

For an arbitrary but fixed $x_\circ \in \mathcal{X}$, we then consider the probability

$$\|G_{x_\circ}^H |\phi_q^H\rangle\|_2^2\,.$$

Understanding $\mathcal{A}^H$ as an algorithm that outputs the measured x together with the state z in register Z, we will denote this probability also by

$$\Pr\big[x = x_\circ \wedge V(x, H(x), z) : (x, z) \leftarrow \mathcal{A}^H\big]\,,$$

understanding V to be a quantum predicate specified by the projections $\Pi_{x,H(x)}$.

2.2 Main technical result

We consider a quantum oracle algorithm $\mathcal{A}$ as formalized above, and we define a two-stage algorithm $\mathcal{S}$ with black-box access to $\mathcal{A}$ as follows. In the first stage, $\mathcal{S}$ tries to predict $\mathcal{A}$'s future output x, and then, upon input a (random) Θ, in the second stage tries to output what $\mathcal{A}$ is supposed to output, but now with respect to Θ instead of $H(x)$.

$\mathcal{S}$ works by running $\mathcal{A}$, but with the following modifications. First, one of the $q + 1$ queries of $\mathcal{A}$ (also counting the final output in register X) is selected uniformly at random and this query is measured, and the measurement outcome x is output by (the first stage of) $\mathcal{S}$. Then, this very query of $\mathcal{A}$ is answered either using the original H *or* using the reprogrammed oracle $H * \Theta x$, with the choice being made at random, while all the remaining queries of $\mathcal{A}$ are answered using oracle $H * \Theta x$.[9] Finally, (the second stage of) $\mathcal{S}$ outputs whatever $\mathcal{A}$ outputs.

Here, the figure of merit is the probability that for a fixed x, both the intermediate measurement and a measurement of the register X return x *and* that the register Z contains a state that satisfies the considered quantum predicate with respect to x and its (now reprogrammed) hash value Θ. Formally, this probability is captured by

$$\mathop{\mathbb{E}}_{\Theta,i,b}\Big[\big\|G_x^\Theta \left(\mathcal{A}_{i+b\to q}^{H*\Theta x}\right) \left(\mathcal{A}_{i\to i+b}^H\right) X|\phi_i^H\rangle\big\|_2^2\Big]$$

---

[9] If it is the final output that is measured then there is nothing left to reprogram.

where here and from now on, we use X as a short hand for the projection $|x\rangle\langle x|$ acting on X. The expectation is taken over $\Theta \in \mathcal{Y}$, $i \in \{0, ..., q\}$ and $b \in \{0, 1\}$ uniformly random. Note that the random bit $b \in \{0, 1\}$ determines whether the measured query is answered with H or with $H*\Theta x$.

We write $\mathcal{S}^\mathcal{A}[H]$ to emphasize that $\mathcal{S}$ only makes black-box access to $\mathcal{A}$ and that it depends on H. Our main technical lemma below then ensures that for any H and for a random $\Theta \in \mathcal{Y}$, the success probability of $\mathcal{S}^\mathcal{A}[H]$ is up to an order-q^2 loss not much smaller than that of $\mathcal{A}^{H*\Theta x}$, and therefore not much smaller than that of $\mathcal{A}^H$ in case of a random H.

Lemma 1 *For any $H : \mathcal{X} \to \mathcal{Y}$ and $x \in \mathcal{X}$, it holds that*

$$\mathop{\mathbb{E}}_{\Theta, i, b}\left[\left\|G_x^\Theta\left(\mathcal{A}_{i+b\to q}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2\right] \geq \frac{\mathbb{E}_\Theta\left[\left\||G_x^\Theta|\phi_q^{H*\Theta x}\rangle\right\|_2^2\right]}{2(q+1)(2q+3)} - \frac{\left\|X|\phi_q^H\rangle\right\|_2^2}{2(q+1)|\mathcal{Y}|}.$$

where the expectation is over random $\Theta \in \mathcal{Y}$, $i \in \{0, \ldots, q\}$ and $b \in \{0, 1\}$.[10]

Proof. We assume that the Y-register of $|\phi_q^H\rangle = \left(\mathcal{A}_{0\to q}^H\right)|\phi_0\rangle$ is $|0\rangle$ no matter what H is; this is without loss of generality since it can always be achieved by an insignificant modification to $\mathcal{A}$, i.e., by swapping Y with a default register within E. For the purpose of the proof, we introduce an additional step $\mathcal{A}_{q\to q+1}^H$ that simply applies $\mathcal{O}^H$, and we expand the notions of $|\phi_j^H\rangle$ and $\mathcal{A}_{i\to j}^H$ to allow $j = q + 1$. Finally, we "enhance" G_x^Θ to[11]

$$\tilde{G}_x^\Theta := G_x^\Theta(\mathbb{1} \otimes |\Theta\rangle\langle\Theta| \otimes \mathbb{1} \otimes \mathbb{1}) = X \otimes |\Theta\rangle\langle\Theta| \otimes \Pi_{x,\Theta} \otimes \mathbb{1}.$$

For any $0 \leq i \leq q$, inserting a resolution of the identity and exploiting that

$$\left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)(\mathbb{1} - X)|\phi_i^H\rangle = \left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)(\mathbb{1} - X)|\phi_i^H\rangle,$$

we can write

$$\left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)|\phi_{i+1}^H\rangle$$
$$= \left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)(\mathbb{1} - X)|\phi_i^H\rangle + \left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)X|\phi_i^H\rangle$$
$$= \left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)(\mathbb{1} - X)|\phi_i^H\rangle \qquad\qquad + \left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)X|\phi_i^H\rangle$$
$$= \left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)|\phi_i^H\rangle - \left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)X|\phi_i^H\rangle + \left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)X|\phi_i^H\rangle$$

---

[10] We consider $|\mathcal{Y}|$ to be superpolynomial in the security parameter, so that $\frac{1}{2(q+1)|\mathcal{Y}|}$ is negligible and can be neglected. In cases where $|\mathcal{Y}|$ is polynomial, the presented bound is not optimal, but an improved bound can be derived with the same kind of techniques.

[11] Informally, these modifications mean that we let $\mathcal{A}$ make one more query to get $H(x)$ into register Y, and $\tilde{G}_x^{H(x)}$ would then check that Y indeed contains $H(x)$.

Rearranging terms, applying $\tilde{G}_x^{\Theta}$ and using the triangle equality, we can thus bound

$$\left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)|\phi_i^H\rangle\right\|_2 \leq \left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)|\phi_{i+1}^H\rangle\right\|_2$$
$$+ \left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i\to q+1}^{H*\Theta x}\right)X|\phi_i^H\rangle\right\|_2$$
$$+ \left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+1\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+1}^H\right)X|\phi_i^H\rangle\right\|_2.$$

Summing up the respective sides of the inequality over $i = 0, \ldots, q$, we get

$$\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^{H*\Theta x}\rangle\right\|_2 \leq \left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^H\rangle\right\|_2 + \sum_{\substack{0\leq i\leq q\\ b\in\{0,1\}}}\left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+b\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2.$$

By squaring both sides, dividing by $2q+3$ (i.e., the number of terms on the right hand side), and using Jensen's inequality on the right hand side, we obtain

$$\frac{\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^{H*\Theta x}\rangle\right\|_2^2}{2q+3} \leq \left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^H\rangle\right\|_2^2 + \sum_{\substack{0\leq i\leq q\\ b\in\{0,1\}}}\left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+b\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2$$

and thus

$$\mathop{\mathbb{E}}_{i,b}\left[\left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+b\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2\right] \geq \frac{\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^{H*\Theta x}\rangle\right\|_2^2}{2(q+1)(2q+3)} - \frac{\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^H\rangle\right\|_2^2}{2(q+1)}. \tag{1}$$

Since both $|\Theta\rangle\langle\Theta|$ and $\mathcal{A}_{q\to q+1}^{H*\Theta x} = \mathcal{O}^{H*\Theta x}$ commute with G_x^{Θ}, we get

$$\left\|\tilde{G}_x^{\Theta}\left(\mathcal{A}_{i+b\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2 \leq \left\|G_x^{\Theta}\left(\mathcal{A}_{i+b\to q+1}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2$$
$$= \left\|G_x^{\Theta}\left(\mathcal{A}_{i+b\to q}^{H*\Theta x}\right)\left(\mathcal{A}_{i\to i+b}^H\right)X|\phi_i^H\rangle\right\|_2^2. \tag{2}$$

Also, because $(X\otimes|\Theta\rangle\langle\Theta|)\mathcal{O}^{H*\Theta x} = (X\otimes\mathbb{1})\mathcal{O}^{H*\Theta x}$, and $\mathcal{O}^{H*\Theta x}$ commutes with G_x^{Θ}, we get

$$\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^{H*\Theta x}\rangle\right\|_2^2 = \left\|G_x^{\Theta}|\phi_q^{H*\Theta x}\rangle\right\|_2^2. \tag{3}$$

Finally,

$$\mathop{\mathbb{E}}_{\Theta}\left[\left\|\tilde{G}_x^{\Theta}|\phi_{q+1}^H\rangle\right\|_2^2\right] \leq \mathop{\mathbb{E}}_{\Theta}\left[\left\|(X\otimes|\Theta\rangle\langle\Theta|)\mathcal{O}^H|\phi_q^H\rangle\right\|_2^2\right] \leq \frac{1}{|\mathcal{Y}|}\left\|X|\phi_q^H\rangle\right\|_2^2. \tag{4}$$

Inserting (2), (3) and (4) into (1) yields the claimed result. $\square$

2.3 Switching notation, and simulating the random oracle

Introducing more algorithmic-probabilistic notation, we write

$$(x, x', z) \leftarrow \langle \mathcal{S}^{\mathcal{A}}[H], \Theta\rangle$$

to specify the probability space determined as follows, relying on the above construction of the two-stage algorithm $\mathcal{S}$ when given $\mathcal{A}$. In the first stage $\mathcal{S}^{\mathcal{A}}[H]$ produces x, and then in the second stage, upon receiving Θ, it produces x' and z, where z may be quantum. Our figure of merit above, i.e., the left hand side of the bound in Lemma 1 (with x replaced by $x_\circ$), is then denoted by

$$\Pr_{\Theta}\left[x = x_\circ \wedge x' = x_\circ \wedge V(x, \Theta, z) : (x, x', z) \leftarrow \langle \mathcal{S}^{\mathcal{A}}[H], \Theta \rangle\right],$$

where the subscript Θ in $\Pr_{\Theta}$ denotes that the probability is averaged over a random choice of Θ.

Using this notation, but also weakening the bound slightly by not requiring $x' = x_\circ$, for any H and $x_\circ$ the bound from Lemma 1 then becomes

$$\Pr_{\Theta}\left[x = x_\circ \wedge V(x, \Theta, z) : (x, z) \leftarrow \langle \mathcal{S}^{\mathcal{A}}[H], \Theta \rangle\right]$$

$$\gtrsim \frac{1}{O(q^2)} \Pr_{\Theta}\left[x = x_\circ \wedge V(x, H(x), z) : (x, z) \leftarrow \mathcal{A}^{H * \Theta x}\right]$$

where the approximate inequality $\gtrsim$ hides the term

$$\frac{1}{2(q+1)|\mathcal{Y}|} \Pr_{H}\left[x = x_\circ : (x, z) \leftarrow \mathcal{A}^H\right].$$

Recall that the output z may be a quantum state, in which case the predicate V is given by a measurement that depends on x, and $H(x)$ or Θ, respectively.

We fix a family $\mathcal{H}$ of $2(q+1)$-wise independent hash functions and average the above inequality over a random choice of $H \in \mathcal{H}$ from this family. We simply write $\mathcal{S}$ for $\mathcal{S}[H]$ with H chosen like that. Furthermore, we observe that, for any fixed x, the family $\{H * \Theta x \mid H \in \mathcal{H}, \Theta \in \{0,1\}^n\}$ is a family of $2(q+1)$-wise independent hash functions as well. Finally, we use that $\mathcal{A}$ (together with the check $V(x, H(x), z)$) cannot distinguish a random function $H * \Theta x$ in that family from a fully random function H [Zha12]. This gives us the following variation of Lemma 1, which we state as our main technical theorem:

Theorem 2 (Measure-and-reprogram) *Let $\mathcal{X}, \mathcal{Y}$ be finite non-empty sets. There exists a black-box polynomial-time two-stage quantum algorithm $\mathcal{S}$ with the following property. Let $\mathcal{A}$ be an arbitrary oracle quantum algorithm that makes q queries to a uniformly random $H : \mathcal{X} \to \mathcal{Y}$ and that outputs some $x \in \mathcal{X}$ and a (possibly quantum) output z. Then, the two-stage algorithm $\mathcal{S}^{\mathcal{A}}$ outputs some $x \in \mathcal{X}$ in the first stage and, upon a random $\Theta \in \mathcal{Y}$ as input to the second stage, a (possibly quantum) output z, so that for any $x_\circ \in \mathcal{X}$ and any predicate[12] V:*

$$\Pr_{\Theta}\left[x = x_\circ \wedge V(x, \Theta, z) : (x, z) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \Theta \rangle\right]$$

$$\gtrsim \frac{1}{O(q^2)} \Pr_{H}\left[x = x_\circ \wedge V(x, H(x), z) : (x, z) \leftarrow \mathcal{A}^H\right],$$

where the $\gtrsim$ hides a term that is bounded by $\frac{1}{2q|\mathcal{Y}|}$ when summed over all $x_\circ$.

---

[12] We recall that in case z is a quantum state, V is given by means of a measurement.

Remark 3 We do not spell out in detail what it means for a quantum algorithm like $\mathcal{S}$ to be *black-box*; see e.g. [Unr17] for a rigorous definition. What we obviously need here is that $\mathcal{S}^{\mathcal{A}}$ has access to $\mathcal{A}$'s initial state $|\phi_0\rangle$ and to q, and is given black-box access to the unitaries A_i. Furthermore, for later purposes, we need the following composition property: if $\mathcal{S}$ is a black-box algorithm with access to $\mathcal{A}$, and $\mathcal{K}$ is a black-box algorithm with access to $\mathcal{S}^{\mathcal{A}}$, then there exists a black-box algorithm $\mathcal{K}^{\mathcal{S}}$ with access to $\mathcal{A}$ so that $(\mathcal{K}^{\mathcal{S}})^{\mathcal{A}} = \mathcal{K}^{(\mathcal{S}^{\mathcal{A}})}$.

3 Security of the Fiat-Shamir Transformation

In this section, we show how to reduce security of the Fiat-Shamir transformation to the security of the underlying Σ-protocol: any dishonest prover attacking the Fiat-Shamir transformation can be turned into a dishonest prover that succeeds to break the underlying Σ-protocol with the same probability up to a polynomial loss. This reduction is obtained by a straightforward application of Theorem 2. Our security reduction holds very generically and is not strongly tight to the considered notion of security, as long as the respective security definitions for the Σ-protocol and the Fiat-Shamir transformation "match up".

3.1 Σ-protocols

We recall the definition of a Σ-protocol.

Definition 4 (Σ-protocol) *A Σ-protocol $\Sigma = (\mathcal{P}, \mathcal{V})$ for a relation $R \subseteq \mathcal{X} \times \mathcal{W}$ is a three-round two-party interactive protocol of the form:*

Prover $\mathcal{P}(x, w)$		*Verifier $\mathcal{V}(x)$*
	$\xrightarrow{\ a\ }$	
	$\xleftarrow{\ c\ }$	$c \xleftarrow{\$} \mathcal{C}$
	$\xrightarrow{\ z\ }$	Accept iff $V(x, a, c, z) = 1$

Using our terminology and notation from above, $\mathcal{P}$ is a two-stage algorithm and we can write

$$(a, z) \leftarrow \langle \mathcal{P}(x, w), c \rangle$$

for the generation of the first message a in the first stage and the reply z in the second stage once given the challenge c.

Remark 5 We allow the set of *instances* $\mathcal{X}$, the set of *witnesses* $\mathcal{W}$ and the relation R to depend on a security parameter η. Similarly, the interactive algorithms $\mathcal{P}$ and $\mathcal{V}$ may depend on η (or have η as part of their input). However, for ease of notation, we suppress these dependencies on η unless they are crucial.

Remark 6 We do not necessarily require a Σ-protocol to be perfectly or statistically correct. This allows us to include protocols that use *rejection sampling*, where with a constant probability, the value z would leak too much information on the witness w and so the prover sends $\perp$ instead. On the other hand, by default we consider the soundness/knowledge error to be negligible, i.e., a dishonest prover succeeds only with negligible probability to make the verifier accept if x is not a valid instance or the prover has no witness for it (depending on the considered soundness notion). Negligible soundness/knowledge error can always be achieved by parallel repetition (see e.g. [Dam10]).

3.2 The Fiat-Shamir transformation

The *Fiat-Shamir transformation* turns a Σ-protocol Σ into a non-interactive proof system, denoted $\mathsf{FS}[\Sigma]$, by replacing the verifier's random choice of $c \in \mathcal{C}$ with $c := H(x, a)$, where $H : \mathcal{X}' \to \mathcal{C}$ is a hash function with a domain $\mathcal{X}'$ that contains all pairs $x' = (x, a)$ with $x \in \mathcal{X}$ and a produced by $\mathcal{P}$. In other words, upon input x and w, the honest FS-prover produces $\pi = (a, z)$ by running the two-stage Σ-protocol prover $\mathcal{P}$ but using $c = H(x, a)$ as challenge (i.e., as input to the second stage). In case Σ is not statistically correct, the above process of producing $\pi = (a, z)$ is repeated sufficiently many times until $V(x, a, H(x, a), z)$ is satisfied (or some bound is reached). In either case, we will write this as

$$\pi = (a, z) \leftarrow P^H_{FS}(x, w).$$

We may write as $V^H_{FS}(x, \pi)$ the FS-verifier's check whether $V(x, a, H(x, a), z)$ is satisfied or not. In the security analysis, the hash function H is modeled by a random oracle, i.e. by oracle access to a uniformly random $H : \mathcal{X}' \to \mathcal{C}$.

When considering an *adversary* $\mathcal{A}$ that tries to *forge* a proof for some instance $x \in \mathcal{X}$, one can distinguish between an *arbitrary but fixed* x, and an x that is *chosen* by $\mathcal{A}$ and output along with a in case of Σ-protocols, respectively along with π in case of the Fiat-Shamir transformation. If x is fixed then the adversary is called *static*, otherwise it is called *adaptive*. For the typical security definitions for Σ-protocols this distinction between a static and an adaptive $\mathcal{A}$ makes no difference (see Lemmas 12 and 15 below), but for the Fiat-Shamir transformation it (potentially) does.

3.3 The generic security reduction

Since an adaptive adversary is clearly not less powerful than a static adversary, we restrict our attention for the moment to the adaptive case. Recall that such an adaptive FS-adversary $\mathcal{A}$ outputs the instance $x \in \mathcal{X}$ along with the proof $\pi = (a, z)$, and the figure of merit is the probability that x, a, z satisfies $V(x, a, H(x, a), z)$. Thus, we can simply apply Theorem 2, with (x, a) playing the role of what is referred to as x in the theorem statement, to obtain the existence of an adaptive Σ-adversary $\mathcal{S}^{\mathcal{A}}$ that produces (x, a) in a first stage, and upon

receiving a random challenge c produces z, such that for any $x_\circ \in \mathcal{X}$

$$\Pr_c\big[x{=}x_\circ \wedge V(x,a,c,z) : (x,a,z) \leftarrow \langle \mathcal{S}^\mathcal{A}, c\rangle\big]$$

$$\gtrsim \frac{1}{O(q^2)} \Pr_H\big[x{=}x_\circ \wedge V(x,a,H(x,a),z) : (x,a,z) \leftarrow \mathcal{A}^H\big],$$

where the approximate inequality hides a term that is bounded by $\frac{1}{2q|\mathcal{C}|}$ when summed over all $x_\circ \in \mathcal{X}$. Understanding that x is given to $\mathcal{V}$ along with the first message a but also treating it as an output of $\mathcal{S}^\mathcal{A}$, while $\mathcal{V}$'s output v is its decision to accept or not, we write this as

$$\Pr\big[x{=}x_\circ \wedge v = accept : (x,v) \leftarrow \langle \mathcal{S}^\mathcal{A}, \mathcal{V}\rangle\big]$$

$$\gtrsim \frac{1}{O(q^2)} \Pr_H\big[x{=}x_\circ \wedge V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\big].$$

Summed over all $x_\circ \in \mathcal{X}$, this in particular implies that

$$\Pr\big[\langle \mathcal{S}^\mathcal{A}, \mathcal{V}\rangle = accept\big] \geq \frac{1}{O(q^2)} \Pr_H\big[V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\big] - \frac{1}{2q|\mathcal{C}|}.$$

Remark 7 We point out that the above arguments extend to a FS-adversary $\mathcal{A}$ that, besides the instance x and the proof $\pi = (a,z)$, also produces some local (possibly quantum) output satisfying some (quantum) predicate that may depend on x, a, z. The resulting Σ-adversary $\mathcal{S}^\mathcal{A}$ is then ensured to produce a local output that satisfies the considered predicate as well, up to the given loss in the probability. Indeed, we can simply include this local output in z and extend the predicate V accordingly.

In a very broad sense, the above means that for *any* FS-adversary $\mathcal{A}$ there exists a Σ-adversary $\mathcal{S}^\mathcal{A}$ that "*achieves the same thing*" up to a $O(q^2)$ loss in success probability. Hence, for matching corresponding security definitions, security of a Σ-protocol (against a dishonest prover) implies security of its Fiat-Shamir transform.

We summarize here the above basic transformation from an adaptive FS-adversary $\mathcal{A}$ to an adaptive Σ-adversary $\mathcal{S}^\mathcal{A}$.

Theorem 8 *There exists a black-box quantum polynomial-time two-stage quantum algorithm $\mathcal{S}$ such that for any adaptive Fiat-Shamir adversary $\mathcal{A}$, making q queries to a uniformly random function H with appropriate domain and range, and for any $x_\circ \in \mathcal{X}$:*

$$\Pr\big[x{=}x_\circ \wedge v = accept : (x,v) \leftarrow \langle \mathcal{S}^\mathcal{A}, \mathcal{V}\rangle\big]$$

$$\gtrsim \frac{1}{O(q^2)} \Pr_H\big[x{=}x_\circ \wedge V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\big],$$

where the $\gtrsim$ hides a term that is bounded by $\frac{1}{2q|\mathcal{C}|}$ when summed over all $x_\circ$.

Below, we apply the above general reduction to the respective standard definitions for *soundness* and *proof of knowledge*. Each property comes in the variants *computational* and *statistical*, for guarantees against computationally bounded or unbounded adversaries respectively, and one may consider the static or the adaptive case.

3.4 Preservation of soundness

Let $\Sigma = (\mathcal{P}, \mathcal{V})$ be a Σ-protocol for a relation R, and let $\mathsf{FS}[\Sigma]$ be its Fiat-Shamir transformation. We set $\mathcal{L} := \{x \in \mathcal{X} \mid \exists\, w \in \mathcal{W} : R(x, w)\}$. It is understood that $\mathcal{P}$ and $\mathcal{V}$, as well as R and thus $\mathcal{L}$, may depend on a security parameter η. We note that in the following definition, we overload notation a bit by writing $\mathcal{A}$ for both for the ordinary static and for the adaptive adversary (even though a given $\mathcal{A}$ is usually either static or adaptive).

Definition 9 Σ *is **(computationally/statistically) sound** if there exists a negligible function $\mu(\eta)$ such that for any (quantum polynomial-time/unbounded) adversary $\mathcal{A}$ and any $\eta \in \mathbb{N}$:*

$$\Pr\left[\langle \mathcal{A}, \mathcal{V}(x) \rangle = accept\right] \leq \mu(\eta)$$

*for all $x \notin \mathcal{L}$; respectively, in case of an **adaptive** $\mathcal{A}$:*

$$\Pr\left[x \notin \mathcal{L} \wedge v = accept : (x, v) \leftarrow \langle \mathcal{A}, \mathcal{V} \rangle\right] \leq \mu(\eta).$$

$\mathsf{FS}[\Sigma]$ *is **(computationally/statistically) sound** if there exists a negligible function $\mu(\eta)$ and a constant e such that for any (quantum polynomial-time/unbounded) adversary $\mathcal{A}$ and any $\eta \in \mathbb{N}$:*

$$\Pr_{H}\left[V_{FS}^{H}(x, \pi) : \pi \leftarrow \mathcal{A}^{H}\right] \leq q^{e}\mu(\eta)$$

*for all $x \notin \mathcal{L}$; respectively, in case of an **adaptive** $\mathcal{A}$:*

$$\Pr_{H}\left[V_{FS}^{H}(x, \pi) \wedge x \notin \mathcal{L} : (x, \pi) \leftarrow \mathcal{A}^{H}\right] \leq q^{e}\mu(\eta).$$

Remark 10 Note that for the soundness of $\mathsf{FS}[\Sigma]$, the adversary $\mathcal{A}$'s success probability may unavoidably grow with the number q of oracle queries, but we require that it grows only polynomially in q.

Remark 11 In line with Section 2, the description of a quantum algorithm $\mathcal{A}$ is understood to include the initial state $|\phi_0\rangle$. As such, when quantifying over all $\mathcal{A}$ it is understood that this includes a quantification over all $|\phi_0\rangle$ as well. This stays true when considering $\mathcal{A}$ to be quantum polynomial-time, which means that the unitaries A_i can be computed by polynomial-time quantum circuits, and q is polynomial in size, but does not put any restriction on $|\phi_0\rangle$.[13] This is in line with [Unr12, Def. 1], which explicitly spells out this quantification.

---

[13] In other words, $\mathcal{A}$ is then *non-uniform* quantum polynomial-time with *quantum* advice.

We consider the following to be folklore knowledge; for completeness, we still give a proof in Appendix A.

Lemma 12 *If Σ is computationally/statistically sound for* static *adversaries then it is also computationally/statistically sound for* adaptive *adversaries.*

The following is now an immediate application of Theorem 8 and the above observation regarding static and adaptive security for Σ-protocols.

Corollary 13 *Let Σ be a Σ-protocol with superpolynomially sized challenge space $\mathcal{C}$. If Σ is computationally/statistically sound against a static adversary then $\mathsf{FS}[\Sigma]$ is computationally/statistically sound against an adaptive adversary.*

Proof. Applying Theorem 8, we find that for any adaptive FS-adversary $\mathcal{A}$, polynomially bounded in the computational setting, there exists an adaptive Σ-protocol adversary $\mathcal{S}^{\mathcal{A}}$, polynomially bounded if $\mathcal{A}$ is, so that

$$\Pr\big[x \notin \mathcal{L} \wedge V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\big]$$

$$= \sum_{x_\circ \notin \mathcal{L}} \Pr\big[x=x_\circ \wedge V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\big]$$

$$\leq O(q^2) \cdot \left(\left(\sum_{x_\circ \notin \mathcal{L}} \Pr\big[x=x_\circ \wedge v = accept : (x,v) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \mathcal{V}\rangle\big]\right) + \frac{1}{2q|\mathcal{C}|}\right)$$

$$= O(q^2) \cdot \left(\Pr\big[x \notin \mathcal{L} \wedge v = accept : (x,v) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \mathcal{V}\rangle\big]\right) + \frac{O(q)}{|\mathcal{C}|}$$

$$\leq O(q^2) \cdot \mu(\eta) + \frac{O(q)}{|\mathcal{C}|}$$

where the last inequality holds for some negligible function $\mu(\eta)$ if Σ is sound against an adaptive adversary. The latter is ensured by the assumed soundness against a static adversary and Lemma 12. This bound can obviously be written as $q^2 \mu'(\eta)$ for another negligible function $\mu'(\eta)$, showing the claimed soundness of $\mathsf{FS}[\Sigma]$. □

3.5 Preservation as a proof of knowledge

We now recall the definition of a proof of knowledge, sometimes also referred to as (witness) extractability, tailored to the case of a negligible "knowledge error". Informally, the requirement is that if $\mathcal{A}$ succeeds in proving an instance x, then by using $\mathcal{A}$ as a black-box only it is possible to extract a witness for x. In case of an arbitrary but fixed x, this property is formalized in a rather straightforward way; however, in case of an adaptive $\mathcal{A}$, the formalization is somewhat subtle, because one can then not refer to *the* x for which $\mathcal{A}$ manages to produce a proof. We adopt the approach (though not the precise formalization) from [Unr17], which requires x to satisfy an arbitrary but fixed predicate.

Definition 14 Σ *is a **(computational/statistical) proof of knowledge** if there exists a quantum polynomial-time black-box 'knowledge extractor' $\mathcal{K}$, a polynomial $p(\eta)$, a constant $d \geq 0$, and a negligible function $\kappa(\eta)$ such that for any (quantum polynomial-time/unbounded) adversary $\mathcal{A}$, any $\eta \in \mathbb{N}$ and any $x \in \mathcal{X}$:*

$$\Pr\left[(x,w) \in R : w \leftarrow \mathcal{K}^{\mathcal{A}}(x)\right] \geq \frac{1}{p(\eta)} \cdot \Pr\left[\langle \mathcal{A}, \mathcal{V}(x)\rangle = accept\right]^d - \kappa(\eta) \,;$$

*respectively, in case of an **adaptive** $\mathcal{A}$:*

$$\Pr\left[x \in X \,\wedge\, (x,w) \in R : (x,w) \leftarrow \mathcal{K}^{\mathcal{A}}\right]$$

$$\geq \frac{1}{p(\eta)} \cdot \Pr\left[x \in X \wedge v = accept : (x,v) \leftarrow \langle \mathcal{A}, \mathcal{V}\rangle\right]^d - \kappa(\eta)$$

for any subset $X \subseteq \mathcal{X}$.

$\mathsf{FS}[\Sigma]$ *is a **(computational/statistical) proof of knowledge** if there exists a polynomial-time black-box 'knowledge extractor' $\mathcal{E}$, a polynomial $p(\eta)$, constants $d, e \geq 0$, and a negligible function $\mu(\eta)$, such that for any (quantum polynomial-time/unbounded) algorithm $\mathcal{A}$, any $\eta \in \mathbb{N}$ and any $x \in \mathcal{X}$:*

$$\Pr\left[(x,w) \in R : (x,w) \leftarrow \mathcal{E}^{\mathcal{A}}\right] \geq \frac{1}{q^e p(\eta)} \cdot \Pr_H\left[V_{FS}^H(x,\pi) : \pi \leftarrow \mathcal{A}^H\right]^d - \mu(\eta). \,;$$

*respectively, in case of an **adaptive** $\mathcal{A}$:*

$$\Pr\left[x \in X \,\wedge\, (x,w) \in R : (x,w) \leftarrow \mathcal{E}^{\mathcal{A}}\right]$$

$$\geq \frac{1}{q^e p(\eta)} \cdot \Pr_H\left[x \in X \wedge V_{FS}^H(x,\pi) : (x,\pi) \leftarrow \mathcal{A}^H\right]^d - \mu(\eta)$$

for any subset $X \subseteq \mathcal{X}$, where q is the number of queries $\mathcal{A}$ makes.

Also here, for Σ-protocols static security implies adaptive security.

Lemma 15 *If Σ is a computational/statistical proof of knowledge for* static $\mathcal{A}$ *then it is also a computational/statistical proof of knowledge for* adaptive $\mathcal{A}$.

Again, the following is now an immediate application of Theorem 8 and the above observation regarding static and adaptive security for Σ-protocols.

Corollary 16 *Let Σ be a Σ-protocol with superpolynomially sized $\mathcal{C}$. If Σ is a computational/statistical proof of knowledge for static adversaries then $\mathsf{FS}[\Sigma]$ is a computational/statistical proof of knowledge for adaptive adversaries.*

Proof. First, we observe that by Lemma 15, we may assume Σ to be a computational/statistical proof of knowledge for *adaptive* adversaries. Let $\mathcal{K}$ be the black-box knowledge extractor. Let $\mathcal{A}$ be an (quantum polynomial-time/unbounded) adaptive FS-adversary $\mathcal{A}$. We define a black-box knowledge extractor $\mathcal{E}$ for $\mathsf{FS}[\Sigma]$ as follows. $\mathcal{E}^{\mathcal{A}}$ simply works by running $\mathcal{K}^{\mathcal{S}^{\mathcal{A}}}$, where $\mathcal{S}^{\mathcal{A}}$ the adaptive Σ-protocol

adversary obtained by invoking Theorem 8. For any subset $X \subseteq \mathcal{X}$, invoking the proof-of-knowledge property of Σ and using Theorem 8, we see that

$$\Pr\big[x \in X \wedge (x, w) \in R : (x, w) \leftarrow \mathcal{E}^{\mathcal{A}}\big]$$

$$= \Pr\big[x \in X \wedge (x, w) \in R : (x, w) \leftarrow \mathcal{K}^{\mathcal{S}^{\mathcal{A}}}\big]$$

$$= \frac{1}{p(\eta)} \cdot \Pr\big[x \in X \wedge v = accept(x, v) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \mathcal{V} \rangle\big]^d - \kappa(\eta)$$

$$= \frac{1}{p(\eta)} \cdot \bigg(\sum_{x_\circ \in X} \Pr\big[x = x_\circ \wedge v = accept(x, v) \leftarrow \langle \mathcal{S}^{\mathcal{A}}, \mathcal{V} \rangle\big] \bigg)^d - \kappa(\eta)$$

$$\geq \frac{1}{p(\eta)} \bigg(\frac{1}{O(q^2)} \sum_{x_\circ \in X} \Pr_H\big[x = x_\circ \wedge V_{FS}^H(x, \pi) : (x, \pi) \leftarrow \mathcal{A}^H\big] - \frac{1}{2q|\mathcal{C}|} \bigg)^d - \kappa(\eta)$$

$$\geq \frac{1}{p(\eta) \cdot O(q^{2d})} \cdot \Pr_H\big[x \in X \wedge V_{FS}^H(x, \pi) : (x, \pi) \leftarrow \mathcal{A}^H\big]^d - \mu(\eta)$$

for some negligible function $\mu(\eta)$. □

Remark 17 We point out that in [Unr17] Unruh considers a stronger notion of extractability than our Definition 14, where it is required that, in some sense, the extractor also recovers any local (possibly quantum) output of the adversary $\mathcal{A}$. In the light of Remark 7, we expect that our result also applies to this stronger notion of extractability.

4   Application to Fiat-Shamir signatures

Any Fiat-Shamir non-interactive proof system can easily be transformed into a public-key signature scheme.[14] The signer simply proves knowledge of a witness (the secret key) for a composite statement $x^* := x\|m$, which includes the public key x as well as the message m. The signature σ then consists of a proof for x^*.

Definition 18 *A* binary relation R with instance generator G *is said to be* hard *if for any quantum polynomial-time algorithm $\mathcal{A}$ we have*

$$\Pr\left[(x, w') \in R : (x, w) \leftarrow G, w' \leftarrow \mathcal{A}(x)\right] \leq \mu(\eta)$$

for some negligible function $\mu(\eta)$, where G is such that it always outputs a pair $(x, w) \in R$.

Definition 19 *A* Fiat-Shamir signature scheme *based on a Σ-protocol $\Sigma = (\mathcal{P}, \mathcal{V})$ for a hard relation R with instance generator G, denoted by $\mathsf{Sig}[\Sigma]$ is defined by the triple (Gen, Sign, Verify), with*

---

[14] In fact, that is how the Fiat-Shamir transform was originally conceived in [FS87]. Only later [BG93] adapted the idea to construct a non-interactive zero-knowledge proof system.

- Gen*: Pick $(x, w) \leftarrow G$, set $sk := (x, w)$ and $pk := x$.*
- Sign$^H(sk, m)$*: Return (m, σ) where $\sigma \leftarrow P_{FS}^H(x\|m, w)$.*
- Verify$^H(pk, m, \sigma)$*: Return $V_{FS}^H(x\|m, \sigma)$.*

Here $(P_{FS}^H, V_{FS}^H) = \mathsf{FS}[\Sigma^*]$, where $\Sigma^* = (\mathcal{P}^*, \mathcal{V}^*)$ is the Σ-protocol obtained from Σ by setting $\mathcal{P}^*(x\|m) = \mathcal{P}(x)$ and $\mathcal{V}^*(x\|m) = \mathcal{V}(x)$ for any m.

Note that by definition of FS in Section 3.2, we use $V_{FS}^H(x\|m, \sigma)$ as shortcut for $V(x\|m, a, H(x\|m, a), z)$.

We investigate the following standard security notions for signature schemes.

Definition 20 (sEUF$-$CMA/EUF$-$NMA) *A signature scheme fulfills* strong existential unforgeability under chosen-message attack (sEUF$-$CMA) *if for all quantum polynomial-time algorithms $\mathcal{A}$ and for uniformly random $H : \mathcal{X}' \rightarrow \mathcal{C}$ it holds that*

$$\Pr\Big[\mathsf{Verify}^H(pk, m, \sigma) \wedge (m, \sigma) \notin \mathbf{Sig-q} : (pk, sk) \leftarrow \mathsf{Gen}, (m, \sigma) \leftarrow \mathcal{A}^{H, \mathbf{Sig}}(pk)\Big]$$

is negligible. Here $\mathbf{Sig}$ is classical oracle which upon classical input m returns $\mathsf{Sign}^H(m, sk)$, and $\mathbf{Sig-q}$ is the list of all queries made to $\mathbf{Sig}$.

Analogously, a signature scheme fulfills existential unforgeability under no-message attack (EUF$-$NMA) *if for all quantum polynomial-time algorithms $\mathcal{A}$ and for uniformly random $H : \mathcal{X}' \rightarrow \mathcal{C}$ it holds that*

$$\Pr\Big[\mathsf{Verify}^H(pk, m, \sigma) : (pk, sk) \leftarrow \mathsf{Gen}, (m, \sigma) \leftarrow \mathcal{A}^H(pk)\Big]$$

is negligible.

The unforgeability (against no-message attacks) of a Fiat-Shamir signature scheme is shown below to follow from the proof-of-knowledge property of the underlying proof system (hence, as we now know, of the underlying Σ-protocol), under the assumption that the relation is hard, i.e. it is infeasible to compute sk from pk.

Theorem 21 *Let Σ be Σ-protocol for some hard relation R, with superpolynomially sized challenge space $\mathcal{C}$ and the proof-of-knowledge property according to Definition 14. Then, the Fiat-Shamir signature scheme $\mathsf{Sig}[\Sigma]$ fulfills EUF$-$NMA security.*

Proof. Let $\mathcal{A}$ be an adversary against EUF$-$NMA, issuing at most q quantum queries to H. We show that

$$\mathrm{Adv}_{\mathsf{Sig}[\Sigma]}^{\mathsf{EUF-NMA}}(\mathcal{A}) := \Pr\Big[\mathsf{Verify}^H(pk, m, \sigma) : (pk, sk) \leftarrow \mathsf{Gen}, (m, \sigma) \leftarrow \mathcal{A}^H(pk)\Big]$$

is negligible.

Recall from Definition 19 of Fiat-Shamir signatures that the Σ-protocol Σ^* is the Σ-protocol Σ where the prover and verifier ignore the message part m of

the instance $x\|m$. A successful forgery (m, σ) is such that $V_{FS}^H(x\|m, \sigma)$ accepts the proof σ. Therefore,

$$\mathrm{Adv}_{\mathsf{Sig}[\Sigma]}^{\mathsf{EUF\text{–}NMA}}(\mathcal{A}) = \mathop{\mathbb{E}}_{(x,w)\leftarrow G} \left[\Pr_H \left[V_{FS}^H(x\|m, \sigma) : (m, \sigma) \leftarrow \mathcal{A}^H(x) \right] \right]. \quad (5)$$

Note that if Σ is a proof of knowledge, so is Σ^*. Our Corollary 16 assures that if Σ^* is a proof of knowledge, then also $\mathsf{FS}[\Sigma^*]$ is a proof of knowledge.

For fixed instance x, let X be the set of instance/message strings $x'\|m$ where $x' = x$. We apply the knowledge extractor from Definition 14 to the adaptive FS-attacker $\mathcal{A}^H(x)$ that has x hard-wired and outputs it along with a message m and the proof/signature σ: There exists a knowledge extractor $\mathcal{E}$, constants d, e and a polynomial p (all independent of x) such that

$$\Pr_H \left[x'\|m \in X \wedge V_{FS}^H(x'\|m, \sigma) : (x'\|m, \sigma) \leftarrow \mathcal{A}^H(x) \right]$$
$$\leq \left(\Pr\!\left[x'\|m \in X \wedge (x', w) \in R : (x'\|m, w) \leftarrow \mathcal{E}^{\mathcal{A}} \right] q^e p(\eta) + \mu(\eta) \right)^{1/d} \quad (6)$$

Finally, taking the expected value of (6) over the choice of the instance x according to the hard-instance generator G, we obtain that the left hand side equals $\mathrm{Adv}_{\mathsf{Sig}[\Sigma]}^{\mathsf{EUF\text{–}NMA}}(\mathcal{A})$. For the right-hand side, we can use the concavity of $(\cdot)^{1/d}$ (note that we can assume without loss of generality that $d > 1$) and apply Jensen's inequality to obtain

$$\mathop{\mathbb{E}}_{x\leftarrow G} \left[\left(\Pr\!\left[x'\|m \in X \wedge (x', w) \in R : (x'\|m, w) \leftarrow \mathcal{E}^{\mathcal{A}} \right] q^e p(\eta) + \mu(\eta) \right)^{1/d} \right]$$
$$\leq \left(\mathop{\mathbb{E}}_{x\leftarrow G} \Pr\!\left[x'\|m \in X \wedge (x', w) \in R : (x'\|m, w) \leftarrow \mathcal{E}^{\mathcal{A}} \right] q^e p(\eta) + \mu(\eta) \right)^{1/d}.$$

Note that the expected probability is the success probability of the extractor to produce a witness w matching the instance x. As long as the relation R is hard according to Definition 18, this success probability is negligible, proving our claim.

$\square$

If we wish for unforgeability *under chosen-message attack*, zero-knowledge is required as well. [Unr17] and [KLS18] contain partial results that formalize this intuition, but they were unable to derive the extractability of the non-interactive proof system. Instead, they modify the Σ-protocol to have a *lossy mode* [AFLT12], i.e. a special key-generation procedure that produces key pairs whose public keys are computationally indistinguishable from the real ones, but under which it is impossible for any (even unbounded) quantum adversary to answer correctly.

Our new result above completes these previous analyses, so that we can now state precise conditions under which a Σ-protocol gives rise to a (strongly) unforgeable Fiat-Shamir signature scheme, without the need for lossy keys.

Theorem 22 *Let* Σ *be* Σ-*protocol for some hard relation* R, *with superpolynomially sized challenge space* $\mathcal{C}$ *and the proof-of-knowledge property according to Definition 14. Assume further that* Σ *is* ε-*perfect (non-abort) honest-verifier zero-knowledge (naHVZK), has* α *bits of min entropy and computationally unique responses as defined in [KLS18]. Then,* $\mathsf{Sig}[\Sigma]$ *fulfills* $\mathsf{sEUF-CMA}$ *security.*

Proof. By Theorem 3.3 of [KLS18], we can use the naHVZK, min-entropy and computationally-unique-response properties of Σ to reduce an $\mathsf{sEUF-CMA}$ adversary to an $\mathsf{EUF-NMA}$ adversary[15]. The conclusion then follows immediately from our Theorem 21 above. □

5 Extractable Σ-protocols from quantum computationally unique responses

In the last section, we have seen that the proof-of-knowledge property of the underlying Σ-protocol is crucial for a Fiat-Shamir signature scheme to be unforgeable. In [Unr12], Unruh proved that special soundness (a witness can be constructed efficiently from two different accepting transcripts) and perfect unique responses are sufficient conditions for a Σ-protocol to achieve this property in the context of quantum adversaries. The perfect-unique-responses property is used to show that the final measurement of the Σ-protocol adversary that produces the response is nondestructive conditioned on acceptance. This property ensures that the extractor can measure the response, and then rewind "as if nothing had happened".

A natural question is therefore which other property except the arguably quite strict condition of perfect unique responses is sufficient to imply extractability together with special soundness. In [ARU14], the authors show that computationally unique responses is insufficient to replace perfect unique responses. A Σ-protocol has computationally unique responses if the verification relation V is collision-resistant from responses to commitment-challenge pairs in the sense that it is computationally hard to find two valid responses for the same commitment-challenge pair.

In [Unr16], Unruh introduced the notion of collapsingness, a quantum generalization of the collision-resistance property for hash functions. It is straightforward to generalize this notion to apply to binary relations instead of just functions.

Definition 23 (generalized from [Unr16]) *Let* $R : \mathcal{X} \times \mathcal{Y} \rightarrow \{0,1\}$ *be a relation with* $|X|$ *and* $|Y|$ *superpolynomial in the security parameter* η, *and define the following two games for polynomial-time two-stage adversaries* $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$,

Game 1 :
$(S, X, Y) \leftarrow \mathcal{A}_1,\ r \leftarrow R(X,Y),\ X \leftarrow \mathcal{M}(X),\ Y \leftarrow \mathcal{M}(Y),\ b \leftarrow \mathcal{A}_2(S,X,Y)$

Game 2 :
$(S, X, Y) \leftarrow \mathcal{A}_1,\ r \leftarrow R(X,Y),\ \qquad\qquad Y \leftarrow \mathcal{M}(Y),\ b \leftarrow \mathcal{A}_2(S,X,Y).$

---

[15] See also Theorem 25 in [Unr17] for a different proof technique.

Here, X and Y are registers of dimension $|X|$ and $|Y|$, respectively, $\mathcal{M}$ denotes a measurement in the computational basis, and applying R to quantum registers is done by computing the relation coherently and measuring it. R is called collapsing from $\mathcal{X}$ to $\mathcal{Y}$, if an adversary cannot distinguish the two experiments if the relation holds, i.e. if for all adversaries $\mathcal{A}$ it holds that

$$\left| \Pr_{\mathcal{A},\ \text{Game 1}}[r = b = 1] - \Pr_{\mathcal{A},\ \text{Game 2}}[r = b = 1] \right| \le \text{negl}(\eta). \tag{7}$$

Note that this definition is equivalent to Definition 23 in [Unr16] for functions, i.e. if $R(x, y) = 1$ if and only if $f(x) = y$ for some function f.

Via the relation that is computed by the second stage of the verifier, the collapsingness property can be naturally defined for Σ-protocols.

Definition 24 (Quantum computationally unique responses) *A* Σ-*protocol has* quantum computationally unique responses*, if the verification predicate* $V(x, \cdot, \cdot, \cdot) : \mathcal{Y} \times \mathcal{C} \times \mathcal{Z} \to \{0, 1\}$ *seen as a relation between* $\mathcal{Y} \times \mathcal{C}$ *and* $\mathcal{Z}$ *is collapsing from* $\mathcal{Z}$ *to* $\mathcal{Y} \times \mathcal{C}$*, where* $\mathcal{Y}$*,* $\mathcal{C}$ *and* $\mathcal{Z}$ *are the commitment, challenge and response spaces of the protocol, respectively.*

Intuitively, for fixed commitment-challenge pairs, no adversary should be able to determine whether a superposition over successful responses z has been measured or not. As in the case of hash functions (where collapsingness is a natural stronger quantum requirement than collision-resistance), quantum computationally unique responses is a natural stronger quantum requirement than computationally unique responses.

The following is a generalization of Theorem 9 in [Unr12] where the assumption of perfect unique responses is replaced by the above quantum computational version. Additionally, we relax the special soundness requirement to *t-soundness*, which requires that for any first message a, for uniformly random chosen challenges $c_1, \ldots, c_t$, and for any responses $z_1, \ldots, z_t$ with $V(x, a_i, c_i, z_i)$ for all $i \in \{1, \ldots, t\}$, a witness w for x can be efficiently computed except with negligible probability (over the choices of the c_i).

Theorem 25 (Generalization of Theorem 9 from [Unr12]) *Let* Π *be a* Σ-*protocol with t-soundness for some constant t and with quantum computationally unique responses. Then* Π *is a computational proof of knowledge as in Definition 14.*

The proof follows very much the proof of Theorem 9 in [Unr12], up to some small extensions; thus, we only give a proof sketch here.

Proof (sketch). We consider the following extractor $\mathcal{K}$. It runs $\mathcal{A}$ to the point where it outputs a. Then, it chooses a random challenge c_1 and sends it to $\mathcal{A}$, and obtains a response z_1 by measuring $\mathcal{A}$'s corresponding register. $\mathcal{K}$ then rewinds $\mathcal{A}$ (on the measured state!) and chooses and sends to $\mathcal{A}$ a fresh random challenge c_2, resulting in a response z_2, etc., up to obtaining response z_t. If $V(x, a_i, c_i, z_i)$

for all $i \in \{1, \ldots, t\}$ then $\mathcal{K}$ can compute w except with negligible probability by the t-soundness property; otherwise, it aborts.

It remains to analyze the probability, denoted by F below, that $V(x, a_i, c_i, z_i)$ for all i. If the Σ-protocol has *perfect* unique responses then measuring the response z is *equivalent* to measuring whether the response satisfies the verification predicate V (with respect to x, a, c). Lemma 29 in Appendix B, which generalizes Lemma 7 in [Unr12], allows us then to control the probability F by means of the probability V that $\mathcal{A}$ succeeds in convincing the verifier in an ordinary run (this holds for an arbitrary but fixed a, and on average over a by means of Jensen's inequality). If the Σ-protocol has *quantum computationally* unique responses instead, then measuring the response z is *computationally indistinguishable* from measuring whether the response satisfies the verification predicate, and so there can only be a negligible loss in the success probability of $\mathcal{K}$ compared to above. $\qed$

We expect the above theorem to be very useful in practice, for the following reason. Usually, Σ-protocols deployed in Fiat-Shamir signature schemes have computationally unique responses to ensure strong unforgeability via Theorem 22 or similar reductions. On the other hand, only very artificial separations between the notions of collision resistance and collapsingness for hash functions are known (e.g. the one presented in [Zha17]). It is therefore plausible that many Σ-protocols deployed in strongly unforgeable Fiat-Shamir signature schemes have quantum computationally unique responses as well. In the next section we take a look at a couple of examples that form the basis of some signature schemes submitted to the NIST competition for the standardization of post-quantum cryptographic schemes.

6 Application to NIST submissions

In the previous sections we gave sufficient conditions for a Fiat-Shamir signature scheme to be existentially unforgeable in the QROM. Several schemes of the Fiat-Shamir kind have made it into the second round of the NIST post-quantum standardization process. In this section we outline how our result might be applied to some of these schemes, and under which additional assumptions. We leave the problems of applying our techniques to the actual (highly optimized) signature schemes and of working out the concrete security bounds for future work.

6.1 Picnic

In order to obtain QROM-security, Picnic uses the *Unruh transform* [Unr15] instead of the Fiat-Shamir transformation, incurring a 1.6x loss in efficiency (according to [CDG+17]) compared to Fish, which is the same scheme under plain Fiat-Shamir.

The underlying sigma-protocol for these schemes is ZKB++ [CDG+17], an optimized version of ZKBoo [GMO16], which uses an arbitrary one-way function

ϕ, a commitment scheme COM and a multi-party computation protocol to prove knowledge of a secret key. Roughly, a prover runs the multi-party protocol 'in its head' (i.e. simulates the three agents from the protocol, see [IKOS07]) to compute $pk := \phi(sk)$. Only a prover who knows the secret key can produce the correct view of all three agents, but the public key suffices to verify the correctness of two of the views. In the first round, the prover uses COM to commit to all three views separately, and sends these commitments to the verifier. The verifier replies with a random challenge $i \in \{1, 2, 3\}$, to which the prover in turn responds by opening the i-th and $i+1$-th commitment.

ZKBoo does not specify a concrete commitment scheme for COM. A natural option is to commit by hashing the input together with some random bits.

Corollary 26 Sig[ZKBoo] *is strongly existentially unforgeable in the QROM when* COM *is instantiated with a hash function* H.

Proof. If we treat H as a quantum-accessible random oracle, then H is collapsing by [Unr16]. Since the response of the prover in the third round consists only of openings to the commitments c_i, c_{i+1}, i.e. preimages of c_i and c_{i+1} under H, and since collapsingness is closed under concurrent composition [Feh18], the collapsingness of H implies that ZKBoo has quantum computational unique responses. ZKBoo further has 3-soundness, and thus the claim follows using Theorems 25 and 22. □

ZKB++ improves on ZKBoo by introducing optimizations specific to the signature context, which complicate the analysis of the overall scheme. We therefore leave the adaption of Corollary 26 to ZKB++ and Fish for future work.

6.2 Lattice-based Fiat-Shamir signature schemes – CRYSTALS-Dilithium and qTesla

In [Lyu09] and [Lyu12], Lyubashevsky developed a Fiat-Shamir signature scheme based on (ring) lattice assumptions. In the following, we explain the lattice case and mention ring-based lattice terms in parentheses. The underlying sigma protocol, which forms the basis of the NIST submissions CRYSTALS-Dilithium and qTesla, can be roughly described as follows. The instance is given by a key pair $((A, T), S)$, with $T = AS$. Here, A and S are matrices of appropriate dimensions over a finite field (polynomials of appropriate degree), and S is *small*. For the first message to the verifier, the prover selects a random short vector (small polynomial) y, and sends over Ay. The second message, from the verifier to the prover, is a random vector (polynomial) c with entries (coefficients) in $\{-1, 0, 1\}$ and a small Hamming weight. The third message, i.e. the response of the prover, is $z = Sc + y$, which is short (small) as well. The prover actually sends z only with a particular probability, which is chosen so as to make the distribution of (sent) z independent of S. Otherwise, it aborts and tries again. Verification is done by checking whether z is indeed short (small), and whether $Az - Ay = Tc$. Let us denote this protocol by LatticeΣ. In the following we

restrict our attention to the lattice case, but we expect that one can do a similar analysis for the ring-based schemes.

The security of the scheme is, in the lattice case, based on the SHORT INTE-GER SOLUTION (SIS) problem, which essentially guarantees that it is hard to find an integral solution to a linear system that has a small norm. The computationally unique responses property for the simple Σ-protocol described above, in fact, follows directly from SIS: If one can find a vector c and two short vectors x_i, $i = 1, 2$ such that $Ax_0 = c = Ax_1$, then the difference $x = x_1 - x_0$ is a short solution to the linear system $Ax = 0$.

Another way to formulate the computationally unique responses property for the above Σ-protocol is as follows. Let $S \subset \mathbb{F}_q^n$ be the set of short vectors. Let $f_A : S \to \mathbb{F}_q^m$ be the restriction to S of the linear map given by the matrix $A \in \mathbb{F}_q^{m \times n}$. The Σ-protocol above has computationally unique responses if and only if f_A is collision resistant. As pointed out at the end of Section 5, the known examples that separate the collision resistance and collapsingness properties are fairly artificial. Hence it is a natural to assume that f_A is collapsing as well.

Assumption 27 *For m, n and q polynomial in the security parameter η, the function family f_A keyed by a uniformly random matrix $A \in \mathbb{F}_q^{m \times n}$ is collapsing.*

Under Assumption 27, LatticeΣ has quantum computational unique responses, and hence gives rise to an unforgeable Fiat-Shamir signature scheme.

Corollary 28 *Under Assumption 27, $\mathsf{Sig}[\mathsf{Lattice\Sigma}]$ is strongly existentially unforgeable in the QROM.*

As mentioned at the end of the introduction, in their concurrent and independent work [LZ19], Lie and Zhandry show that f_A satisfies their notion of *weak-collapsingness* (assuming hardness of LWE), which roughly says that there is some non-negligible probability that the adversary *does not* notice a measurement. Weak-collapsingness implies a similarly weakened variant of our property 'quantum computational responses', which is still sufficient to let the proof of Theorem 25 go through, albeit with a worse but still non-negligible success probability for the knowledge-extractor.

7 Acknowledgement

References

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

AFLT12. Michel Abdalla, Pierre-Alain Fouque, Vadim Lyubashevsky, and Mehdi Tibouchi. Tightly-secure signatures from lossy identification schemes. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, pages 572–590, Berlin, Heidelberg, 2012. Springer.

ARU14. A. Ambainis, A. Rosmanis, and D. Unruh. Quantum attacks on classical proof systems: The hardness of quantum rewinding. In *2014 IEEE 55th Annual Symposium on Foundations of Computer Science*, pages 474–483, Oct 2014.

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

BDK⁺18. J. Bos, L. Ducas, E. Kiltz, T. Lepoint, V. Lyubashevsky, J. M. Schanck, P. Schwabe, G. Seiler, and D. Stehle. CRYSTALS - Kyber: A CCA-secure module-lattice-based KEM. In *2018 IEEE European Symposium on Security and Privacy (EuroS P)*, pages 353–367, April 2018.

BG93. Mihir Bellare and Oded Goldreich. On defining proofs of knowledge. In Ernest F. Brickell, editor, *Advances in Cryptology — CRYPTO' 92*, pages 390–420, Berlin, Heidelberg, 1993. Springer.

BR93. Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *Proceedings of the 1st ACM conference on Computer and communications security*, pages 62–73. ACM, 1993.

CDG⁺17. Melissa Chase, David Derler, Steven Goldfeder, Claudio Orlandi, Sebastian Ramacher, Christian Rechberger, Daniel Slamanig, and Greg Zaverucha. Post-quantum zero-knowledge and signatures from symmetric-key primitives. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 1825–1842, New York, NY, USA, 2017. ACM.

CGH04. Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *Journal of the ACM*, 51(4):557–594, jul 2004.

Dam10. Ivan Damgard. On sigma-protocols, lecture notes, faculty of science aarhus university, department of computer science, 2010.

DFG13. Özgür Dagdelen, Marc Fischlin, and Tommaso Gagliardoni. The fiat–shamir transformation in a quantum world. In Kazue Sako and Palash Sarkar, editors, *Advances in Cryptology - ASIACRYPT 2013*, pages 62–81, Berlin, Heidelberg, 2013. Springer.

ES15. Edward Eaton and Fang Song. Making existential-unforgeable signatures strongly unforgeable in the quantum random-oracle model. In *10th Conference on the Theory of Quantum Computation, Communication and Cryptography*, page 147, 2015.

Feh18. Serge Fehr. Classical proofs for the quantum collapsing property of classical hash functions. *Theory of Cryptography Conference - TCC2018, volume 11240 of Lecture Notes in Computer Science*, pages 315–338, 2018.

FKMV12. Sebastian Faust, Markulf Kohlweiss, Giorgia Azzurra Marson, and Daniele Venturi. On the Non-malleability of the Fiat-Shamir Transform. In *Indocrypt 2012*, volume 7668 LNCS, pages 60–79. 2012.

FS87. Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In Andrew M. Odlyzko, editor, *Advances in Cryptology — CRYPTO' 86*, pages 186–194, Berlin, Heidelberg, 1987. Springer.

GMO16. Irene Giacomelli, Jesper Madsen, and Claudio Orlandi. Zkboo: Faster zero-knowledge for boolean circuits. In *25th USENIX Security Symposium (USENIX Security 16)*, pages 1069–1083, Austin, TX, 2016. USENIX Association.

IKOS07. Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Zero-knowledge from secure multiparty computation. *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing - STOC '07*, page 21, 2007.

KLS18. Eike Kiltz, Vadim Lyubashevsky, and Christian Schaffner. A concrete treatment of Fiat-Shamir signatures in the quantum random-oracle model. In Jesper Buus Nielsen and Vincent Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2018*, pages 552–586, Cham, 2018. Springer.

Lyu09. Vadim Lyubashevsky. Fiat-Shamir with aborts: Applications to lattice and factoring-based signatures. In Mitsuru Matsui, editor, *Advances in Cryptology – ASIACRYPT 2009*, pages 598–616, Berlin, Heidelberg, 2009. Springer.

Lyu12. Vadim Lyubashevsky. Lattice signatures without trapdoors. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, pages 738–755, Berlin, Heidelberg, 2012. Springer.

LZ19. Qipeng Liu and Mark Zhandry. Revisiting post-quantum fiat-shamir. Cryptology ePrint Archive, Report 2019/262, 2019. https://eprint.iacr.org/2019/262.

SXY18. Tsunekazu Saito, Keita Xagawa, and Takashi Yamakawa. Tightly-secure key-encapsulation mechanism in the quantum random oracle model. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 520–551. Springer, 2018.

Unr12. Dominique Unruh. Quantum proofs of knowledge. In David Pointcheval and Thomas Johansson, editors, *Advances in Cryptology – EUROCRYPT 2012*, pages 135–152, Berlin, Heidelberg, 2012. Springer.

Unr14. Dominique Unruh. Quantum position verification in the random oracle model. In Juan A. Garay and Rosario Gennaro, editors, *Advances in Cryptology – CRYPTO 2014*, pages 1–18, Berlin, Heidelberg, 2014. Springer.

Unr15. Dominique Unruh. Non-interactive zero-knowledge proofs in the quantum random oracle model. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015*, pages 755–784, Berlin, Heidelberg, 2015. Springer.

Unr16. Dominique Unruh. Computationally binding quantum commitments. In Marc Fischlin and Jean-Sébastien Coron, editors, *Advances in Cryptology – EUROCRYPT 2016*, pages 497–527, Berlin, Heidelberg, 2016. Springer.

Unr17. Dominique Unruh. Post-quantum security of Fiat-Shamir. In Tsuyoshi Takagi and Thomas Peyrin, editors, *Advances in Cryptology – ASIACRYPT 2017*, pages 65–95, Cham, 2017. Springer.

Zha12. Mark Zhandry. How to construct quantum random functions. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 679–687. IEEE, oct 2012.

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

Zha17.   Mark Zhandry. Quantum lightning never strikes the same state twice. http://arxiv.org/abs/1711.02276, 2017.

Zha18.   Mark Zhandry. How to record quantum queries, and applications to quantum indifferentiability. Cryptology ePrint Archive, Report 2018/276, 2018. https://eprint.iacr.org/2018/276.

A   Proof of Lemma 12 and 15

Proof (of Lemma 12). Let $\mathcal{A}$ be an adaptive Σ-protocol adversary, producing x and a in the first stage, and z in the second stage. We then consider the following algorithms. $\mathcal{A}_{init}$ runs the first stage of $\mathcal{A}$ (using the same initial state), outputting x and a. Let $|\psi_{x,a}\rangle$ be the corresponding internal state at this point. Furthermore, for any possible x and a, $\mathcal{A}_{x,a}$ is the following static Σ-protocol adversary. Its initial state is $|\psi_{x,a}\rangle|a\rangle$ and in the first stage it simply outputs a, and in the second stage, after having received the verifier's challenge, it runs the second stage of $\mathcal{A}$. We then see that

$$\Pr\big[x \notin \mathcal{L} \wedge v = accept : (x,v) \leftarrow \langle \mathcal{A}, \mathcal{V} \rangle\big]$$

$$= \sum_{x_\circ \notin \mathcal{L}} \Pr\big[x = x_\circ \wedge v = accept : (x,v) \leftarrow \langle \mathcal{A}, \mathcal{V} \rangle\big]$$

$$= \sum_{x_\circ \notin \mathcal{L}} \sum_a \Pr\big[\mathcal{A}_{init} = (x_\circ, a)\big] \Pr\big[\langle \mathcal{A}_{x_\circ,a}, \mathcal{V}(x_\circ)\rangle = accept\big].$$

Since $\Pr\big[\langle \mathcal{A}_{x_\circ,a}, \mathcal{V}(x_\circ)\rangle = accept\big]$ is bounded by a negligible function, given that $\mathcal{A}_{x,a}$ is a (quantum polynomial-time/unbounded) static adversary, the claim follows.   □

Proof (of Lemma 15). Let $\mathcal{A}$ be an adaptive Σ-protocol adversary, producing x and a in the first stage, and z in the second stage. We construct a black-box knowledge extractor $\mathcal{K}_{ad}$ that works for any such $\mathcal{A}$. In a first step, $\mathcal{K}_{ad}^{\mathcal{A}}$ runs the first stage of $\mathcal{A}$ using the black-box access to $\mathcal{A}$ (and having access to the initial state of $\mathcal{A}$). Below, we call this first stage of $\mathcal{A}$ as $\mathcal{A}_{init}$. This produces x and a, and we write $|\psi_{x,a}\rangle$ for the corresponding internal state. Then, it runs $\mathcal{K}_{na}^{\mathcal{A}^{x,a}}$, where $\mathcal{K}_{na}$ is the knowledge extractor guaranteed to exist for static adversaries, and $\mathcal{A}^{x,a}$ is the static adversary that works as follows. It's initial state is $|\psi_{x,a}\rangle|a\rangle$ and in the first stage it simply outputs a, and in the second stage it runs the second stage of $\mathcal{A}$ on the state $|\psi_{x,a}\rangle$. Note that having obtained x and a and the state $|\psi_{x,a}\rangle$ as first step of $\mathcal{K}_{ad}^{\mathcal{A}}$, $\mathcal{K}_{na}^{\mathcal{A}^{x,a}}$ can then be executed with black box access to (the second stage of) $\mathcal{A}$. For any subset $X \subseteq \mathcal{X}$, we now see that

$$\Pr\big[x \in X \wedge (x,w) \in R : (x,w) \leftarrow \mathcal{K}_{ad}^{\mathcal{A}}\big]$$

$$= \sum_{x \in X} \sum_a \Pr\big[\mathcal{A}_{init} = (x,a)\big] \Pr\big[(x,w) \in R : w \leftarrow \mathcal{K}_{na}^{\mathcal{A}^{x,a}}\big]$$

$$\geq \sum_{x \in X} \sum_{a} \Pr\big[\mathcal{A}_{init} = (x, a)\big] \cdot \frac{1}{p(\eta)} \cdot \Pr\big[\langle \mathcal{A}^{x,a}, \mathcal{V}(x)\rangle = accept\big]^d - \kappa(\eta)$$

$$\geq \frac{1}{p(\eta)} \Big(\sum_{x \in X} \sum_{a} \Pr\big[\mathcal{A}_{init} = (x, a)\big] \Pr\big[\langle \mathcal{A}^{x,a}, \mathcal{V}(x)\rangle = accept\big] \Big)^d - \kappa(\eta)$$

$$= \frac{1}{p(\eta)} \Pr\big[x \in X \wedge v = 1 : (x, v) \leftarrow \langle \mathcal{A}^{x,a}, \mathcal{V}(x)\rangle\big]^d - \kappa(\eta),$$

where the first inequality is because of the static proof-of-knowledge property, and the second is Jensen's inequality, noting that we may assume without loss of generality that $d \geq 1$. $\qquad\square$

B   Generalization of Lemma 7 from [Unr12]

Lemma 29 *Let $P_1, \ldots, P_n$ be projections and $|\psi\rangle$ a state vector, and set*

$$V := \frac{1}{n} \sum_i \langle \psi | P_i | \psi \rangle = \frac{1}{n} \sum_i \| P_i |\psi\rangle \|^2 \qquad and \qquad F := \frac{1}{n^t} \sum_{i_1 \cdots i_t} \| P_{i_t} \cdots P_{i_1} |\psi\rangle \|^2.$$

Then $F \geq V^{2t-1}$.

The case $t = 2$ was proven in [Unr12, Lemma 7]. We show here how to extend the proof to $t = 3$; the general case works along the same lines.

Proof (of the case $t = 3$). For convenience, set $A := \frac{1}{n} \sum_i P_i$ and $|\psi_{ijk}\rangle := P_k P_j P_i |\psi\rangle$. Then, using convexity of the function $x \mapsto x^5$ to argue the first inequality, we get

$$V^5 = (\langle \psi | A | \psi \rangle)^5 = \langle \psi | A^5 | \psi \rangle = \frac{1}{n^5} \sum_{ijk\ell m} \langle \psi | P_i P_j P_k P_\ell P_m | \psi \rangle$$

$$= \frac{1}{n^5} \sum_{ijk\ell m} \langle \psi_{ijk} | \psi_{m\ell k} \rangle = \frac{1}{n} \sum_k \Big(\frac{1}{n^2} \sum_{ij} \langle \psi_{ijk} | \Big) \Big(\frac{1}{n^2} \sum_{\ell m} | \psi_{m\ell k} \rangle \Big)$$

$$= \frac{1}{n} \sum_k \Big\| \frac{1}{n^2} \sum_{ij} | \psi_{ijk} \rangle \Big\|^2 \leq \frac{1}{n^3} \sum_{ijk} \| | \psi_{ijk} \rangle \|^2 = F,$$

where the last inequality is Claim 2 in the proof of Lemma 7 in [Unr12]. $\qquad\square$