Online-Extractability in the Quantum Random-Oracle Model[⋆]

Jelle Don[1], Serge Fehr[1,2], Christian Majenz[3], and Christian Schaffner[4,5]

[1] Centrum Wiskunde & Informatica (CWI), Amsterdam, Netherlands
[2] Mathematical Institute, Leiden University, Netherlands
[3] Cyber Security Section, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby, Denmark
[4] Informatics Institute, University of Amsterdam, Amsterdam, Netherlands
[5] QuSoft, Amsterdam, Netherlands
`jelle.don@cwi.nl`, `serge.fehr@cwi.nl`, `chmaj@dtu.dk`, `c.schaffner@uva.nl`

Abstract. We show the following generic result: When a quantum query algorithm in the quantum random-oracle model outputs a classical value t that is promised to be in some tight relation with $H(x)$ for some x, then x can be efficiently extracted with almost certainty. The extraction is by means of a suitable simulation of the random oracle and works *online*, meaning that it is *straightline*, i.e., without rewinding, and *on-the-fly*, i.e., during the protocol execution and (almost) without disturbing it.
The technical core of our result is a new commutator bound that bounds the operator norm of the commutator of the unitary operator that describes the evolution of the compressed oracle (which is used to simulate the random oracle above) and of the measurement that extracts x.
We show two applications of our generic online extractability result. We show *tight* online extractability of commit-and-open Σ-protocols in the quantum setting, and we offer the first complete post-quantum security proof of the *textbook* Fujisaki-Okamoto transformation, i.e, without adjustments to facilitate the proof, including concrete security bounds.

1 Introduction

Background. *Extractability* plays an important role in cryptography. In an extractable protocol, an algorithm $\mathcal{A}$ sends messages that depend on some secret s, and while the secret remains private in an honest run of the protocol, an *extractor* can learn s via some form of enhanced access to $\mathcal{A}$. The probably most prominent example is that of zero-knowledge *proofs* (or *arguments*) *of knowledge*, for which, by definition, there must exist an extractor that manages to extract a witness from any successful prover. Another example are *extractable commitments*, which have a wide range of applications. Hash-based extractable commitments are extremely simple to construct and prove secure in the random-oracle model (ROM) [22]. Indeed, when the considered hash function H is modelled as a random oracle, the hash input x for the commitment $c = H(x)$, where $x = s\|r$

[⋆] Full version available at `https://eprint.iacr.org/2021/280`.

consists of the actual secret s and randomness r, can be extracted simply by finding a query x to the random oracle that yielded c as an output.

The general notion of extractability comes in different flavors. The most well-known example is extraction by *rewinding*. Here, the extractor is allowed to run $\mathcal{A}$ several times, on the same private input and using different randomness. This is the notion usually considered in the context of proofs/arguments of knowledge. In some contexts, extraction via rewinding access is not possible. For example, the UC security model prohibits the simulator to rewind the adversary. In other occasions, rewinding may be possible but not desirable due to a loss of efficiency, which stems from having to run $\mathcal{A}$ multiple times. In comparison, so-called *straightline* extraction works with a single ordinary run of $\mathcal{A}$, without rewinding. Instead, the extractor is then assumed to know some trapdoor information, or it is given enhanced control over some part of the setting. For instance, in the above construction of an extractable commitment, the extractor is given "read access" to $\mathcal{A}$'s random-oracle queries.

Another binary criterion is whether the extraction takes place *on-the-fly*, i.e., during the run of the protocol, or *after-the-fact*, i.e., at the end of the execution. For instance, in the context of proving CCA security for an encryption scheme, to simulate decryption queries without knowing the secret key, it is necessary to extract the plaintext for a queried ciphertext on-the-fly; otherwise, the attacker may abort and not produce the output for which the reduction is waiting.

The extractability of our running example of an extractable commitment in the ROM is *both*, straightline and on-the-fly; we refer to this combination as *online* extraction. This notion is what we are aiming for: online extractability of (general) hash-based commitments, but now with *post-quantum security*.

For post-quantum security, the ROM needs to be replaced by the *quantum* random-oracle model (QROM) [4] to reflect the fact that attackers can implement hash functions on a quantum computer. Here, adversaries have quantum superposition access to the random oracle. Many ROM techniques fail in the QROM due to fundamental features of quantum information, such as the so-called *no-cloning principle*. In particular, it is impossible to maintain a query transcript (a fact sometimes referred to as the *recording barrier*), and so one cannot simply "search for a query x to the random oracle", as was exploited for the (classical) RO-security of the extractable-commitment example.

A promising step in the right direction is the compressed-oracle technique, recently developed by Zhandry [27]. This technique enables to maintain *some sort* of a query transcript, but now in the form of a quantum state. This state can be inspected via quantum measurements, offering the possibility to learn some information about the interaction history of the random oracle. However, since quantum measurements disturb the state to which they are applied, and this disturbance is often hard to control, this inspection of the query transcript can *per-se*, i.e., without additional argumentation, only be done at the end of the execution (see the Related Work paragraph for more on this).

Our Results. Our main contribution is the following generic extractability result in the QROM: We consider an arbitrary quantum query algorithm $\mathcal{A}$ in the QROM, which announces during its execution some classical value t that is supposed to be equal to $f(x, H(x))$ for some x. Here, f is an arbitrary fixed function, subject to that it must tie t sufficiently to x and $H(x)$, e.g., there must not be too many y's with $f(x, y) = t$; a canonical example is the function $f(x, y) = y$ so that t is supposed to be $t = H(x)$. In general, it is helpful to think of $t = f(x, H(x))$ as a commitment to x. We then show that x can be *efficiently extracted* with almost certainty. The extraction works *online* and is by means of a simulator $\mathcal{S}$ that simulates the quantum random oracle, but which additionally offers an *extraction interface* that produces a guess $\hat{x}$ for x when queried with t. The simulation is statistically indistiguishable from the real quantum random oracle, and $\hat{x}$ is such that whenever $\mathcal{A}$ outputs x with $f(x, H(x)) = t$ at some later point, $\hat{x} = x$ holds except with negligible probability, while $\hat{x} = \emptyset$ (some special symbol) indicates that $\mathcal{A}$ will not be able to output such an x.

The simulator $\mathcal{S}$ simulates the random oracle using Zhandry's compressed-oracle technique, and extraction is done via a suitable measurement of the compressed oracle's internal register. The technical core of our result is a new bound for the operator norm $\|[O, M]\|$ of the commutator of O, the unitary operator that describes the evolution of the compressed oracle, and of M, the extraction measurement. This bound allows us to show that the extraction measurement only negligibly disturbs the behavior of the compressed oracle, and so can indeed be performed *on-the-fly*. At first glance, our technical result has some resemblance with Lemma 39 in [27], which also features an almost-commutativity property, and, indeed, with Lemma 3 we use (a reformulated version of) Lemma 39 in [27] as a first step in our proof. However, the challenging part of the main proof consists of lifting the almost-commutativity property of the "local" projectors $\mathit{\Pi}^x$ from Lemma 3 to the "global" measurement M.

We emphasize that even though the existence of the simulator with its extraction interface is proven using the compressed-oracle technique, our presentation is in terms of a black-box simulator $\mathcal{S}$ with certain interfaces and with certain promises on its behavior, abstracting away all the (mainly internal) quantum workings. This makes our generic result applicable (e.g. for the applications discussed below) without the need to understand the underlying quantum aspects.

A first concrete application of our generic result is in the context of so-called commit-and-open Σ-protocols. These are (typically honest-verifier zero-knowledge) interactive proofs of a special form, where the prover first announces a list of commitments and is then asked to open a subset of them, chosen at random by the verifier. We show that, when implementing the commitments with a typical hash-based commitment scheme (like committing to s by $H(s\|r)$ with a random r), such Σ-protocols allow for *online* extraction of a witness in the QROM, with a *smaller security loss* than witness extraction via rewinding.

Equipped with our extractable RO-simulator $\mathcal{S}$, the idea for the above online extraction is very simple: we simulate the random oracle using $\mathcal{S}$ and use its extraction interface to extract the prover's commitments from the first message of the Σ-protocol. As we work out in detail, this procedure gives rise to an online witness extractor that has a polynomial additive overhead in running time compared to the considered prover, and that outputs a valid witness with a probability that is *linear* in the difference of the prover's success probability and the trivial cheating probability, up to an additive error. Using rewinding techniques, on the other hand, incurs a *square-root* loss in success probability classically and a *cube-root* loss quantumly for special-sound Σ-protocols, and typically an even worse loss in case of weaker soundness guarantees, like a k-th-root loss classically and a $(2k+1)$-th-root loss quantumly for k-sound protocols.

Our second application is a security reduction for the Fujisaki-Okamoto (FO) transformation. We offer the first complete post-quantum security proof of the *textbook* FO transformation [13], with concrete security bounds. Most of the prior post-quantum security proofs had to adjust the transformation to facilitate the proof (like [16]); those security proofs either consider a FO variant that employs an *implicit-rejection* routine, or have to resort to an additional "key confirmation" hash [24] that is appended to the ciphertex, thus increasing the ciphertext size. The *unmodified* FO transformation was analyzed in [27] and [18]; however, as we explain in detail in Appendix A of the full version, the given post-quantum security proofs are incomplete, both having the same gap.

Beyond its theoretical relevance of showing that no adjustment is necessary, the security of the original unmodified FO transformation with explicit rejection in particular ensures that the conservative variant with implicit rejection remains secure even when the decapsulation algorithm is not implemented carefully enough and admits a side-channel attack that reveals information on whether the submitted ciphertext is valid or not.

The core idea of our proof for the textbook FO transformation is to use the extractability of the RO-simulator to handle the decryption queries. Indeed, letting $f(x,y)$ be the encryption $Enc_{pk}(x;y)$ of the message x under the randomness y, a "commitment" $t = f(x, H(x))$ is then the encryption of x under the derandomized scheme, and so the extraction interface recovers x.

Related Work. The compressed-oracle technique has proven to be a powerful tool for lifting classical ROM proofs to the QROM setting. Examples are [19, 10] for quantum query complexity lower bounds and [15] for space-time tradeoff bounds, [9] for the security of succinct arguments, [1] for quantum-access security, and [3] for a new "double-sided" O2H lemma in the context of the FO transformation. In these cases, the argument exploits the possibility to extract information on the interaction history of the algorithm $\mathcal{A}$ and the (compressed) oracle *after-the-fact*, i.e., at the very end of the run.

In addition, some tools have been developed that allow measuring (the internal state of) the compressed oracle *on-the-fly*, which then causes the state, and thus the behavior of the oracle, to change. In some cases, the disturbance is significant yet asymptotically good enough for the considered application, causing "only" a polynomial blow-up of a negligible error term, as, e.g., in [20] for proving the security of the Fiat-Shamir transformation. In other cases [27, 11], it is shown for some limited settings that certain measurements do not render the simulation of the random oracle distinguishable (except for negligible advantage). The indifferentiability result in [11], for example, only uses measurements that have an almost certain outcome.

In particular, [27] contains a security reduction for the FO transformation that implicitly uses a measurement similar to the one we analyze in Section 3, but without analyzing the disturbance it causes. We discuss this in more detail in Appendix A of the full version. The same gap exists in follow-up work by Katsumata, Kwiatkowski, Pintore and Prest [18], who follow the FO proof outline from [27].

2 Preliminaries

For Sect. 3 and 4 (only), we assume some familiarity with the mathematics of quantum information as well as with the compressed-oracle technique of [27]. Below, we summarize the concepts that will be of particular importance. For a function or algorithm f, we write $\mathrm{Time}[f]$ to denote the time complexity of (an algorithm computing) f.

2.1 Mathematical Preliminaries

Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space. We use the standard bra-ket notation for the vectors in $\mathcal{H}$ and its dual space. We write $\||\varphi\rangle\|$ for the (Euclidean) norm $\||\varphi\rangle\| = \sqrt{\langle\varphi|\varphi\rangle}$ of $|\varphi\rangle \in \mathcal{H}$. Furthermore, for an operator $A \in \mathcal{L}(\mathcal{H})$, we denote by $\|A\|$ its *operator norm*, i.e., $\|A\| = \max_{|\psi\rangle} \|A|\psi\rangle\|$, where the max is over all $|\psi\rangle \in \mathcal{H}$ with norm 1. We assume the reader to be familiar with basic properties of these norms, like triangle inequality, $\||\varphi\rangle\langle\psi|\| = \||\varphi\rangle\|\||\psi\rangle\|$, $\|A|\varphi\rangle\| \le \|A\|\||\varphi\rangle\|$, $\|AB\| \le \|A\|\|B\|$, etc. Less well known may be the inequality[1]

$$\||\varphi\rangle\langle\psi| - |\psi\rangle\langle\varphi|\| \le \||\varphi\rangle\|\||\psi\rangle\| \,. \tag{1}$$

Another basic yet important property that we will exploit is the following.

Lemma 1. *Let A and B be operators in $\mathcal{L}(\mathcal{H})$ with $A^\dagger B = 0$ and $AB^\dagger = 0$. Then, $\|A + B\| \le \max\{\|A\|, \|B\|\}$.*

Exploiting that $\|A \otimes B\| = \|A\|\|B\|$, the following is a direct consequence.

Corollary 1. *If $A = \sum_x |x\rangle\langle x| \otimes A^x$ then $\|A\| \le \max_x \|A^x\|$.*

Definition 1. *For $A, B \in \mathcal{L}(\mathcal{H})$, the* commutator *is $[A, B] := AB - BA$.*

Some obvious properties of the commutator are:

$$[B, A] = -[A, B] = [A, \mathbb{1} - B] \,, \quad [A \otimes \mathbb{1}, B \otimes C] = [A, B] \otimes C \tag{2}$$

$$\text{and} \qquad [AB, C] = A[B, C] + [A, C]B \,. \tag{3}$$

[1] It is immediate for normalized $|\phi\rangle$ and $|\psi\rangle$ when expanding both vectors in an orthonormal basis containing $|\varphi\rangle$ and $\frac{|\psi\rangle - \langle\varphi|\psi\rangle|\varphi\rangle}{\sqrt{1-|\langle\varphi|\psi\rangle|^2}}$, and the general case then follows by homogeneity of the norms.

Combining the right equality in (2) with basic properties of the operator norm, if $\|C\| \leq 1$, e.g., if C is a unitary of a projection, we have

$$\|[A \otimes \mathbb{1}, B \otimes C]\| = \|[A, B]\| \|C\| \leq \|[A, B]\| \,. \tag{4}$$

It is common in quantum information science to write A_X to emphasize that the operator A acts on *register* X, i.e., on a Hilbert space $\mathcal{H}_X$ that is labeled by the X. It is then understood that when applied to registers X and Y, say, A_X acts as A on register X and as identity $\mathbb{1}$ on register Y, i.e., A_X is identified with $A_X \otimes \mathbb{1}_Y$. Property (4) would then e.g. be written as $\|[A_X, B_X \otimes C_Y]\| \leq \|[A_X, B_X]\|$. In this work, we will write or not write these subscripts emphasizing the register(s) at our convenience; typically we write them when the argument crucially depends on the registers, and we may omit them otherwise.

Another important matrix norm is the *trace norm*, $\|A\|_1 = \mathrm{tr}\big[\sqrt{A^\dagger A}\big]$. For density matrices ρ and σ, the *trace distance* is defined as $\delta(\rho, \sigma) = \frac{1}{2}\|\rho - \sigma\|_1$. By equation (9.110) in [21], for any norm-1 vectors $|\varphi\rangle$ and $|\psi\rangle$,

$$\delta(|\varphi\rangle\langle\varphi|, |\psi\rangle\langle\psi|) \leq \| |\varphi\rangle - |\psi\rangle \| \,. \tag{5}$$

For probability distributions p and q, we write $\delta(p, q)$ for the *total variational distance*; this is justified as $\|\rho_0 - \rho_1\|_1 = \delta(p_0, q_1)$ for $\rho_i = \sum_x p_i(x)|x\rangle\langle x|$, $i = 0, 1$. In case of a hybrid classical-quantum state, consisting of a randomized classical value x that follows a distribution p and of a quantum register W with a state ρ_W^x that depends on x, we write $[x, W] = \sum_x p(x)|x\rangle\langle x| \otimes \rho_W^x$.[2] When the distribution p and the density operators ρ_W^x are implicitly given by a game (or experiment) $\mathcal{G}$ then we may write $[x, W]_\mathcal{G}$, in particular when considering and comparing different such games. For instance, we write $\delta\big([x, W]_\mathcal{G}, [x, W]_{\mathcal{G}'}\big)$ for the trace distance of the respective density matrices in game $\mathcal{G}$ and in game $\mathcal{G}'$.

2.2 The (Compressed) Random Oracle

The (quantum) random-oracle model. In the *random-oracle model*, a cryptographic hash function $H : \mathcal{X} \to \mathcal{Y}$ is treated as an oracle RO that the adversary needs to query on $x \in \mathcal{X}$ to learn $H(x)$. The random oracle answers these queries by means of a uniformly random function $H : \mathcal{X} \to \mathcal{Y}$. For concreteness, we restrict here to $\mathcal{Y} = \{0, 1\}^n$; on the other hand, we do not further specify the domain $\mathcal{X}$ except that we assume it to have an efficiently computable order, so one may well think of $\mathcal{X}$ as $\mathcal{X} = \{1, \ldots, M\}$ for some positive $M \in \mathbb{Z}$ or as bit strings of bounded size. We then often write $RO(x)$ instead of $H(x)$ to emphasize that $H(x)$ is obtained by querying the random oracle and/or to emphasize the randomized nature of H. In the *quantum* random oracle model (QROM), a quantum algorithm $\mathcal{A}$ may make *superposition queries* to RO, meaning that the oracle acts as unitary $|x\rangle|y\rangle \mapsto |x\rangle|y \oplus H(x)\rangle$. The QROM still admits *classical* queries, which are queries with the query register set to $|x\rangle|0\rangle$ for some x, and the second register is subsequently measured to obtain the classical output y.

[2] In this equality and at other occasions, we use the same letter, here x, for the considered *random variable* as well as for a *particular value*.

The compressed oracle. We recall here (some version of) the *compressed* oracle, as introduced in [27], which offers a powerful tool for QROM proofs. For this purpose, we consider the multi-register $D = (D_x)_{x\in\mathcal{X}}$, where the state space of D_x is given by $\mathcal{H}_{D_x} = \mathbb{C}[\{0,1\}^n \cup \{\bot\}]$, meaning that it is spanned by an orthonormal set of vectors $|y\rangle$ labelled by $y \in \{0,1\}^n \cup \{\bot\}$. The initial state is set to be $|\bot\rangle_D := \bigotimes_x |\bot\rangle_{D_x}$. Consider the unitary F defined by

$$F|\bot\rangle = |\phi_0\rangle\ , \quad F|\phi_0\rangle = |\bot\rangle \quad \text{and} \quad F|\phi_y\rangle = |\phi_y\rangle\ \forall\, y \in \{0,1\}^n \setminus \{0^n\}\,,$$

where $|\phi_y\rangle := H|y\rangle$ with H the Hadamard transform on $\mathbb{C}[\{0,1\}^n] = (\mathbb{C}^2)^{\otimes n}$. Exploiting the relation $|y\rangle = 2^{-n/2}\sum_\eta (-1)^{\eta\cdot y}|\phi_\eta\rangle$, we see that

$$F|y\rangle = |y\rangle + 2^{-n/2}\left(|\bot\rangle - |\phi_0\rangle\right)\ . \tag{6}$$

When the oracle is queried, a unitary O_{XYD}, acting on the query registers X and Y and the oracle register D, is applied, given by

$$O_{XYD} = \sum_x |x\rangle\langle x|_X \otimes O^x_{YD_x} \qquad \text{with} \qquad O^x_{YD_x} = F_{D_x}\mathrm{CNOT}_{YD_x}F_{D_x}\ , \tag{7}$$

where $\mathrm{CNOT}|y\rangle|y_x\rangle = |y \oplus y_x\rangle|y_x\rangle$ for $y, y_x \in \{0,1\}^n$, and $\mathrm{CNOT}|y\rangle|\bot\rangle = |y\rangle|\bot\rangle$.

As long as no other operations are applied to the state of D, the compressed oracle exactly simulates the quantum random oracle. Also, the support of the state of D_x then remains orthogonal to $|\phi_0\rangle$ for all x. However, these properties may change when, e.g., measurements are performed on D. The oracle may then behave differently than the quantum random oracle, and the state of D may have a non-trivial overlap with $|\phi_0\rangle$. Note that, by the convention on CNOT to act trivially for control registers in state $|\bot\rangle$, it holds that $O^x_{YD_x}|y\rangle|\phi_0\rangle = |y\rangle|\phi_0\rangle$.

We briefly discuss the behavior of the compressed oracle under a *classical* query, i.e., a query with the XY-register in state $|x\rangle|0\rangle$ for some x, and where the Y-register is then measured after the application of O_{XYD}. If D_x is in state ρ then a classical query on x will give response h with probability $\mathrm{tr}(|h\rangle\langle h|F\rho F)$ — unless ρ has nontrivial overlap with $|\phi_0\rangle$ and $h = 0$, in which a classical query on x will give response 0 with probability $\mathrm{tr}(|0\rangle\langle 0|F\rho F) + \mathrm{tr}(|\bot\rangle\langle\bot|F\rho F)$. The latter is an artifact of CNOT defined to act trivially on $|y\rangle|\bot\rangle$, which has the effect that $|\phi_0\rangle$ is treated like $F|0\rangle$. We note that, for any $h \in \mathcal{Y}$ and $\rho = |h\rangle\langle h|$,

$$\begin{aligned}\mathrm{tr}(|h\rangle\langle h|F\rho F) = |\langle h|F|h\rangle|^2 &= \Big|\langle h|\Big(|h\rangle + 2^{-n/2}(|\bot\rangle - |\phi_0\rangle)\Big)\Big|^2 \\ &= \Big|1 - 2^{-n/2}\langle h|\phi_0\rangle\Big|^2 = \Big|1 - 2^{-n}\Big|^2 \geq 1 - 2\cdot 2^{-n}\,.\end{aligned} \tag{8}$$

Vice-versa, after a classical query on x with response h, the state of D_x is $F|h\rangle$ — unless, the state of D_x prior to the query had a nontrivial overlap with $|\phi_0\rangle$ and $h = 0$, then the state after the query is supported by $F|0\rangle$ and $F|\bot\rangle = |\phi_0\rangle$.

Efficient representation of the compressed oracle. Following [27], one can make the (above variant of the) compressed oracle efficient. Indeed, by applying the standard classical sparse encoding to quantum states with the right choice of basis, one can *efficiently* maintain the state D, compute the unitary O_{XYD}, and extract information from D. More details are given in Appendix B of the full version. For simplicity, we mostly use the inefficient variant in this paper.

3 Main Technical Result: A Commutator Bound

3.1 Setup and the Technical Statement

Throughout this section, we consider an arbitrary but fixed relation $R \subset \mathcal{X} \times \{0,1\}^n$. A crucial parameter of the relation R is the number of y's that fulfill the relation together with x, maximized over all possible $x \in \mathcal{X}$:

$$\Gamma_R := \max_{x\in\mathcal{X}} \left|\left\{y \in \{0,1\}^n \middle| (x,y) \in R\right\}\right| . \tag{9}$$

Given the relation R, we consider the following projectors:

$$\Pi^x_{D_x} := \sum_{\substack{y \text{ s.t.}\\ (x,y)\in R}} |y\rangle\!\langle y|_{D_x} \quad \text{and} \quad \Pi^\emptyset_D := \mathbb{1}_D - \sum_{x\in\mathcal{X}} \Pi^x_{D_x} = \bigotimes_{x\in\mathcal{X}} \bar{\Pi}^x_{D_x} \tag{10}$$

with $\bar{\Pi}^x_{D_x} := \mathbb{1}_{D_x} - \Pi^x_{D_x}$. Informally, $\Pi^x_{D_x}$ checks whether register D_x contains a value $y \neq \perp$ such that $(x,y) \in R$. We then define the measurement $\mathcal{M} = \mathcal{M}^R$ to be given by the projectors

$$\Sigma^x := \bigotimes_{x'<x} \bar{\Pi}^{x'}_{D_{x'}} \otimes \Pi^x_{D_x} \quad \text{and} \quad \Sigma^\emptyset := \mathbb{1} - \sum_{x'} \Sigma^{x'} = \bigotimes_{x'} \bar{\Pi}^{x'}_{D_{x'}} = \Pi^\emptyset \tag{11}$$

where x ranges over all $x \in \mathcal{X}$. Informally, a measurement outcome x means that register D_x is the first that contains a value y such that $(x,y) \in R$; outcome $\emptyset$ means that no register contains such a value. For technical reasons, we consider the *purified* measurement $M_{DP} = M^R_{DP} \in \mathcal{L}(\mathcal{H}_D \otimes \mathcal{H}_R)$ given by the unitary[3]

$$M_{DP} := \sum_{x\in\mathcal{X}\cup\{\emptyset\}} \Sigma^x \otimes \mathsf{X}^x : |\varphi\rangle_D|w\rangle_P \mapsto \sum_{x\in\mathcal{X}\cup\{\emptyset\}} \Sigma^x|\varphi\rangle_D|w+x\rangle_P\,. \tag{12}$$

The following main technical result is a bound on the norm of $[O_{XYD}, M_{DP}]$.

Theorem 1. *For any relation $R \subset \mathcal{X} \times \{0,1\}^n$ and Γ_R as defined in Eq. (9), the purified measurement M_{DP} defined in Eq. (12) almost commutes with the oracle unitary O_{XYD}:*

$$\left\| [O_{XYD}, M_{DP}] \right\| \le 8 \cdot 2^{-n/2}\sqrt{2\Gamma_R}\,.$$

[3] Both in X^x and in $w + x$ we understand $x \in \mathcal{X} \cup \{\emptyset\}$ to be encoded as an element in $\mathbb{Z}/(|\mathcal{X}|+1)\mathbb{Z}$, $\dim(\mathcal{H}_P) = d := |\mathcal{X}| + 1$, and $\mathsf{X} \in \mathcal{L}(\mathcal{H}_P)$ is the generalized Pauli of order d that maps $|w\rangle$ to $|w+1\rangle$.

We note that Lemma 8 in [9] (with the subsequent discussion there) also provides a bound on a commutator involving O_{XYD}; however, there are various differences that make the two bounds incomparable. E.g., we consider a specific *measurement* whereas Lemma 8 in [9] is for a rather general *projector*. See further down for a comparison with Lemma 39 in [27].

Corollary 2. *For any state vector $|\psi\rangle \in \mathcal{H}_{WXYDP}$, with W an arbitrary additional register, $|\psi'\rangle := O_{XYD}M_{DP}|\psi\rangle$ and $|\psi''\rangle := M_{DP}O_{XYD}|\psi\rangle$ satisfy*

$$\delta\big(|\psi'\rangle\langle\psi'|, |\psi''\rangle\langle\psi''|\big) \le 8 \cdot 2^{-n/2}\sqrt{2\Gamma_R}\,.$$

The same holds for mixed states $\rho' := O_{XYD}M_{DP}\rho M_{DP}^\dagger O_{XYD}^\dagger$ and $\rho'' := M_{DP}O_{XYD}\rho O_{XYD}^\dagger M_{DP}^\dagger$.

Proof. By elementary properties and applying Theorem 1, we have that

$$\big\||\psi'\rangle - |\psi''\rangle\big\| \le \big\|[O_{XYD}, M_{DP}]\big\| \le 8 \cdot 2^{-n/2}\sqrt{2\Gamma_R}\,,$$

and the claim on the trace distance then follows from (5). The claim for mixed states follows from purification. □

3.2 The Proof

We prove the Theorem 1 by means of the following two lemmas.

Lemma 2. *Let F and $O^x_{YD_x}$ be the unitaries introduced in Sect. 2.2, and let $\Pi^x_{D_x}$ and $\Pi^\emptyset_D$ be as in (10). Set $\Gamma_x := \big|\{y \in \{0,1\}^n \big| (x,y) \in R\}\big|$. Then*

$$\big\|\big[F_{D_x}, \Pi^x_{D_x}\big]\big\| \le 2^{-n/2}\sqrt{2\Gamma_x}\,, \qquad \textit{as well as}$$

$$\big\|\big[O^x_{YD_x}, \Pi^x_{D_x}\big]\big\| \le 2 \cdot 2^{-n/2}\sqrt{2\Gamma_x} \quad \textit{and} \quad \big\|\big[O^x_{YD_x}, \Pi^\emptyset_D\big]\big\| \le 2 \cdot 2^{-n/2}\sqrt{2\Gamma_x}\,.$$

The bound on $\|[F, \Pi^x]\|$ can be considered a compact reformulation of Lemma 39 in [27]. We state it here in this form, and (re-)prove it in Appendix C of the full version, for convenience and completeness. The conceptually new and technically challenging ingredient to the proof of Theorem 1 is Lemma 3 below.[4]

Lemma 3. *The purified measurement M_{DP} defined in Equation (12) satisfies*

$$\big\|[F_{D_x}, M_{DP}]\big\| \le 3\big\|[F_{D_x}, \Pi^x_D]\big\| + \big\|[F_{D_x}, \Pi^\emptyset_D]\big\| \qquad \textit{and}$$

$$\big\|[O^x_{YD_x}, M_{DP}]\big\| \le 3\big\|[O^x_{YD_x}, \Pi^x_D]\big\| + \big\|[O^x_{YD_x}, \Pi^\emptyset_D]\big\|\,.$$

[4] The challenging aspect of Lemma 3 is that M_{DP} is made up of an exponential number of projectors Π^x, and thus the obvious approach of using triangle inequality leads to an exponential blow-up of the error term.

Proof. We do the proof for the second claim. The first is proven exactly the same way: the sole property we exploit from $O^x_{YD_x}$ is that it acts only on the D_x register within D, which holds for F_{D_x} as well. Let

$$\bar{\varDelta}^\xi := \bigotimes_{\xi' < \xi} \bar{\varPi}^{\xi'}_{D_{\xi'}}$$

be the projection that accepts if no register $D_{\xi'}$ with $\xi' < \xi$ contains a value y' with $(\xi', y') \in R$, and let $\varDelta^\xi$ be the complement. We then have, using that $\varPi^\xi$ and $\bar{\varDelta}^\xi$ act on disjoint registers,

$$\varSigma^\xi = \bar{\varDelta}^\xi \otimes \varPi^\xi = \varPi^\xi \bar{\varDelta}^\xi = \bar{\varDelta}^\xi \varPi^\xi \,. \tag{13}$$

We also observe that, with respect to the Loewner order, $\bar{\varDelta}^{\xi'} \geq \bar{\varDelta}^\xi$ for $\xi' < \xi$. Taking it as understood that $O^x_{YD_x}$ acts on registers Y and D_x, we can write

$$[O^x, M_{DP}] = \sum_\xi [O^x, \varSigma^\xi] \otimes \mathsf{X}^\xi + [O^x, \varSigma^\emptyset] \otimes \mathsf{X}^\emptyset \,. \tag{14}$$

Exploiting basic properties of the operator norm and recalling that $\varSigma^\emptyset = \varPi^\emptyset_D$, we see that the norm of the last term is bounded by $\|[O^x, \varSigma^\emptyset]\| = \|[O^x, \varPi^\emptyset]\|$.

To deal with the sum in (14), we use $\mathbb{1} = \varDelta^\xi + \bar{\varDelta}^\xi$ to further decompose

$$[O^x, \varSigma^\xi] = \bar{\varDelta}^\xi [O^x, \varSigma^\xi] \bar{\varDelta}^\xi + \bar{\varDelta}^\xi O^x, \varSigma^\xi] \varDelta^\xi + \varDelta^\xi [O^x, \varSigma^\xi] \bar{\varDelta}^\xi + \varDelta^\xi [O^x, \varSigma^\xi] \varDelta^\xi \,. \tag{15}$$

We now analyze the four different terms. For the first one, using (13) we see that

$$\bar{\varDelta}^\xi [O^x, \varSigma^\xi] \bar{\varDelta}^\xi = \bar{\varDelta}^\xi \big(O^x \varSigma^\xi - \varSigma^\xi O^x\big) \bar{\varDelta}^\xi = \bar{\varDelta}^\xi O^x \varPi^\xi \bar{\varDelta}^\xi - \bar{\varDelta}^\xi \varPi^\xi O^x \bar{\varDelta}^\xi = \bar{\varDelta}^\xi [O^x, \varPi^\xi] \bar{\varDelta}^\xi \,,$$

which vanishes for $\xi \neq x$, since then O^x and $\varPi^\xi$ act on different registers and thus commute. For $\xi = x$, its norm is upper bounded by $\|[O^x, \varPi^x]\|$.

We now consider the second term; the third one can be treated the same way by symmetry, and the fourth one vanishes, as will become clear immediately from below. Using (13) and $\bar{\varDelta}^\xi \varDelta^\xi = 0$, so that $\bar{\varDelta}^\xi \varSigma^\xi = 0$, we have

$$\bar{\varDelta}^\xi [O^x, \varSigma^\xi] \varDelta^\xi = \bar{\varDelta}^\xi \big(O^x \varSigma^\xi - \varSigma^\xi O^x\big) \varDelta^\xi = \varSigma^\xi O^x \varDelta^\xi =: N_\xi \,. \tag{16}$$

Looking at (14), we want to control the norm of the sum $N := \sum_\xi N_\xi \otimes X^\xi$. To this end, we show that N_ξ and $N_{\xi'}$ have orthogonal images and orthogonal support, i.e., $N^\dagger_{\xi'} N_\xi = 0 = N_{\xi'} N^\dagger_\xi$, for all $\xi \neq \xi'$. We first observe that if $x \geq \xi$ then O^x commutes with $\varDelta^\xi$, since they act on different registers then, and thus

$$N_\xi = \varSigma^\xi O^x \varDelta^\xi = \varSigma^\xi \varDelta^\xi O^x = \varPi^\xi \bar{\varDelta}^\xi \varDelta^\xi O^x = 0 \,,$$

exploiting once more that $\bar{\varDelta}^\xi \varDelta^\xi = 0$. Therefore, we only need to consider $N_\xi, N_{\xi'}$ for $\xi, \xi' > x$ (see Fig. 1 top left), where we may assume $\xi > \xi'$. For the orthogonality of the images, we observe that

$$\varPi^{\xi'} \bar{\varDelta}^\xi = 0 \tag{17}$$

by definition of $\bar{\Delta}^\xi$ as a tensor product with $\bar{\Pi}^{\xi'}$ being one of the components. Therefore,

$$(\Sigma^{\xi'})^\dagger \Sigma^\xi = \Sigma^{\xi'} \Sigma^\xi = \bar{\Delta}^{\xi'} \Pi^{\xi'} \bar{\Delta}^\xi \Pi^\xi = 0\,,$$

and $N_{\xi'}^\dagger N_\xi = 0$ follows directly (see also Fig. 1 top right). For the orthogonality of the supports, we recall that $\bar{\Delta}^{\xi'} \geq \bar{\Delta}^\xi$, and thus $\Delta^{\xi'} \leq \Delta^\xi$, from which it follows that $\Delta^\xi \Delta^{\xi'} = \Delta^{\xi'}$. $N_{\xi'} N_\xi^\dagger = 0$ then follows by exploiting (17) again (see Fig. 1 bottom).

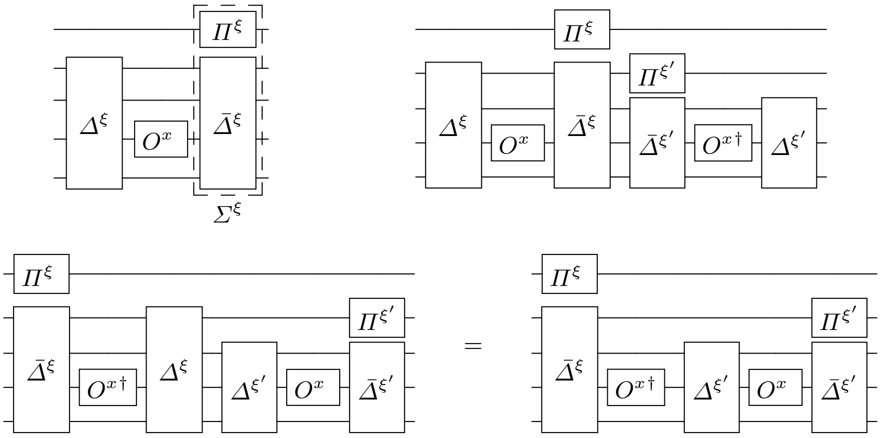

Fig. 1. Operators N_ξ (top left), $N_{\xi'}^\dagger N_\xi$ (top right), and $N_{\xi'} N_\xi^\dagger$ (bottom), for $x < \xi' < \xi$.

These orthogonality properties for the images and supports of the N_ξ immediately extend to $N_\xi \otimes X^\xi$, so we have

$$\|N\| \leq \max_{\xi > x} \|N_\xi \otimes \mathsf{X}^\xi\| \leq \max_{\xi > x} \|N_\xi\|$$

by Lemma 1. Recall from (16) that $N_\xi = \bar{\Delta}^\xi[\Sigma^\xi, O^x]\Delta^\xi$. Furthermore, we exploit that, by definition, Σ^ξ is in tensor-product form and O^x acts trivially on all components in this tensor product except for the component $\bar{\Pi}^x$, so that $[\Sigma^\xi, O^x] = [\bar{\Pi}^x, O^x]$ by property (4). Thus, $\|N_\xi\| \leq \|[\Sigma^\xi, O^x]\| = \|[\bar{\Pi}^x, O^x]\| = \|[\Pi^x, O^x]\|$. Using the triangle inequality with respect to the sum versus the last term in (14), and another triangle inequality with respect to the decomposition (15), we obtain the claimed inequality. □

The proof of Theorem 1 is now an easy consequence.

Proof (of Theorem 1). Since O_{XYD} is a control unitary $O_{XYD} = \sum_x |x\rangle\langle x| \otimes O^x_{YD_x}$, controlled by $|x\rangle$, while M_{DP} does not act on register X, it follows that

$$\big\|[O_{XYD}, M_{DP}]\big\| \leq \max_x \big\|[O^x_{YD_x}, M_{DP}]\big\|\,.$$

The claim now follows by combining Lemma 3 with Lemma 2. □

3.3 A First Immediate Application

As an immediate application of the commutator bound of Theorem 1, we can easily derive the following generic query-complexity bound for finding x with $(x, H(x)) \in R$ and Γ_R as defined in Eq. (9). Applied to $R = \mathcal{X} \times \{0^n\}$, where $\Gamma_R = 1$, we recover the famous lower bound for search in a random function.

Proposition 1. *For any algorithm $\mathcal{A}$ that makes q queries to the random oracle RO,*

$$\Pr_{x \leftarrow \mathcal{A}^{RO}} \big[\big(x, RO(x)\big) \in R\big] \leq 152(q+1)^2 \Gamma_R / 2^n \,. \tag{18}$$

Proof. Consider the modified algorithm $\mathcal{A}'$ that runs $\mathcal{A}$ to obtain output x, makes a query to obtain $RO(x)$ and outputs $(x, RO(x))$. By Lemma 5 in [27], we have that[5]

$$\sqrt{\Pr_{x \leftarrow \mathcal{A}'^H} [(x, RO(x)) \in R]} \leq \sqrt{\Pr_{x' \leftarrow G^R} [x' \neq \emptyset]} + 2^{-n/2}, \tag{19}$$

where G^R is the following procedure/game: (1) run $\mathcal{A}'$ using the compressed oracle, and (2) apply the measurement $\mathcal{M}^R$ to obtain $x' \in \mathcal{X} \cup \{\emptyset\}$, which is the same as preparing a register P, applying $M_{DP} = M^R_{DP}$, and measuring P.

In other words, writing $|\psi\rangle_{WXY}$ for the initial state of $\mathcal{A}'$ and V_{WXY} for the unitary applied between any two queries of $\mathcal{A}'$(which we may assume to be fixed), and setting $U_{WXYD} := V_{WXY} O_{XYD}$, $\Pi_P := \mathbb{1}_P - |\emptyset\rangle\langle\emptyset|_P$ and $|\Psi\rangle := |\psi\rangle_{WXY} \otimes |\bot\rangle_D^{\otimes|\mathcal{X}|} \otimes |0\rangle_P$, we have, omitting register subscripts,

$$\begin{aligned}
\sqrt{\Pr[x' \neq \emptyset]} &= \big\| \Pi M U^{q+1} |\Psi\rangle \big\| \\
&\leq \sum_{i=1}^{q+1} \big\| \Pi U^{i-1} [M, U] U^{q+1-i} |\Psi\rangle \big\| + \big\| \Pi U^{q+1} M |\Psi\rangle \big\| \\
&\leq (q+1) \big\| [M_{DP}, O_{XYD}] \big\| + \big\| \Pi_P M_{DP} |\Psi\rangle \big\| \\
&= (q+1) \big\| [M_{DP}, O_{XYD}] \big\| \leq 8 \cdot 2^{-n/2} (q+1) \sqrt{2\Gamma_R} \,,
\end{aligned}$$

where the last equation exploits that $\Pi_P M_{DP}$ applied to $|\bot\rangle_D^{\otimes|\mathcal{X}|} \otimes |0\rangle_P$ vanishes, and the final inequality is by Theorem 1. Observing $(8\sqrt{2}+1)^2 = 129 + 16\sqrt{2} \approx 151.6$ finishes the proof. □

4 Extraction of Random-Oracle Based Commitments

Throughout this Sect. 4, let $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{T}$ be an arbitrary fixed function with $\mathcal{Y} = \{0,1\}^n$. For a hash function $H : \mathcal{X} \to \mathcal{Y}$, which will then be modelled as a random oracle RO, we will think and sometimes speak of $f(x, H(x))$ as a *commitment* of x (though we do not require it to be a commitment scheme in the strict sense). Typical examples are $f(x, y) = y$ and $f(x, y) = \mathsf{Enc}_{pk}(x; y)$, where the latter is the encryption of x under public key pk with randomness y.

[5] Lemma 5 in [27] applies to an algorithm $\mathcal{A}$ that outputs both x and what is supposed to be its hash value; this is why we need to do this additional query.

4.1 Informal Problem Description

Consider a query algorithm $\mathcal{A}^{RO}$ in the random oracle model, which, during the course of its run, announces some $t \in \mathcal{T}$. This t is supposed to be $t = f(x, RO(x))$ for some x, and, indeed, $\mathcal{A}^{RO}$ may possibly reveal x later on. Intuitively, for the required relation between x and t to hold, we expect that $\mathcal{A}^{RO}$ *first* has to query RO on x and only *then* can output t; thus, one may hope to be able to extract x from RO *early on*, i.e., at the time $\mathcal{A}^{RO}$ announces t.

This is clearly true when $\mathcal{A}$ is restricted to classical queries, simply by checking all the queries made so far. This observation was first made and utilized by Pass [22] and only requires looking at the query transcript (it can be done in the *non-programmable* ROM). As the extractor does not change the course of the experiment, it works on-the-fly.

In the setting considered here, $\mathcal{A}^{RO}$ may query the random oracle in *superposition* over various choices of x, making it impossible to maintain a classical query transcript. On the positive side, since the output t is required to be classical, $\mathcal{A}^{RO}$ has to perform a measurement before announcing t, enforcing such a superposition to collapse.[6] We show here that early extraction of x is indeed possible in this quantum setting as well.

Note that if the goal is to extract *the same* x as $\mathcal{A}$ will (potentially) output, which is what we aim for, then we must naturally assume that it is hard for $\mathcal{A}$ to find $x \neq x'$ that are both consistent with the same t, i.e., we must assume the commitment to be binding. Formally, we will think of $\Gamma(f)$ and $\Gamma'(f)$, defined as follows, to be small compared to 2^n. When f is fixed, we simply write Γ and Γ'.

Definition 2. *For $f : \mathcal{X} \times \{0,1\}^n \to \mathcal{T}$, let $\Gamma(f) := \max_{x,t} |\{y \mid f(x,y) = t\}$ and $\Gamma'(f) := \max_{x \neq x', y'} |\{y \mid f(x,y) = f(x',y')\}|$.*

For the example $f(x,y) = y$, we have $\Gamma(f) = 1 = \Gamma'(f)$. For the example $f(x,y) = \mathsf{Enc}_{pk}(x;y)$, they both depend on the choice of the encryption scheme but typically are small, e.g. $\Gamma(f) = 1$ if Enc is injective as a function of the randomness y and $\Gamma'(f) = 0$ if there are no decryption errors.

4.2 The Extractable RO-Simulator $\mathcal{S}$

Towards formalizing the above goal, we introduce a simulator $\mathcal{S}$ that replaces RO and tries to extract x early on, right after $\mathcal{A}$ announces t. In more detail, $\mathcal{S}$ acts as a black-box oracle with two interfaces, the *RO-interface* $\mathcal{S}.RO$ providing access to the simulated random oracle, and the *extraction interface* $\mathcal{S}.E$ providing the functionality to extract x early on (see Fig. 3, left). In principle, both interfaces can be accessed quantumly, i.e., in superposition over different classical inputs, but in our applications we only use classical access to $\mathcal{S}.E$. We stress that $\mathcal{S}$ is per-se *stateful* and thus may change its behavior from query to query.

Formally, the considered simulator $\mathcal{S}$ is defined to work as follows. It simulates the random oracle and answers queries to $\mathcal{S}.RO$ by means of the compressed

[6] We can also think of this measurement being done by the interface that receives t.

oracle. For the $\mathcal{S}.E$ interface, upon a classical input $t \in \mathcal{T}$, $\mathcal{S}$ applies the measurement $\mathcal{M}^t := \mathcal{M}^{R_t}$ from (11) for the relation $R_t := \{(x,y) \,|\, f(x,y) = t\}$ to obtain $\hat{x} \in \mathcal{X} \cup \{\emptyset\}$, which it then outputs (see Fig. 2). In case of a *quantum* query to $\mathcal{S}.E$, the above is performed coherently: given the query registers TP, the unitary $\sum_t |t\rangle\langle t|_T \otimes M^{R_t}_{DP}$ is applied to TPD, and TP is then returned.

The extractable RO-oracle $\mathcal{S}$:

Initialization: $\mathcal{S}$ prepares its internal register D to be in state $|\bot\rangle_D := \bigotimes_x |\bot\rangle_{D_x}$.

$\mathcal{S}$.RO-query: Upon a (quantum) RO-query, with query registers XY, $\mathcal{S}$ applies O_{XYD} to registers XYD.

$\mathcal{S}$.E-query: Upon a classical extraction-query with input t, $\mathcal{S}$ applies $\mathcal{M}^t$ to D and returns the outcome $\hat{x}$.

Fig. 2. The (inefficient version of) simulator $\mathcal{S}$, restricted to classical extraction queries.

As described here, the simulator $\mathcal{S}$ is inefficient, having to maintain an exponential number of qubits; however, using the sparse representation of the internal state D, as discussed in Appendix B of the full version, $\mathcal{S}$ can well be made efficient without affecting its query-behavior (see Theorem 2 for details).

The following statement captures the core properties of $\mathcal{S}$. We refer to two subsequent queries as being *independent* if they can in principle be performed in either order, i.e., if the input to one query does not depend on the output of the other. More formally, e.g., two $\mathcal{S}.RO$ queries are independent if they can be captured by first preparing the two in-/output registers XY and $X'Y'$, and then doing the two respective queries with XY and $X'Y'$. The commutativity claim then means that the order does not matter. Furthermore, whenever we speak of a *classical* query (to $\mathcal{S}.RO$ or to $\mathcal{S}.E$), we consider the obvious classical variant of the considered query, with a classical input and a classical response. Finally, the almost commutativity claims are in terms of the trace distance of the (possibly quantum) output of any algorithm interacting with $\mathcal{S}$ arbitrarily and doing the two considered independent queries in one or the other order.

Theorem 2. *The extractable RO-simulator $\mathcal{S}$ constructed above, with interfaces $\mathcal{S}.RO$ and $\mathcal{S}.E$, satisfies the following properties.*

1. *If $\mathcal{S}.E$ is unused, $\mathcal{S}$ is perfectly indistinguishable from the random oracle RO.*

2.a *Any two subsequent independent queries to $\mathcal{S}.RO$ commute. Thus, two subsequent* classical *$\mathcal{S}.RO$-queries with the same input x give identical responses.*

2.b *Any two subsequent independent queries to $\mathcal{S}.E$ commute. Thus, two subsequent* classical *$\mathcal{S}.E$-queries with the same input t give identical responses.*

2.c *Any two subsequent independent queries to $\mathcal{S}.E$ and $\mathcal{S}.RO$ ε-almost-commute with $\varepsilon = 8\sqrt{2\Gamma(f)/2^n}$.*

3.a *Any classical query $\mathcal{S}.RO(x)$ is idempotent.*[7]

[7] I.e., applying it twice has the same effect on the state of $\mathcal{S}$ as applying it once.

3.b Any classical query $\mathcal{S}.E(t)$ is idempotent.

4.a If $\hat{x} = \mathcal{S}.E(t)$ and $\hat{h} = \mathcal{S}.RO(\hat{x})$ are two subsequent classical queries then

$$\Pr[f(\hat{x},\hat{h}) \neq t \wedge \hat{x} \neq \emptyset] \leq \Pr[f(\hat{x},\hat{h}) \neq t \mid \hat{x} \neq \emptyset] \leq 2 \cdot 2^{-n} \Gamma(f) \tag{20}$$

4.b If $h = \mathcal{S}.RO(x)$ and $\hat{x} = \mathcal{S}.E(f(x,h))$ are two subsequent classical queries such that no prior query to $\mathcal{S}.E$ has been made, then

$$\Pr[\hat{x} = \emptyset] \leq 2 \cdot 2^{-n}. \tag{21}$$

Furthermore, the total runtime of $\mathcal{S}$, when implemented using the sparse representation of the compressed oracle described in Sect. 2.2, is bounded as

$$T_{\mathcal{S}} = O\big(q_{RO} \cdot q_E \cdot \mathrm{Time}[f] + q_{RO}^2\big)\,,$$

where q_E and q_{RO} are the number of queries to $\mathcal{S}.E$ and $\mathcal{S}.RO$, respectively.

Proof. All the properties follow rather directly by construction of $\mathcal{S}$. Indeed, without $\mathcal{S}.E$-queries, $\mathcal{S}$ is simply the compressed oracle, known to be perfectly indistinguishable from the random oracle, confirming 1. Property 2.a follows because the unitaries O_{XYD} and $O_{X'Y'D}$, acting on the same register D but on distinct query registers, are both controlled unitaries with control register D, conjugated by a fixed unitary $(F^{\otimes|\mathcal{X}|})$. They thus commute. For 2.b, the claim follows because the unitaries M^t_{DP} and $M^{t'}_{DP'}$ commute, as they are both controlled unitaries with control register D. 2.c is a direct consequence of our main technical result Theorem 1 (in the form of Cor. 2). 3.a follows because a classical $\mathcal{S}.RO$ query with input x acts as a projective measurement on register D_x, which is, as any projective measurement, idempotent. Thus, so is the measurement $\mathcal{M}^t$, confirming 3.b.

To prove 4.a, consider the state $\rho_{D_{\hat{x}}}$ of register $D_{\hat{x}}$ after the measurement $\mathcal{M}^t$ that is performed by the extraction query $\hat{x} = \mathcal{S}.E(t)$, assuming $\hat{x} \neq \emptyset$. Let $|\psi\rangle$ be a purification of $\rho_{D_{\hat{x}}}$. By definition of $\mathcal{M}^t$, it holds that $\Pi^{\hat{x}}_{D_{\hat{x}}}|\psi\rangle = |\psi\rangle$. Then, understanding that all operators act on register $D_{\hat{x}}$, by definition of $\bar{\Pi}^{\hat{x}}$ the probability of interest is bounded as[8]

$$\begin{aligned}\Pr[f(\hat{x},\hat{h}) \neq t \,|\, \hat{x} \neq \emptyset] \leq \big\|\bar{\Pi}^{\hat{x}}F|\psi\rangle\big\|^2 = \big\|\bar{\Pi}^{\hat{x}}F\Pi^{\hat{x}}|\psi\rangle\big\|^2 &\leq \big\|\bar{\Pi}^{\hat{x}}F\Pi^{\hat{x}}\big\|^2 \\ &\leq \big\|[F,\Pi^{\hat{x}}]\big\|^2\,,\end{aligned}$$

where the last inequality exploits that $\bar{\Pi}^{\hat{x}}\Pi^{\hat{x}} = 0$. The claim now follows from Lemma 2.

For 4.b, we first observe that, given that there were no prior extraction queries, the state of D_x before the $h = \mathcal{S}.RO(x)$ query has no overlap with $|\phi_0\rangle$, and thus the state after the query is $F|h\rangle$ (see the discussion above Equation (8)). For the purpose of the argument, instead of applying the measurement

[8] The first inequality is an artefact of the $|\bot\rangle\langle\bot|$-term in $\bar{\Pi}^{\hat{x}}$ contributing to the probability of $\hat{h} = 0$, as discussed in Sect. 2.2.

$\mathcal{M}^{f(x,h)}$ to answer the $\mathcal{S}.E(f(x,h))$ query, we may equivalently consider a measurement in the basis $\{|\mathbf{y}\rangle\}$, and then set $\hat{x}$ to be the smallest element $\mathcal{X}$ so that $f(\hat{x}, y_{\hat{x}}) = t := f(x,h)$, with $\hat{x} = \emptyset$ if no such element exists. Then,

$$\begin{aligned}\Pr[\hat{x} \neq \emptyset] = \Pr[\exists\, \xi : f(\xi, y_\xi) = t] &\geq \Pr[f(x,y_x) = t] \\ &\geq \Pr[y_x = h] = |\langle h|F|h\rangle|^2 \geq 1 - 2 \cdot 2^{-n}\end{aligned}$$

where the last two (in)equalities are by Equation (8).

□

4.3 Two More Properties of $\mathcal{S}$

On top of the above basic features of our extractable RO-simulator $\mathcal{S}$, we show the following two additional, more technical, properties, which in essence capture that the extraction interface cannot be used to bypass query hardness results.

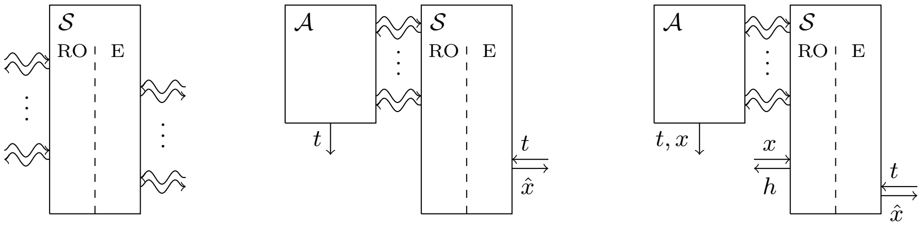

Fig. 3. The extractable RO-simulator $\mathcal{S}$, with its $\mathcal{S}.RO$ and $\mathcal{S}.E$ interfaces, distinguished here by queries from the left and right (left), and the games considered in Prop. 2 (middle) and 3 (right) for $\ell = 1$. Waved arrows denote quantum queries, straight arrows denote classical queries.

The first property is easiest to understand in the context of the example $f(x,y) = y$, where $\mathcal{S}.E(t)$ tries to extract a hash-preimage of t, and where the relations R and R' in Prop. 2 below then coincide. In this case, recall from Prop. 1 that, informally, if Γ_R is small then it is hard to find $x \in \mathcal{X}$ so that $t := RO(x)$ satisfies $(x,t) \in R$. The statement below ensures that this hardness cannot be bypassed by first selecting a "good" hash value t and then trying to extract a preimage by means of $\mathcal{S}.E$ (Fig. 3, middle).

Proposition 2. *Let $R' \subseteq \mathcal{X} \times \mathcal{T}$ be a relation. Consider a query algorithm $\mathcal{A}$ that makes q queries to the $\mathcal{S}.RO$ interface of $\mathcal{S}$ but no query to $\mathcal{S}.E$, outputting some $\mathbf{t} \in \mathcal{T}^\ell$. For each i, let $\hat{x}_i$ then be obtained by making an additional query to $\mathcal{S}.E$ on input t_i (see Fig. 3, middle). Then*

$$\Pr_{\substack{\mathbf{t} \leftarrow \mathcal{A}^{\mathcal{S}.RO} \\ \hat{x}_i \leftarrow \mathcal{S}.E(t_i)}}[\exists\, i : (\hat{x}_i, t_i) \in R'] \leq 128 \cdot q^2 \Gamma_R / 2^n\,,$$

where $R \subseteq \mathcal{X} \times \mathcal{Y}$ is the relation $(x,y) \in R \Leftrightarrow (x, f(x,y)) \in R'$ and Γ_R as in (9).

Proof. The considered experiment is like the experiment G^R in the proof of Prop. 1, the only difference being that in G^R the measurement $\mathcal{M}^R$ is applied to register D to obtain x' (see Fig. 4, middle), while here we have ℓ measurements $\mathcal{M}^{t_i}$ that are applied to obtain $\hat{x}_i$ (see Fig. 4, left). Since all measurements are defined by means of projections that are diagonal in the same basis $\{|\mathbf{y}\rangle\}$ with $|\mathbf{y}\rangle$ ranging over $\mathbf{y} \in (\mathcal{Y} \cup \{\perp\})^{\mathcal{X}}$, we may equivalently measure D in that basis to obtain $\mathbf{y}$ (see Fig. 4, right), and let $\hat{x}_i$ be minimal so that $f(\hat{x}_i, y_{\hat{x}_i}) = t_i$ (and $\hat{x}_i = \emptyset$ if no such value exists), and let x' be minimal so that $(x', y_{x'}) \in R$ (and $x' = \emptyset$ if no such value exists). By the respective definitions of $\mathcal{M}^t_i$ and $\mathcal{M}^R$, both pairs of random variables $(\hat{\mathbf{x}}, \mathbf{t})$ and $(x', \mathbf{t})$ then have the same distributions as in the respective original two games. But now, we can consider their joint distribution and argue that

$$\begin{aligned}\Pr[\exists\, i : (\hat{x}_i, t_i) \in R'] &= \Pr[\exists\, i : (\hat{x}_i, f(\hat{x}_i, y_{\hat{x}_i})) \in R'] \\ &= \Pr[\exists\, i : (\hat{x}_i, y_{\hat{x}_i}) \in R] \leq \Pr[\exists\, x : (x, y_x) \in R] = \Pr[x' \neq \emptyset]\,.\end{aligned}$$

The bound on $\Pr[x' \neq \emptyset]$ from the proof of Prop. 1 concludes the proof. □

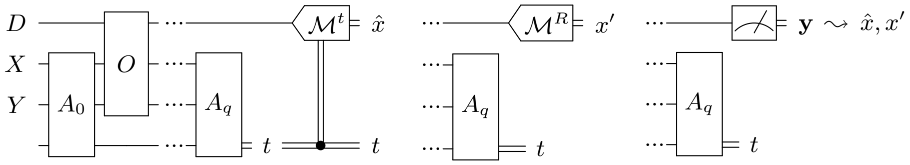

Fig. 4. Quantum circuit diagrams of the experiments in the proof of Prop. 2 for $\ell = 1$.

In a somewhat similar spirit, the following ensures that if it is hard in the QROM to find x and x' with $f(x, RO(x)) = f(x', RO(x'))$ then this hardness cannot be bypassed by, say, first choosing x, querying $h = \mathcal{S}.RO(x)$, computing $t := f(x, h)$, and then extracting $\hat{x} := \mathcal{S}.E(t)$. The latter will most likely give $\hat{x} = x$, except, intuitively, if $\mathcal{S}.RO$ has additionally been queried on a colliding x'.

Proposition 3. *Consider a query algorithm $\mathcal{A}$ that makes q queries to $\mathcal{S}.RO$ but no query to $\mathcal{S}.E$, outputting some $t \in \mathcal{T}$ and $x \in \mathcal{X}$. Let h then be obtained by making an additional query to $\mathcal{S}.RO$ on input x, and $\hat{x}$ by making an additional query to $\mathcal{S}.E$ on input t (see Fig. 3, right). Then*

$$\Pr_{\substack{t,\, x \,\leftarrow\, \mathcal{A}^{\mathcal{S}.RO} \\ h \,\leftarrow\, \mathcal{S}.RO(x) \\ \hat{x} \,\leftarrow\, \mathcal{S}.E(t)}} [\hat{x} \neq x \wedge f(x,h) = t] \leq \frac{40e^2(q+2)^3\Gamma'(f) + 2}{2^n}\,.$$

More generally, if $\mathcal{A}$ outputs ℓ-tuples $\mathbf{t} \in \mathcal{T}^\ell$ and $\mathbf{x} \in \mathcal{X}^\ell$, and $\mathbf{h} \in \mathcal{Y}^\ell$ is obtained by querying $\mathcal{S}.RO$ component-wise on $\mathbf{x}$, and $\hat{\mathbf{x}} \in (\mathcal{X} \cup \{\emptyset\})^\ell$ by querying $\mathcal{S}.E$

component-wise on $\mathbf{t}$*, then*

$$\Pr_{\substack{\mathbf{t},\mathbf{x} \leftarrow \mathcal{A}^{\mathcal{S}.RO} \\ \mathbf{h} \leftarrow \mathcal{S}.RO(\mathbf{x}) \\ \hat{\mathbf{x}} \leftarrow \mathcal{S}.E(\mathbf{t})}} [\exists\, i : \hat{x}_i \neq x_i \wedge f(x_i, h_i) = t] \leq \frac{40e^2(q+\ell+1)^3 \Gamma'(f) + 2}{2^n} \,.$$

The proof is similar in spirit to the proof of Prop. 2, but relying on the hardness of collision finding rather than on (the proof of) Prop. 1, and so is moved to Appendix C in the full version.

Remark 1. The claim of Prop. 3 stays true when the queries $\mathcal{S}.RO(x_i)$ are not performed as *additional* queries *after* the run of $\mathcal{A}$ but are explicitly *among* the q queries that are performed by $\mathcal{A}$ *during* its run. Indeed we observe that the proof does not exploit that these queries are performed at the end, which additionally shows that in this case the ℓ-term on the right hand side of the bound vanishes, i.e., scales as $(q+1)^3$ rather than as $(q+\ell+1)^3$.

4.4 Early Extraction

We consider here the following concrete setting. Let $\mathcal{A}$ be a two-round query algorithm, interacting with the random oracle RO and behaving as follows. At the end of the first round, $\mathcal{A}^{RO}$ outputs some $t \in \mathcal{T}$, and at the end of the second round, it outputs some $x \in \mathcal{X}$ that is supposed to satisfy $f(x, RO(x)) = t$; on top, $\mathcal{A}^{RO}$ may have some additional (possibly quantum) output W.

We now show how the extractable RO-simulator $\mathcal{S}$ provides the means to extract x early on, i.e., right after $\mathcal{A}$ has announced t. To formalize this claim, we consider the following experiment, which we denote by $G^{\mathcal{A}}_{\mathcal{S}}$. The RO-interface $\mathcal{S}.RO$ of $\mathcal{S}$ is used to answer all the oracle queries made by $\mathcal{A}$. In addition, as soon as $\mathcal{A}$ outputs t, the interface $\mathcal{S}.E$ is queried on t to obtain $\hat{x} \in \mathcal{X} \cup \{\emptyset\}$, and after $\mathcal{A}$ has finished, $\mathcal{S}.RO$ is queried on $\mathcal{A}$'s final output x to generate h.

Informally, we want that $\mathcal{A}$ does not notice any difference when RO is replaced by $\mathcal{S}.RO$, and that $\hat{x} = x$ whenever $f\big(x, h\big) = t$, while $\hat{x} = \emptyset$ implies that $\mathcal{A}$ will fail to output x with $f\big(x, h\big) = t$. This situation is captured by the following statement.

Corollary 3. *The extractable RO-simulator* $\mathcal{S}$ *is such that the following holds. For any* $\mathcal{A}$ *that outputs* t *after* q_1 *queries and* $x \in \mathcal{X}$ *and* W *after an additional* q_2 *queries, setting* $q = q_1 + q_2$*, it holds that*

$$\delta\big([t, x, RO(x), W]_{\mathcal{A}^{RO}}, [t, x, h, W]_{G^{\mathcal{A}}_{\mathcal{S}}}\big) \leq 8(q_2+1)\sqrt{2\Gamma/2^n} \qquad \textit{and}$$

$$\Pr_{G^{\mathcal{A}}_{\mathcal{S}}}\big[x \neq \hat{x} \wedge f(x, h) = t\big] \leq 8(q_2+1)\sqrt{2\Gamma/2^n} + \frac{40e^2(q+2)^3\Gamma'(f) + 2}{2^n} \,,$$

Proof. The first claim follows because the trace distance vanishes when $\mathcal{S}.E(t)$ is performed at the very end, after the $\mathcal{S}.RO(x)$-query, in combination with the (almost-)commutativity of the two interfaces (Theorem 2, 2.a to 2.c). Similarly, the second claim follows from Prop. 3 when considering the $\mathcal{S}.E(t)$ query to be performed at the very end, in combination with the (almost-)commutativity of the interfaces again. □

The statements above extend easily to *multi*-round algorithms $\mathcal{A}^{RO}$ that output $t_1, \ldots, t_\ell$ in (possibly) different rounds, and $x_1, \ldots, x_\ell \in \mathcal{X}$ and some (possibly quantum) output W at the end of the run. We then extend the definition of $G_{\mathcal{S}}^{\mathcal{A}}$ in the obvious way: $\mathcal{S}.E$ is queried on each output t_i to produce $\hat{x}_i$, and at the end of the run $\mathcal{S}.RO$ is queried on each of the final outputs $x_1, \ldots, x_\ell$ of $\mathcal{A}$ to obtain $\mathbf{h} = (h_1, \ldots, h_\ell) \in \mathcal{Y}^\ell$. As a minor extension, we allow some of the x_i to be $\bot$, i.e., $\mathcal{A}^{RO}$ may decide to not output certain x_i's; the $\mathcal{S}.RO$ query on x_i is then not done and h_i is set to $\bot$ instead, and we declare that $RO(\bot) = \bot$ and $f(\bot, h_i) \neq t_i$. To allow for a compact notation, we write $RO(\mathbf{x}) = (RO(x_1), \ldots, RO(x_\ell))$ for $\mathbf{x} = (x_1, \ldots, x_\ell)$.

Corollary 4. *The extractable RO-simulator $\mathcal{S}$ is such that the following holds. For any $\mathcal{A}$ that makes q queries in total, it holds that*

$$\delta\big([\mathbf{t}, \mathbf{x}, RO(\mathbf{x}), W]_{\mathcal{A}^{RO}}, [\mathbf{t}, \mathbf{x}, \mathbf{h}, W]_{G_{\mathcal{S}}^{\mathcal{A}}}\big) \leq 8\ell(q+\ell)\sqrt{2\Gamma/2^n} \qquad \text{and}$$

$$\Pr_{G_{\mathcal{S}}^{\mathcal{A}}}\big[\exists\, i : x_i \neq \hat{x}_i \wedge f(x_i, h_i) = t_i\big] \leq 8\ell(q+1)\sqrt{2\Gamma/2^n} + \frac{40e^2(q+\ell+1)^3\Gamma'(f)+2}{2^n}.$$

5 Extractability of Commit-And-Open Σ-protocols

5.1 Commit-and-Open Σ-protocols

We assume the reader to be familiar with the concept of an interactive proof for a language $\mathcal{L}$ or a relation R, and specifically with the notion of a *Σ-protocol*.

Here, we consider the notion of a *commit-and-open* Σ-protocol, which is as follows. The prover begins by sending commitments $a_1, ..., a_\ell$ to the prover, computed as $a_i = H(x_i)$ for $x_1, ..., x_\ell \in \mathcal{X}$, where $H : \mathcal{X} \to \{0,1\}^n$ is a hash function. Here, x_i can either be the actual message m_i to be committed, or m_i concatenated with randomness. The verifier answers by sending a challenge c, which is a subset $c \subseteq [\ell] = \{1, ..., \ell\}$, picked uniformly at random from a challenge set $C \subseteq 2^{[\ell]}$, upon which the prover sends the response $z = (x_i)_{i \in c}$. Finally, the verifier checks whether $H(x_i) = a_i$ for every $i \in c$, computes an additional verification predicate $V(c, z)$ and outputs 1 if both check out, 0 otherwise. Such (usually zero-knowledge) protocols have been known since the concept of zero-knowledge proofs was developed [5, 14].

Commit-and-open Σ-protocols are (classically) extractable in a straight-forward manner as soon as a witness can be computed from sufficiently many of the x_i's: rewind the prover a few times until it has opened every commitment a_i at least once.[9] There is, however, an alternative (classical) *online* extractor if the hash function H is modelled as a random oracle: simply look at the query transcript

[9] Naturally, we can assume $[\ell] = \bigcup_{c \in C} c$

of the prover to find preimages of the commitments $a_1, ..., a_\ell$. As the challenge is chosen independently, the extractability and collision resistance of the commitments implies that for a prover with a high success probability, the ℓ extractions succeed simultaneously with good probability. This is roughly how the proof of online extractability of the ZK proof system for graph 3-coloring by Goldreich, Micali and Wigderson [14], instantiated with random-oracle based commitments, works that was announced in [22] and shown in [23] (Prop. 5).

Equipped with our extractable RO-simulator $\mathcal{S}$, we can mimic the above in the quantum setting. Indeed, the only change is that the look-ups in the transcript are replaced with the additional interface of the simulator $\mathcal{S}$. Cor. 4 can then be used to prove the success of extraction using essentially the same extractor as in the classical case.

5.2 Notions of Special Soundness

The property that allows such an extraction is most conveniently expressed in terms of special soundness and its variants. Because there are, next to special and k-soundness, a number of additional variants in the literature (e.g. in the context of Picnic2/Picnic3 [17] or MQDSS [8]), we begin by formulating a generalized notion of special soundness that captures in a broad sense that a witness can be computed from correct responses to "*sufficiently many*" challenges.[10] While the notions introduced below can be formulated for arbitrary public-coin interactive proof systems, we present them here tailored to commit-and-open Σ-protocols. In [26], Wikström considers a similar notion of general special soundness (but then for arbitrary multi-round public-coin interactive proof systems); however, the formalism in [26] is more restrictive in that it requires the set system we call $\mathfrak{S}_{\min}$ below to form the set of bases of a matroid. As a consequence, the r-fold parallel repetition of a k-sound protocol is for instance not captured by the formalism suggested by Wikström.

In the remainder, Π is thus assumed to be an arbitrary commit-and-open Σ-protocol for a relation R with associated language $\mathcal{L}$, and C is the challenge space of Π. Furthermore, we consider a non-empty, monotone increasing set $\mathfrak{S}$ of subsets $S \subseteq C$, i.e., such that $S \in \mathfrak{S} \wedge S \subseteq S' \;\Rightarrow\; S' \in \mathfrak{S}$, and we let $\mathfrak{S}_{\min} := \{S \in \mathfrak{S} \,|\, S_\circ \subsetneq S \Rightarrow S_\circ \notin \mathfrak{S}\}$ consist of the minimal sets in $\mathfrak{S}$.

Definition 3. *Π is called* $\mathfrak{S}$-sound *if there exists an efficient algorithm $\mathcal{E}_{\mathfrak{S}}(I, x_1, \ldots, x_\ell, S)$ that takes as input an instance $I \in \mathcal{L}$, strings $x_1, \ldots, x_\ell \in \mathcal{X}$ and a set $S \in \mathfrak{S}_{\min}$, and outputs a witness for I whenever $V(c, (x_i)_{i \in c}) = 1$ for all $c \in S$, and outputs $\perp$ otherwise.*[11]

Note that there is no correctness requirement on the x_i's with $i \notin \bigcup_{c \in S} c$; thus, those x_i's may just as well be set to be empty strings.

[10] Using the language from secret sharing, we consider an arbitrary access structure $\mathfrak{S}$, while the k-soundness case corresponds to a threshold access structure.

[11] The restriction for S to be in $\mathfrak{S}_{\min}$, rather than in $\mathfrak{S}$, is only to avoid an exponentially sized input. When C is constant in size, we may admit any $S \in \mathfrak{S}$.

This property generalizes k-soundness, which is recovered for $\mathfrak{S} = \mathfrak{T}_k := \{S \subseteq C \,|\, |S| \geq k\}$, but it also captures more general notions. For instance, the r-fold parallel repetition of a k-sound protocol is not k-sound anymore, but it is $\mathfrak{T}_k^{\vee r}$-sound with $\mathfrak{T}_k^{\vee r}$ consisting of those subsets of challenge-sequences $(c_1, \ldots, c_r) \in C^r$ for which the restriction to at least one of the positions is a set in $\mathfrak{T}_k$. This obviously generalizes to the parallel repetition of an arbitrary $\mathfrak{S}$-sound protocol, with the parallel repetition then being $\mathfrak{S}^{\vee r}$-sound with

$$\mathfrak{S}^{\vee r} := \{S \subseteq C^r \,|\, \exists\, i : S_i \in \mathfrak{S}\}\,,$$

where $S_i := \{c \in C \,|\, \exists\,(c_1, ..., c_r) \in S : c_i = c\}$ is the i-th *marginal* of S.

For our result to apply, we need a strengthening of the above soundness condition where $\mathcal{E}_{\mathfrak{S}}$ has to find the set S himself. This is clearly the case for $\mathfrak{S}$-sound protocols that have a *constant sized* challenge space C, but also for the parallel repetition of $\mathfrak{S}$-sound protocols with a constant sized challenge space. Formally, we require the following strengthened notion of $\mathfrak{S}$-sound protocols.

Definition 4. *Π is called* $\mathfrak{S}$-sound* *if there exists an efficient algorithm $\mathcal{E}^*_{\mathfrak{S}}(I, x_1, \ldots, x_\ell)$ that takes as input an instance $I \in \mathcal{L}$ and strings $x_1, \ldots, x_\ell \in \mathcal{X}$, and outputs a witness for I whenever there exists $S \in \mathfrak{S}$ with $V(c, (x_i)_{i \in c}) = 1$ for all $c \in S$, and outputs $\perp$ otherwise.*

$\mathfrak{S}$-sound Σ-protocols may — and often do — have the property that a dishonest prover can pick any set $\hat{S} = \{\hat{c}_1, \ldots, \hat{c}_m\} \notin \mathfrak{S}$ of challenges $\hat{c}_i \in C$ and then prepare $\hat{x}_1, \ldots, \hat{x}_\ell$ in such a way that $V(c, (\hat{x}_i)_{i \in c}) = 1$ if $c \in \hat{S}$, i.e., after having committed to $\hat{x}_1, \ldots, \hat{x}_\ell$ the prover can successfully answer challenge c if $c \in \hat{S}$. We call this a *trivial* attack. The following captures the largest success probability of such a trivial attack, maximized over the choice of $\hat{S}$:

$$p^{\mathfrak{S}}_{triv} := \frac{1}{|C|} \max_{\hat{S} \notin \mathfrak{S}} |\hat{S}|\,. \tag{22}$$

When there is no danger of confusion, we omit the superscript $\mathfrak{S}$. Looking ahead, our result will show that for any prover that does better than the trivial attack by a non-negligible amount, online extraction is possible. For special sound commit-and-open Σ-protocols, $p_{triv} = 1/|C|$, and for k-sound protocols, $p_{triv} = (k-1)/|C|$. Furthermore, our definition of $\mathfrak{S}$-soundness allows a straightforward parallel repetition lemma on the combinatorial level providing an expression for p_{triv} of parallel-repeated commit-and-open Σ-protocols (the proof is an easy computation).

Lemma 4. *Let Π be $\mathfrak{S}$-sound. Then* $p^{\mathfrak{S}^{\vee r}}_{triv} = \left(p^{\mathfrak{S}}_{triv}\right)^r$.

5.3 Online Extractability in the QROM

We are now ready to define our extractor and prove that it succeeds. Equipped with the results from the previous section, the intuition is very simple. Given a (possibly dishonest) prover $\mathcal{P}$, running the considered Σ-protocol in the QROM, we use the simulator $\mathcal{S}$ to answer $\mathcal{P}$'s queries to the random oracle but also to extract the commitments $a_1, \ldots, a_\ell$, and if the extracted $\hat{x}_1, \ldots, \hat{x}_\ell$ satisfy the verification predicate V for sufficiently many challenges, we can compute a witness by applying $\mathcal{E}^*_{\mathfrak{S}}$.

The following relates the success probability of this extraction procedure to the success probability of the (possibly dishonest) prover.

Theorem 3. *Let Π be an $\mathfrak{S}$-sound* commit-and-open Σ-protocol where the first message consists of ℓ commitments. Then it admits an online extractor $\mathcal{E}$ in the QROM that succeeds with probability*

$$\Pr[\mathcal{E} \text{ succeeds}] \geq \frac{1}{1-p_{triv}}\big(\Pr[\mathcal{P}^{RO} \text{ succeeds}] - p_{triv} - \varepsilon\big) \qquad \textit{where} \tag{23}$$

$$\varepsilon = 8\sqrt{2}\,\ell(2q+\ell+1)/\sqrt{2^n} + \frac{40e^2(q+\ell+1)^3\Gamma'(f)+2}{2^n}$$

and p_{triv} is defined in Eq. (22). For $q \geq \ell+1$, the bound simplifies to

$$\varepsilon \leq 34\ell q/\sqrt{2^n} + 2365q^3/2^n\,.$$

Furthermore, the running time of $\mathcal{E}$ is bounded as $T_{\mathcal{E}} = T_{\mathcal{P}_1} + T_{\mathcal{E}^*_{\mathfrak{S}}} + O(q_1^2)$, where $T_{\mathcal{P}_1}$ and $T_{\mathcal{E}^*_{\mathfrak{S}}}$ are the respective runtimes of $\mathcal{P}_1$ and $\mathcal{E}^*_{\mathfrak{S}}$.

Recall that $p_{triv} = (k-1)/|C|$ for k-soundness, giving a corresponding bound. We note that the bound (23) is tight, and the additive term ε has a matching attack for some schemes, see Section 5.4 of the full version.

Proof. We begin by describing the extractor $\mathcal{E}$. First, using $\mathcal{S}.RO$ to answer $\mathcal{P}$'s queries, $\mathcal{E}$ runs the prover $\mathcal{P}$ until it announces $a_1, \ldots, a_\ell$, and then it uses $\mathcal{S}.E$ to extract $\hat{x}_1, ..., \hat{x}_\ell$. I.e., $\mathcal{E}$ acts as $\mathcal{S}$ in Cor. 4 for the function $f(x,h) = h$ and runs the game $G^{\mathcal{P}}_{\mathcal{S}}$ to the point where $\mathcal{S}.E$ outputs $\hat{x}_1, ..., \hat{x}_\ell$ on input $a_1, \ldots, a_\ell$. As a matter of fact, for the purpose of the analysis, we assume that $G^{\mathcal{P}}_{\mathcal{S}}$ is run until the end, with the challenge c chosen uniformly at random, and where $\mathcal{P}$ then outputs x_i for all $i \in c$ (and $\perp$ for $i \notin c$) at the end of $G^{\mathcal{P}}_{\mathcal{S}}$; we also declare that $\mathcal{P}$ additionally outputs c and $a_1, \ldots, a_\ell$ at the end. Then, upon having obtained $\hat{x}_1, ..., \hat{x}_\ell$, the extractor $\mathcal{E}$ runs $\mathcal{E}^*_{\mathfrak{S}}$ on $\hat{x}_1, ..., \hat{x}_\ell$ to try to compute a witness. By definition, this succeeds if $\hat{S} := \{\hat{c} \in C \,|\, V(\hat{c}, (\hat{x}_i)_{i\in\hat{c}}) = 1\}$ is in $\mathfrak{S}$.

It remains to relate the success probability of $\mathcal{E}$ to that of the prover $\mathcal{P}^{RO}$. By the first statement of Cor. 4, writing $\mathbf{x}_c = (x_i)_{i\in c}$, $RO(\mathbf{x}_c) = (RO(x_i))_{i\in c}$, $\mathbf{a}_c = (a_i)_{i\in c}$, etc., we have, writing $V(c, \mathbf{x}_c)$ instead of $V(c, \mathbf{x}_c) = 1$ for brevity,

$$\begin{aligned}\Pr[\mathcal{P}^{RO} \text{ succeeds}] &= \Pr_{\mathcal{P}^{RO}}[V(c,\mathbf{x}_c) \wedge RO(\mathbf{x}_c) = \mathbf{a}_c] \\ &\leq \Pr_{G^{\mathcal{P}}_{\mathcal{S}}}[V(c,\mathbf{x}_c) \wedge \mathbf{h}_c = \mathbf{a}_c] + \delta_1\end{aligned} \tag{24}$$

with $\delta_1 = 8\sqrt{2}\,\ell(q+\ell)/\sqrt{2^n}$. Omitting the subscript $G^{\mathcal{P}}_{\mathcal{S}}$ now,

$$\begin{aligned}&\Pr[V(c,\mathbf{x}_c) \wedge \mathbf{h}_c = \mathbf{a}_c] \\ &\quad\leq \Pr[V(c,\mathbf{x}_c) \wedge \mathbf{h}_c = \mathbf{a}_c \wedge \mathbf{x}_c = \hat{\mathbf{x}}_c] + \Pr[\mathbf{h}_c = \mathbf{a}_c \wedge \mathbf{x}_c \neq \hat{\mathbf{x}}_c] \\ &\quad\leq \Pr[V(c,\hat{\mathbf{x}}_c)] + \Pr[\exists\, j \in c : x_j \neq \hat{x}_j \wedge h_j = a_j] \leq \Pr[V(c,\hat{\mathbf{x}}_c)] + \delta_2\end{aligned} \tag{25}$$

with $\delta_2 = 8\sqrt{2}\,\ell(q+1)/\sqrt{2^n} + \frac{40e^2(q+\ell+1)^3\Gamma'(f)+2}{2^n}$, where the last inequality is by the second statement of Cor. 4, noting that, by choice of f, the event $h_j = a_j$ is equal to $f(x_j, h_j) = a_j$. Recalling the definition of $\hat{S}$,

$$\begin{aligned}\Pr[V(c, \hat{\mathbf{x}}_c) = 1] = \Pr[c \in \hat{S}] &\le \Pr[\hat{S} \in \mathfrak{S}] + \Pr[c \in \hat{S} \,|\, \hat{S} \notin \mathfrak{S}] \Pr[\hat{S} \notin \mathfrak{S}] \\ &\le \Pr[\mathcal{E} \text{ succeeds}] + p_{triv}(1 - \Pr[\mathcal{E} \text{ succeeds}]).\end{aligned} \tag{26}$$

The last inequality holds as c is chosen at random independent of the $\hat{x}_i$, and hence independent of the event $\hat{S} \notin \mathfrak{S}$. Combining (24), (25) and (26), we obtain

$$\Pr[\mathcal{P}^{RO} \text{ succeeds}] \le \Pr[\mathcal{E} \text{ succeeds}] + p_{triv}(1 - \Pr[\mathcal{E} \text{ succeeds}]) + \delta_1 + \delta_2$$

and solving for $\Pr[\mathcal{E}$ succeeds$]$ gives the claimed bound. □

Application to Fiat Shamir Signatures. In Appendix E of the full version, we discuss the impact on Fiat Shamir signatures, in particular on the round-3 signature candidate Picnic [7] in the NIST standardization process for post-quantum cryptographic schemes. In short, a crucial part in the chain of arguments to prove security of Fiat Shamir signatures is to prove that the underlying Σ-protocol is a proof of knowledge. For post-quantum security, so far this step relied on Unruh's rewinding lemma, which leads (after suitable generalization), to a $(2k+1)$-th root loss for a k-sound protocols. For commit-and-open Σ-protocols, Theorem 3 can replace Unruhs rewinding lemma when working in the QROM, making this step in the chain of arguments tight up to unavoidable additive errors.

As an example, Theorem 3 implies a sizeable improvement over the current best QROM security proof of Picnic2 [7, 17, 6]. Indeed, Unruh's rewinding lemma implies a 6-th root loss for the variant of special soundness the underlying Σ-protocol possesses [12], while Theorem 3 is tight.

6 QROM-Security of Textbook Fujisaki-Okamoto

6.1 The Fujisaki-Okamoto Transformation

The Fujisaki-Okamoto (FO) transform [13] is a general method to turn any public-key encryption scheme secure against *chosen-plaintext attacks* (CPA) into a key-encapsulation mechanism (KEM) that is secure against *chosen-ciphertext attacks* (CCA). We can start either from a scheme with one-way security (OW-CPA) or from one with indistinguishability (IND-CPA), and in both cases obtain an IND-CCA secure KEM. We recall that a KEM establishes a shared key, which can then be used for symmetric encryption.

We include the (standard) formal definitions of a public-key encryption scheme and of a KEM in Appendix F of the full version, and we recall the notions of δ-correctness and γ-spreadness there. In addition, we define a relaxed version of the latter property, *weak* γ-spreadness (see Def. F.4), where the ciphertexts are only required to have high min-entropy when averaged over key generation[12]. The security games for OW-CPA security of a public-key encryption scheme and for IND-CCA security of a KEM are given in Section 6.1 of the full version. The formal specification of the FO transformation, mapping a public-key encryption scheme $\mathsf{PKE} = (\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ and two suitable hash functions H and G (which will then be modeled as random oracles) into a key encapsulation mechanism $\mathsf{FO}[\mathsf{PKE}, H, G] = (\mathsf{Gen}, \mathsf{Encaps}, \mathsf{Decaps})$, is given in Fig. 5.

```
Gen
    $(sk, pk) \leftarrow \mathsf{Gen}$
    return $(sk, pk)$

Encaps$(pk)$
    $m \xleftarrow{\$} \mathcal{M}$
    $c \leftarrow \mathsf{Enc}_{pk}(m; H(m))$
    $K := G(m)$
    return $(K, c)$

Decaps$_{sk}(c)$
    $m := \mathsf{Dec}_{sk}(c)$
    if $m = \perp$ or $\mathsf{Enc}_{pk}(m; H(m)) \neq c$
        return $\perp$
    else return $K := G(m)$
```

Fig. 5. The KEM $\mathsf{FO}[\mathsf{PKE}, H, G]$, obtained by applying the FO transformation to PKE.

6.2 Post-Quantum Security of FO in the QROM

Our main contribution here is a new security proof for the FO transformation in the QROM. In contrast to most previous works on the topic, our result applies to the *standard* FO transformation, without any adjustments. Next to being CPA secure, we require the underlying public-key encryption scheme to be so that ciphertexts have a lower-bounded amount of min-entropy (resulting from the encryption randomness), captured by the mentioned spreadness property. This seems unavoidable for the FO transformation with explicit rejection and without any adjustment, like an additional key confirmation hash (as e.g. in [24]).

Theorem 4. *Let* PKE *be a* δ*-correct public-key encryption scheme satisfying weak* γ*-spreadness. Let* $\mathcal{A}$ *be any* $\mathsf{IND\text{-}CCA}$ *adversary against* $\mathsf{FO}[\mathsf{PKE}, H, G]$*, making* $q_D \geq 1$ *queries to the decapsulation oracle* DECAPS *and* q_H *and* q_G *queries to* $H : \mathcal{M} \to \mathcal{R}$ *and* $G : \mathcal{M} \to \mathcal{K}$*, respectively, where* H *and* G *are modeled as random oracles. Let* $q := q_H + q_G + 2q_D$*. Then, there exists a* $\mathsf{OW\text{-}CPA}$ *adversary* $\mathcal{B}$ *against* PKE *with*

$$\mathsf{ADV}[\mathcal{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}} \leq 2q\sqrt{\mathsf{ADV}^{\mathsf{OW\text{-}CPA}}_{\mathrm{PKE}}[\mathcal{B}]} + 24q^2\sqrt{\delta} + 24q\sqrt{qq_D} \cdot 2^{-\gamma/4}\,.$$

Furthermore, $\mathcal{B}$ *has a running time* $T_{\mathcal{B}} \leq T_{\mathcal{A}} + O\big(q_H \cdot q_D \cdot \mathrm{Time}[\mathsf{Enc}] + q^2\big)$*.*

We start with a proof outline, which is simplified in that it treats $\mathsf{FO}[\mathsf{PKE}, H, G]$ as an encryption scheme rather than as a KEM. We will transform the adversary $\mathcal{A}$ of the $\mathsf{IND\text{-}CCA}$ game into a $\mathsf{OW\text{-}CPA}$ adversary against the PKE in a number

[12] This seems relevant e.g. for lattice-based schemes, where the ciphertext has little (or even no) entropy for certain very unlikely choices of the key (like being all 0).

of steps. There are two main challenges to overcome. (1) We need to switch from the *deterministic* challenge ciphertext $c^* = \mathsf{Enc}_{pk}(m^*; H(m^*))$ that $\mathcal{A}$ attacks to a *randomized* challenge ciphertext $c^* = \mathsf{Enc}_{pk}(m^*; r^*)$ that $\mathcal{B}$ attacks. We do this by re-programming $H(m^*)$ to a random value right after the computation of c^*, which is equivalent to keeping H but choosing a random r^* for computing c^*. For reasons that we explain later, we do this switch from H to its re-programmed variant, denoted $H^\diamond$, in two steps, where the first step (from **Game 0** to **1**) will be "for free", and the second step (from **Game 1** to **2**) is argued using the O2H lemma ([25], we use the version given in [2], Theorem 3). (2) We need to answer decryption queries without knowing the secret key. At this point our extractable RO-simulator steps in. We replace $H^\diamond$, modelled as a random oracle, by $\mathcal{S}$, and we use its extraction interface to extract m from any correctly formed encryption $c = \mathsf{Enc}_{pk}(m; H^\diamond(m))$ and to identify incorrect ciphertexts.

One subtle issue in the argument above is the following. The O2H lemma ensures that we can find m^* by measuring one of the queries to the random oracle. However, given that also the decryption oracle makes queries to the random oracle (for performing the re-encryption check), it could be the case that one of those decryption queries is the one selected by the O2H extractor. This situation is problematic since, once we switch to $\mathcal{S}$ to deal with the decryption queries, some of these queries will be dropped (namely when $\mathcal{S}.E(c) = \emptyset$). This is problematic because, per-se, we cannot exclude that this is the one query that will give us m^*. We avoid this problem by our two-step approach for switching from H to $H^\diamond$, which ensures that the only ciphertext c that would bring us in the above unfortunate situation is the actual (randomized) *challenge ciphertext* $c^* = \mathsf{Enc}_{pk}(m^*; r^*)$, which is forbidden by the specification of the security game.

Proof (of Theorem 4). **Games 0** to **8** below gradually turn $\mathcal{A}$ into $\mathcal{B}$ (in the full version we provide pseudocode for the hybrids that compactly illustrates the change from hybrid to hybrid.). We analyze the sequence of hybrids for a fixed key pair (sk, pk). For a key pair (sk, pk), let $\mathsf{ADV}_{sk}[\mathsf{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}}$ be A's advantage, δ_{sk} the maximum decryption error probability and g_{sk} the maximum probability of any ciphertext, so that $\mathbb{E}\big[\delta_{sk}\big] \leq \delta$ and $\mathbb{E}\big[g_{sk}\big] \leq 2^{-\gamma}$, with the expectation over $(sk, pk) \leftarrow \mathsf{Gen}$.[13]

Game 0 is the $\mathsf{IND\text{-}CCA}$ game for KEMs, except that we provide G and H via arandom oracle F, by setting $H(x) := F(0||x)$ and $G(x) := F(1||x)$.[14] When convenient, we still refer to $F(0\|\cdot)$ as H and $F(1\|\cdot)$ as G. This change does not affect the view of the adversary nor the outcome of the game; therefore,

$$\Pr[b = b' \text{ in } \textbf{Game 0}] = 1/2 + \mathsf{ADV}_{sk}[\mathsf{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}}.$$

In **Game 1**, we introduce a new oracle $F^\diamond$ by setting $F^\diamond(0\|m^*) := r^\diamond$ and $F^\diamond(1\|m^*) := k^\diamond$ for uniformly random $r^\diamond \in \mathcal{R}$ and $k^\diamond \in \mathcal{K}$, while letting $F^\diamond(b\|m) := F(b\|m)$ for $m \neq m^*$ and $b \in \{0, 1\}$. We note that while the

[13] We can assume without loss of generality that pk is included in sk.

[14] These assignments seem to suggest that $\mathcal{R} = \mathcal{K}$, which may not be the case. Indeed, we understand here that $F : \mathcal{M} \to \{0, 1\}^n$ with n large enough, and $F(0||x)$ and $F(1||x)$ are then cut down to the right size.

joint behavior of $F^\diamond$ and F depends on the choice of the challenge message m^*, each one individually is a purely random function, i.e., a random oracle. In line with F, we write $H^\diamond$ for $F^\diamond(0\|\cdot)$ and $G^\diamond$ for $F^\diamond(1\|\cdot)$ when convenient.

Using these definitions, **Game 1** is obtained from **Game 0** via the following modifications. After m^* and c^* have been produced and before $\mathcal{A}$ is executed, we compute $c^\diamond := \mathsf{Enc}_{pk}(m^*; r^\diamond) = \mathsf{Enc}_{pk}(m^*; H^\diamond(m^*))$, making a query to $H^\diamond$ to obtain $r^\diamond$. Furthermore, for every decapsulation query by $\mathcal{A}$, we let DECAPS use $H^\diamond$ and $G^\diamond$ instead of H and G for checking correctness of the queried ciphertexts c_i and for computing the key K_i, *except* when $c_i = c^\diamond$ (which we may assume to happen at most once), in which case DECAPS still uses H and G. We claim that

$$\Pr[b = b' \text{ in } \textbf{Game 1}] = \Pr[b = b' \text{ in } \textbf{Game 0}] = \frac{1}{2} + \mathsf{ADV}_{sk}[\mathsf{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}}.$$

Indeed, for any decryption query c_i, we either have $\mathsf{Dec}_{sk}(c_i) =: m_i \neq m^*$ and thus $F^\diamond(b\|m_i) = F(b\|m_i)$, or else $m_i = m^*$; in the latter case we then either have $c_i = c^\diamond$, where nothing changes by definition of the game, or else $\mathsf{Enc}_{pk}(m^*; H(m^*)) = c^* \neq c_i \neq c^\diamond = \mathsf{Enc}_{pk}(m^*; H^\diamond(m^*))$, and hence the re-encryption check fails and $K_i := \perp$ in either case, without querying G or $G^\diamond$. Therefore, the input-output behavior of Decaps is not affected.

In **Game 2**, all oracle calls by Decaps (also for $c_i = c^\diamond$) and all calls by $\mathcal{A}$ are now to $F^\diamond$. Only the challenge ciphertext $c^* = \mathsf{Enc}_{pk}(m^*; H(m^*))$ is still computed using H, and thus with randomness $r^* = H(m^*)$ that is random and independent of m^* and $F^\diamond$. Hence, looking ahead, we can think of c^* as the input to the OW-CPA game that the to-be-constructed attacker $\mathcal{B}$ will attack. Similarly, $K_0^* = G(m^*)$ is random and independent of m^* and $F^\diamond$, exactly as K_1^* is, which means that $\mathcal{A}$ can only win with probability $\frac{1}{2}$.

By the O2H lemma ([2], Theorem 3), the difference between the respective probabilities of $\mathcal{A}$ guessing b in **Game 1** and **2** gives a lower bound on the success probability of a particular procedure to find an input on which F and $F^\diamond$ differ, and thus to find m^*. Formally,

$$\begin{aligned}2(q_H + q_G + 2)&\sqrt{\Pr[m' = m^* \text{ in } \textbf{Game 3}]}\\ &\geq |\Pr[b' = b \text{ in } \textbf{Game 1}] - \Pr[b' = b \text{ in } \textbf{Game 2}]|\\ &= \frac{1}{2} + \mathsf{ADV}_{sk}[\mathsf{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}} - \frac{1}{2} = \mathsf{ADV}_{sk}[\mathsf{A}]^{\mathsf{IND\text{-}CCA}}_{\mathrm{KEM}}\end{aligned}$$

where **Game 3** is identical to **Game 2** above, except that we introduce and consider a new variable m' (with the goal that $m' = m^*$), obtained as follows. Either one of the $q_H + q_G$ queries from $\mathcal{A}$ to $H^\diamond$ and $G^\diamond$ is measured, or one of the two respective queries from DECAPS to $H^\diamond$ and $G^\diamond$ upon a possible decryption query $c^\diamond$ is measured, and, in either case, m' is set to be the corresponding measurement outcome. The choice of which of these $q_H + q_G + 2$ queries to measure is done uniformly at random.[15]

[15] If this choice instructs to measure DECAPS's query to $H^\diamond$ or to $G^\diamond$ for the decryption query $c^\diamond$, but there is no decryption query $c_i = c^\diamond$, $m' := \perp$ is output instead.

We note that, since we are concerned with the measurement outcome m' only, it is irrelevant whether the game stops right after the measurement, or it continues until $\mathcal{A}$ outputs b'. Also, rather than actually measuring Decaps' classical query to $H^\diamond$ or $G^\diamond$ upon decryption query $c_i = c^\diamond$ (if instructed to do so), we can equivalently set $m' := m_i = \mathsf{Dec}_{sk}(c^\diamond)$.

For **Game 4**, we consider the function $f : \mathcal{M} \times \mathcal{R} \to \mathcal{C}$, $(m, r) \mapsto \mathsf{Enc}_{pk}(m; r)$, and we replace the random oracle $H^\diamond$ with the extractable RO-simulator $\mathcal{S}$ from Theorem 2. Furthermore, *at the very end* of the game, we invoke the extractor interface $\mathcal{S}.E$ to compute $\hat{m}_i := \mathcal{S}.E(c_i)$ for each c_i that A queried to Decaps in the course of its run. By the first statement of Theorem 2, given that the $\mathcal{S}.E$ queries take place only *after* the run of $\mathcal{A}$,

$$\Pr[m' = m^* \text{ in } \textbf{Game 4}] = \Pr[m' = m^* \text{ in } \textbf{Game 3}]\,.$$

Applying Prop. 2 for $R' := \{(m, c) : \mathsf{Dec}_{sk}(c) \neq m\}$, we get that the event $P^\dagger := \big[\forall i : \hat{m}_i = m_i \vee \hat{m}_i = \emptyset\big]$ holds except with probability $\varepsilon_1 := 128(q_H + q_D)^2 \Gamma_R / |\mathcal{R}|$ for Γ_R as in Prop. 2, which here means that $\Gamma_R / |\mathcal{R}| = \delta_{sk}$. Thus

$$\Pr[m' = m^* \wedge P^\dagger \text{ in } \textbf{Game 4}] \geq \Pr[m' = m^* \text{ in } \textbf{Game 4}] - \varepsilon_1\,.$$

In **Game 5**, we query $\mathcal{S}.E(c_i)$ *at runtime*, that is, as part of the Decaps procedure upon input c_i, right after $\mathcal{S}.RO(m)$ has been invoked as part of the re-encryption check. Since $\mathcal{S}.RO(m)$ and $\mathcal{S}.E(c_i)$ now constitute two subsequent classical queries, it follows from the contraposition of 4.b of Theorem 2 that except with probability $2 \cdot 2^{-n}$, $\hat{m}_i = \emptyset$ implies $\mathsf{Enc}_{pk}(m_i; \mathcal{S}.RO(m_i)) \neq c_i$. Applying the union bound, we find that $P^\dagger$ implies $P := \big[\forall i : \hat{m}_i = m_i \vee (\hat{m}_i = \emptyset \wedge \mathsf{Enc}_{pk}(m_i; \mathcal{S}.RO(m_i)) \neq c_i)\big]$ except with probability $q_D \cdot 2 \cdot 2^{-n}$. Furthermore, By 2.c of that same Theorem 2, each swap of a $\mathcal{S}.RO$ with a $\mathcal{S}.E$ query affects the final probability by at most $8\sqrt{2\Gamma(f)/|\mathcal{R}|} = 8\sqrt{2g_{sk}}$. Thus, setting $\varepsilon_2 := 2q_D \cdot \big((q_H + q_D) \cdot 4\sqrt{2g_{sk}} + 2^{-n}\big)$,

$$\Pr[m' = m^* \wedge P \text{ in } \textbf{Game 5}] \geq \Pr[m' = m^* \wedge P^\dagger \text{ in } \textbf{Game 4}] - \varepsilon_2$$

In **Game 6**, Decaps uses $\hat{m}_i$ instead of m_i to compute K_i. That is, it sets $K_i := \perp$ if $\hat{m}_i = \emptyset$ and $K_i := G^\diamond(\hat{m}_i)$ otherwise. Also, if instructed to output $m' := m_i$ where $c_i = c^\diamond$, then the output is set to $m' := \hat{m}_i$ instead. In all cases, Decaps still queries $\mathcal{S}.RO(m_i)$, so that the interaction pattern between Decaps and $\mathcal{S}.RO$ remains as in **Game 5**. Here, we note that if the event $P_i := \big[\hat{m}_i = m_i \vee (\hat{m}_i = \emptyset \wedge \mathsf{Enc}_{pk}(m_i; \mathcal{S}.RO(m_i)) \neq c_i)\big]$ holds for a given i then the above change will not affect Decaps' response K_i, and thus neither the probability of P_{i+1}. Therefore, by induction, $\Pr[P \text{ in } \textbf{Game 6}] = \Pr[P \text{ in } \textbf{Game 5}]$, and since conditioned on the event P the two games are identical, we have

$$\Pr[m' = m^* \wedge P \text{ in } \textbf{Game 6}] = \Pr[m' = m^* \wedge P \text{ in } \textbf{Game 5}].$$

In **Game 7**, instead of obtaining m' by measuring a random query of $\mathcal{A}$ to either $\mathcal{S}.RO$ or G, or outputting $\hat{m}_i$ with $c_i = c^\diamond$, here m' is obtained by measuring a random query of $\mathcal{A}$ to either $\mathcal{S}.RO$ or G, or outputting $\hat{m}_i$ for a *random* $i \in \{1, \ldots, q_D\}$, where the first cases is chosen with probability $(q_H + q_G)/(q_H + q_G + 2q_D)$, and the second otherwise. As conditioned on choosing the first case, or the second one with $i = i_\diamond$, **Game 7** equals **Game 6**, we have

$$\Pr[m' = m^* \text{ in } \textbf{Game 7}] \geq \frac{q_H + q_G + 2}{q_H + q_G + 2q_D} \cdot \Pr[m' = m^* \text{ in } \textbf{Game 6}]\,.$$

In **Game 8**, we observe that the response to the query $\mathcal{S}.RO(m^*)$, introduced in **Game 1** to compute $c^\diamond$, and the responses to the queries that Decaps makes to $\mathcal{S}.RO$ on input m_i do not affect the game anymore, so we can drop these queries, or, equivalently, move them to the end of the game's execution. Invoking 2.c of Theorem 2 again and setting $\varepsilon_3 = (q_D + 1) \cdot q_H \cdot 8\sqrt{2g_{sk}}$, we get

$$\Pr[m' = m^* \text{ in } \textbf{Game 8}] \geq \Pr[m' = m^* \text{ in } \textbf{Game 7}] - \varepsilon_3\,,$$

We now see that **Game 8** works without knowledge of the secret key sk, and thus constitutes a OW-CPA attacker $\mathcal{B}$ against PKE, which takes as input a public key pk and an encryption c^* of a random message $m^* \in \mathcal{M}$, and outputs m^* with the given probability, i.e, $\mathsf{ADV}_{sk}[\mathsf{B}]^{\textsf{OW-CPA}}_{\text{PKE}} \geq \Pr[m' = m^* \text{ in } \textbf{Game 8}]$. We note that the oracle $G^\diamond$ can be simulated using standard techniques. Backtracking all the above (in)equalities and setting $\varepsilon_{23} := \varepsilon_2 + \varepsilon_3$, $q_{HG} := q_H + q_G$ etc. and $q := q_H + q_G + 2q_D$, we get the following bounds,

$$\begin{aligned}
\mathsf{ADV}_{sk}[\mathcal{A}]^{\textsf{IND-CCA}}_{\text{KEM}} &\leq 2(q_{HG}+2)\sqrt{\frac{q_{HG}+2q_D}{q_{HG}+2}\left(\mathsf{ADV}_{sk}[\mathsf{B}]^{\textsf{OW-CPA}}_{\text{PKE}} + \varepsilon_3\right) + \varepsilon_1 + \varepsilon_2} \\
&\leq 2(q_{HG}+2q_D)\sqrt{\mathsf{ADV}_{sk}[\mathsf{B}]^{\textsf{OW-CPA}}_{\text{PKE}} + \varepsilon_{23}} + 2(q_{HG}+2)\sqrt{\varepsilon_1} \\
&\leq 2q\left(\sqrt{\mathsf{ADV}_{sk}[\mathsf{B}]^{\textsf{OW-CPA}}_{\text{PKE}}} + \sqrt{\varepsilon_{23}} + \sqrt{\varepsilon_1}\right) \qquad (27)
\end{aligned}$$

and

$$\begin{aligned}
\sqrt{\varepsilon_{23}} &= \sqrt{2q_D \cdot \left(4\big((q_H+q_D) + (q_D+1)q_H\big)\sqrt{2g_{sk}} + 2^{-n}\right)} \\
&\leq 6\sqrt{q_H q_D} \cdot \left(g_{sk}^{1/4} + 2^{-n/2}\right) \leq 12\sqrt{q q_D} \cdot g_{sk}^{1/4}\,, \qquad (28)
\end{aligned}$$

where we have used that $2^{-n} \leq g_{sk} \leq 1$ in the last line. Taking the expectation over $(sk, pk) \leftarrow \mathsf{Gen}$, applying Jensen's inequality and using $q_H + q_D \leq q$ once more, we get the claimed bound. Finally, we note that the runtime of $\mathcal{B}$ is given by $T_{\mathcal{B}} = T_{\mathcal{A}} + T_{\text{Decaps}} + T_G + T_{\mathcal{S}}$, where apart from its oracle queries Decaps runs in time linear in q_D, and by Theorem 2 $\mathcal{S}$ and G can be simulated in time $T_{\mathcal{S}} = O\big(q_{RO} \cdot q_E \cdot \text{Time}[f] + q_{RO}^2\big) = O\big(q_H \cdot q_D \cdot \text{Time}[\mathsf{Enc}] + q^2\big)$. □

Acknowledgements

The authors thank Andreas Hülsing and Kathrin Hövelmanns for helpful discussions, and Eike Kiltz and anonymous referees for helpful comments on an earlier version of this article. JD was funded by ERC-ADG project 740972 (ALGSTRONGCRYPTO). SF was partly supported by the EU Horizon 2020 Research and Innovation Program Grant 780701 (PROMETHEUS). CM was funded by a NWO VENI grant (Project No. VI.Veni.192.159). CS was supported by a NWO VIDI grant (Project No. 639.022.519).

References

1. G. Alagic, C. Majenz, A. Russell, and F. Song. Quantum-access-secure message authentication via blind-unforgeability. In A. Canteaut and Y. Ishai, editors, *Advances in Cryptology – EUROCRYPT 2020*, pages 788–817, Cham, 2020. Springer International Publishing.
2. A. Ambainis, M. Hamburg, and D. Unruh. Quantum security proofs using semi-classical oracles. In A. Boldyreva and D. Micciancio, editors, *Advances in Cryptology – CRYPTO 2019*, pages 269–295, Cham, 2019. Springer International Publishing.
3. N. Bindel, M. Hamburg, K. Hövelmanns, A. Hülsing, and E. Persichetti. Tighter proofs of cca security in the quantum random oracle model. In D. Hofheinz and A. Rosen, editors, *Theory of Cryptography*, pages 61–90, Cham, 2019. Springer International Publishing.
4. D. Boneh, Ö. Dagdelen, M. Fischlin, A. Lehmann, C. Schaffner, and M. Zhandry. Random oracles in a quantum world. In D. H. Lee and X. Wang, editors, *Advances in Cryptology – ASIACRYPT 2011*, pages 41–69, Berlin, Heidelberg, 2011. Springer.
5. G. Brassard, D. Chaum, and C. Crépeau. Minimum disclosure proofs of knowledge. *Journal of Computer and System Sciences*, 37(2):156 – 189, 1988.
6. M. Chase, D. Derler, S. Goldfeder, J. Katz, V. Kolesnikov, C. Orlandi, S. Ramacher, C. Rechberger, D. Slamanig, X. Wang, and G. Zaverucha. The picnic signature scheme, design document v2.1, 2019.
7. M. Chase, D. Derler, S. Goldfeder, C. Orlandi, S. Ramacher, C. Rechberger, D. Slamanig, and G. 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.
8. M.-S. Chen, A. Hülsing, J. Rijneveld, S. Samardjiska, and P. Schwabe. From 5-pass mq-based identification to mq-based signatures. In J. H. Cheon and T. Takagi, editors, *Advances in Cryptology – ASIACRYPT 2016*, pages 135–165, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.
9. A. Chiesa, P. Manohar, and N. Spooner. Succinct arguments in the quantum random oracle model. In D. Hofheinz and A. Rosen, editors, *Theory of Cryptography*, pages 1–29, Cham, 2019. Springer International Publishing.
10. K.-M. Chung, S. Fehr, Y.-H. Huang, and T.-N. Liao. On the compressed-oracle technique, and post-quantum security of proofs of sequential work. Cryptology ePrint Archive, Report 2020/1305, 2020. https://eprint.iacr.org/2020/1305, to appear in *Advances in Cryptology – EUROCRYPT 2021*.
11. J. Czajkowski, C. Majenz, C. Schaffner, and S. Zur. Quantum lazy sampling and game-playing proofs for quantum indifferentiability. Cryptology ePrint Archive, Report 2019/428, 2019. https://eprint.iacr.org/2019/428.

12. J. Don, S. Fehr, C. Majenz, and C. Schaffner. Security of the Fiat-Shamir transformation in the quantum random-oracle model. In A. Boldyreva and D. Micciancio, editors, *Advances in Cryptology – CRYPTO 2019*, pages 356–383, Cham, 2019. Springer International Publishing.
13. E. Fujisaki and T. Okamoto. How to enhance the security of public-key encryption at minimum cost. In *Public Key Cryptography*, pages 53–68, Berlin, Heidelberg, 1999. Springer Berlin Heidelberg.
14. O. Goldreich, S. Micali, and A. Wigderson. Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems. *J. ACM*, 38(3):690–728, July 1991.
15. Y. Hamoudi and F. Magniez. Quantum time-space tradeoff for finding multiple collision pairs, 2020.
16. D. Hofheinz, K. Hövelmanns, and E. Kiltz. A modular analysis of the Fujisaki-Okamoto transformation. In Y. Kalai and L. Reyzin, editors, *Theory of Cryptography*, pages 341–371, Cham, 2017. Springer International Publishing.
17. D. Kales and G. Zaverucha. Improving the performance of the picnic signature scheme. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, pages 154–188, 2020.
18. S. Katsumata, K. Kwiatkowski, F. Pintore, and T. Prest. Scalable ciphertext compression techniques for post-quantum kems and their applications. In *Advances in Cryptology – ASIACRYPT 2020*, pages 289–320. Springer International Publishing, 2020.
19. Q. Liu and M. Zhandry. On finding quantum multi-collisions. In Y. Ishai and V. Rijmen, editors, *Advances in Cryptology – EUROCRYPT 2019*, pages 189–218, Cham, 2019. Springer International Publishing.
20. Q. Liu and M. Zhandry. Revisiting post-quantum Fiat-Shamir. In A. Boldyreva and D. Micciancio, editors, *Advances in Cryptology – CRYPTO 2019*, pages 326–355, Cham, 2019. Springer International Publishing.
21. M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, New York, NY, USA, 10th edition, 2011.
22. R. Pass. On deniability in the common reference string and random oracle model. In D. Boneh, editor, *Advances in Cryptology - CRYPTO 2003*, pages 316–337, Berlin, Heidelberg, 2003. Springer Berlin Heidelberg.
23. R. Pass. *Alternative variants of zero-knowledge proofs*. PhD thesis, KTH Stockholm, 2004.
24. E. E. Targhi and D. Unruh. Post-quantum security of the Fujisaki-Okamoto and OAEP transforms. In M. Hirt and A. Smith, editors, *Theory of Cryptography*, pages 192–216, Berlin, Heidelberg, 2016. Springer Berlin Heidelberg.
25. D. Unruh. Revocable quantum timed-release encryption. In P. Q. Nguyen and E. Oswald, editors, *Advances in Cryptology – EUROCRYPT 2014*, pages 129–146, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.
26. D. Wikström. Special soundness revisited. Cryptology ePrint Archive, Report 2018/1157, 2018. https://ia.cr/2018/1157.
27. M. Zhandry. How to record quantum queries, and applications to quantum indifferentiability. In A. Boldyreva and D. Micciancio, editors, *Advances in Cryptology – CRYPTO 2019*, pages 239–268, Cham, 2019. Springer International Publishing. Full Version (1 March 2019): https://eprint.iacr.org/2018/276/20190301:184107.