The Impossibility of Post-Quantum Public Indifferentiability for Merkle-Damgård*

Akinori Hosoyamada

NTT Social Informatics Laboratories, Tokyo, Japan
NTT Research Center for Theoretical Quantum Information, Atsugi, Japan
akinori.hosoyamada@ntt.com

Abstract. The Merkle-Damgård construction (in its strengthened form as used in SHA-256 and SHA-512, the untruncated members of SHA-2) is not classically indifferentiable from a Variable-Input-Length (VIL) random oracle because of the length-extension attack. Nevertheless, Dodis, Ristenpart, and Shrimpton showed that Merkle-Damgård is publicly indifferentiable, a weaker notion that still justifies replacing a VIL random oracle by Merkle-Damgård in many security proofs when all inputs to a random oracle are public (e.g., Fiat-Shamir and full-domain-hash signatures).
In this paper, we show that this replacement fails in the post-quantum setting: (Strengthened) Merkle-Damgård is not publicly indifferentiable from a VIL random oracle against quantum distinguishers with superposition access to the underlying primitive (while construction queries remain classical), even if the compression function is ideally random. We first formalize post-quantum public indifferentiability so that the corresponding composition theorem extends to the quantum random oracle model. We also introduce a post-quantum version of sequential indifferentiability, an even weaker notion. We then prove that (strengthened) Merkle-Damgård satisfies neither notion by showing that an explicit quantum distinguisher achieves non-negligible advantage against any efficient simulator, using Zhandry's compressed-oracle technique. We thus obtain an explicit, conjecture-free separation between an indifferentiability-style notion in the classical setting and its post-quantum analogue.

Keywords: symmetric cryptography · Merkle-Damgård · post-quantum security · public indifferentiability · sequential indifferentiability.

1 Introduction

Security proofs for cryptosystems often rely on idealized models such as the Random Oracle Model (ROM) [2]. In practice, however, random oracles are instantiated by concrete hash functions with nontrivial internal structure—for example, the Merkle-Damgård construction [11,26] underlying SHA-2 [28]. It is

* This article is the full version of the paper with the same title accepted to Crypto 2026, ©IACR 2026.

non-trivial whether ROM proofs remain sound once the random oracle is replaced with a structured construction. Re-proving security for each cryptosystem under each specific hash structure is costly, so a systematic justification is desirable.

This issue can be addressed using the notion of indifferentiability, which was introduced by Maurer et al [25]. If a construction is indifferentiable from a random oracle, then the composition theorem [25,31] allows it to safely replace the random oracle in security proofs, as long as the security notion can be defined via a single-stage game. For instance, the sponge construction [4,5] is indifferentiable from a Variable-Input-Length (VIL) random oracle, which justifies the use of SHA-3 [29] in many cryptosystems proven secure in the ROM.

By contrast, the Merkle-Damgård construction most widely used in practice—namely, the strengthened variant that pads a message by simply appending a length-encoding at its end as in SHA-256 and SHA-512, the untruncated members of SHA-2—is not indifferentiable from a VIL random oracle due to the length-extension attack [10]. It can be made indifferentiable if some modifications are made (e.g., prefix-free encoding is used) or input lengths are fixed [10]. However, widely deployed standards such as SHA-256, SHA-512, and RIPEMD-160 (standardized by ISO/IEC [19] and used in Bitcoin [27]) are typically used as-is, without such modifications. Moreover, many applications apply hash functions to variable-length inputs—for example, in signature schemes.

Nevertheless, many cryptosystems appear to remain secure in practice even when the (strengthened) Merkle-Damgård is used in place of a VIL random oracle. To bridge this gap, Dodis et al. [13] introduced a weaker notion called *public indifferentiability*. Roughly speaking, when a cryptosystem only hashes public values, public indifferentiability suffices (via an appropriate composition theorem) to preserve many security proofs in the ROM. Prominent examples of such cryptosystems include Fiat-Shamir signatures [16] and Full Domain Hash (FDH) signatures [2]. Dodis et al. proved that the Merkle-Damgård construction satisfies public indifferentiability, thereby ensuring the security of these schemes even when a VIL random oracle is instantiated by Merkle-Damgård.

(Post-)Quantum Security. Nowadays, post-quantum security is crucial in the analysis of practical cryptographic systems; hash functions are no exception. In the Quantum Random Oracle Model (QROM) [6], an adversary is a quantum algorithm that may query the random oracle in superposition. When analyzing a structured hash construction built from a smaller primitive (e.g., the compression function in Merkle-Damgård or the permutation in sponge), it is natural to model that adversaries have oracle access to that primitive. In the post-quantum setting, we allow these primitive queries to be in quantum superposition.

In this model, the sponge construction has been proven indifferentiable against quantum adversaries [1]. Merkle-Damgård is also quantum indifferentiable from a random oracle *if a modification is made so a prefix-free encoding is used, or input lengths are fixed* [35,1]. However, since the most practically common, strengthened Merkle-Damgård (which is used in various standards [19,28] including SHA-

2 and typically deployed without such modifications) is not indifferentiable from a VIL random oracle classically, it cannot be quantum indifferentiable either.

While public indifferentiability is known to hold for the (strengthened) Merkle-Damgård classically, whether it extends to the (post-)quantum setting has remained open. This question is important not only theoretically but also practically. For example, many papers (e.g., [6,34,32,20,15,21,14,33,17,22]) prove the post-quantum security of Fiat-Shamir or FDH signatures in the (VIL) QROM. If Merkle-Damgård were post-quantum public indifferentiable from a VIL random oracle, then a broader set of practical candidates (including SHA-2) could instantiate quantum random oracles in those schemes, while automatically preserving the soundness of the security proofs. As a related concrete example, the pre-hash variant HashML-DSA in the recent NIST standard [30] allows the use of SHA-256 or SHA-512 to pre-hash messages before signing. For such a pre-hashing step, a post-quantum public-indifferentiability theorem for Merkle-Damgård would provide a generic justification for using SHA-2.

Motivated by this, we investigate the following question:

> *Does the (strengthened) Merkle-Damgård construction satisfy public indifferentiability in the (post-)quantum setting?*

1.1 Our Contributions

We answer this question in the negative: the (strengthened) Merkle-Damgård is *not* post-quantum publicly indifferentiable from a variable-input-length (VIL) random oracle, even when the compression function is an ideal random function. The details are as follows.

Formalizing Post-Quantum Public and Sequential Indifferentiability. Although public indifferentiability is well understood classically, we are not aware of any prior work that formalizes a post-quantum analogue that considers quantum adversaries with superposition oracle access. Thus, we first formalize post-quantum public indifferentiability in a setting where a distinguisher $\mathcal{D}$ may make quantum superposition queries to the underlying primitive (and hence simulators are fully quantum), while $\mathcal{D}$ makes only classical queries to the construction oracle. This formulation preserves the composition theorem for typical QROM proofs for classical cryptosystems, which assume honest parties are classical. Consequently, any hash construction satisfying our notion can soundly instantiate a quantum random oracle in many post-quantum security proofs when all hashed inputs are public (e.g., those for Fiat-Shamir and FDH signatures).

We also formalize a post-quantum analogue of sequential indifferentiability [24], which is a notion even weaker than public indifferentiability, by naturally restricting the class of distinguishers for post-quantum public indifferentiability.

Impossibility for (Strengthened) Merkle-Damgård to Achieve Post-Quantum Public/Sequential Indifferentiability. We then prove that the

(strengthened) Merkle-Damgård construction satisfies neither post-quantum public indifferentiability nor post-quantum sequential indifferentiability, even if the underlying compression function is ideally random. More precisely, we explicitly construct an efficient distinguisher $\mathcal{D}$ that breaks the post-quantum sequential indifferentiability, and thereby post-quantum public indifferentiability, of the (strengthened) Merkle-Damgård.

As a result, we obtain an explicit, conjecture-free separation between a classical indifferentiability-style notion and its post-quantum analogue: The (strengthened) Merkle-Damgård satisfies classical public/sequential indifferentiability, yet provably fails to achieve their post-quantum versions.

Overview of the Distinguisher. Let $\mathsf{MD}[h]$ be the (strengthened) Merkle-Damgård construction, where $h : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$ is the underlying compression function. Let $IV \in \{0,1\}^n$ and pad be the initial value and the padding function (Merkle-Damgård strengthening), respectively[1].

Our distinguisher $\mathcal{D}$ runs one of two subroutines, $\mathcal{D}_0$ and $\mathcal{D}_1$, at random. Roughly, $\mathcal{D}_0$ succeeds if a simulator $\mathcal{S}$ "records" (i.e., makes non-negligible quantum correlation with) the query input by $\mathcal{D}_0$ to $\mathcal{S}$ in superposition during the initial interactions, whereas $\mathcal{D}_1$ succeeds if $\mathcal{S}$ avoids creating such correlation throughout those interactions. Crucially, $\mathcal{S}$ must commit to this behavior before it can tell which of the two subroutines it is facing, yielding a non-negligible distinguishing advantage.

More specifically, both $\mathcal{D}_0$ and $\mathcal{D}_1$ first put a variable $a \in \{0,1\}^{m/2}$ in uniform superposition and query $(\mathsf{pad}(a), IV)$ to the primitive oracle, which is h in the real world and $\mathcal{S}$ in the ideal world. In the real world, this response (denoted by c) is exactly $\mathsf{MD}[h](a)$.

The subroutine $\mathcal{D}_0$ then tries to uncompute this interaction and checks, by applying Hadamards to the a-register, whether $\mathcal{S}$ has created any non-negligible correlation with a. If it has, the interference is disturbed and $\mathcal{D}_0$ detects it.

Meanwhile, the other subroutine $\mathcal{D}_1$ makes another query to the primitive oracle, using c as the chaining value, so that the response (denoted by d) will match $\mathsf{MD}[h](\mathsf{pad}(a)||0^{m/2})$ in the real world. It then queries both a and $\mathsf{pad}(a)||0^{m/2}$ to the construction oracle (which is $\mathsf{MD}[h]$ in the real world and a VIL random oracle RO in the ideal world), and checks that the responses match c and d, respectively. In the ideal world, passing this second check requires $\mathcal{S}$ to know which value of a the chaining value c came from; otherwise the response d to the second primitive query is essentially independent of $\mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$.

Thus $\mathcal{D}_0$ catches simulators that learn a, whereas $\mathcal{D}_1$ catches simulators that do not. Since the two subroutines are indistinguishable to the simulator until after the critical early interactions, no efficient simulator can satisfy both requirements simultaneously. The formal proof implements this intuition using Zhandry's compressed oracle technique [35] to handle superposition queries and entanglement.

[1] See Section 2 for the formal and concrete definition.

Implications. Our attack does not immediately break real-world schemes that instantiate random oracles via the (strengthened) Merkle-Damgård. Such schemes may remain secure under tailored analyses. What fails is the generic substitution guarantee afforded by classical public indifferentiability: Even when only public values are hashed, the (strengthened) Merkle-Damgård cannot be treated unconditionally as a VIL random oracle in the post-quantum setting. Accordingly, instantiating a quantum random oracle with Merkle-Damgård calls for extra care (e.g., use the prefix-free padding) and/or scheme-specific proofs.

1.2 Related Works

Quantum indifferentiability has been studied in a number of works. Carstens et al. [9] argued that, under the conjecture that quantum indifferentiability requires classical indifferentiability with stateless simulators, the sponge and Feistel constructions are not quantum indifferentiable. They further showed that this conjecture holds for *perfect* quantum indifferentiability (i.e., zero distinguishing advantage) assuming an information-theoretic conjecture. Subsequently, however, the sponge construction was proven quantum indifferentiable [1]. Zhandry [36] showed the existence of a construction which is built from an ideal cipher and quantum indifferentiable from an arbitrary-size random oracle, and that generic groups are quantum (reset) indifferentiable from ideal ciphers. In later work, Zhandry [37] discussed indifferentiability from a random unitary operator and showed that the pseudo-random unitary construction of [23] is not indifferentiable. Carolan et al. proved that one-round sponge is (quantum) indifferentiable even if precomputation is allowed for adversaries [8]. Benedikt and Fischlin investigated, mainly from an attack perspective, how the order of hashing different types of inputs (e.g., public keys and messages) affects the post-quantum security of Fiat-Shamir schemes when using the Merkle-Damgård or sponge construction [3]. In [3], they also mentioned that it is unknown whether the public indifferentiability of Merkle-Damgård carries over to the (post-)quantum setting.

Coron et al. [10] showed that the Merkle-Damgård construction can also be indifferentiable by truncating some output bits, instead of using a prefix-free padding. Concrete instances of such constructions include some members of the SHA-2 family, e.g., SHA-384. It is not known whether this indifferentiability lifts to the (post-)quantum setting.

2 Preliminaries

For a bit string x, its length is denoted by $|x|$. The empty string has length 0. For an integer $\ell \geq 1$, $\left(\{0,1\}^{\ell}\right)^{+}$ denotes the set of bit strings whose length is a positive multiple of ℓ, and $\{0,1\}^{\leq \ell}$ denotes the set of bit strings of length at most ℓ. We identify $\{0,1\}^{m} \times \{0,1\}^{n}$ with $\{0,1\}^{m+n}$ via concatenation: $(w,z) \in \{0,1\}^{m} \times \{0,1\}^{n}$ corresponds to $w||z \in \{0,1\}^{m+n}$. We also view $\{0,1\}^{n}$ as an n-dimensional vector space over $\mathbb{F}_2$, where addition is bitwise XOR.

We write $y \leftarrow \mathcal{A}(x)$ for the event that a (classical or quantum) algorithm $\mathcal{A}$, on input x, outputs the classical bit string y. A classical or quantum algorithm is efficient if it runs in time polynomial in the security parameter. For a vector $|\phi\rangle$ in a Hilbert space (resp., a linear operator L on a Hilbert space), $\| |\phi\rangle \|$ (resp., $\|L\|_{\mathrm{op}}$) denotes its Euclidean norm (resp., operator norm). If $\| |\phi\rangle - |\psi\rangle \| \leq \varepsilon$ (resp., $\|L - L'\|_{\mathrm{op}} \leq \varepsilon$), then we say that $|\phi\rangle$ is ε-close to $|\psi\rangle$ (resp., L is ε-close to L'), and we write $|\phi\rangle \overset{\varepsilon}{\approx} |\psi\rangle$ (resp., $L \overset{\varepsilon}{\approx} L'$).

For a fixed-input-length (FIL) function $g : \{0,1\}^m \to \{0,1\}^n$, we model its quantum oracle as the unitary O_g defined by

$$O_g |x\rangle_X |y\rangle_Y = |x\rangle_X |y \oplus g(x)\rangle_Y ,$$

where X and Y are an m-qubit register and an n-qubit register, respectively. The quantum oracle of a VIL function $f : \{0,1\}^{\leq \ell} \to \{0,1\}^n$ is defined analogously, assuming that register X has sufficiently many qubits and that the input length is implicitly encoded in x.

To simplify notation, we occasionally permute tensor factors without further comment. For example, a state $|\psi\rangle = \sum_{x,y,z} \alpha_{xyz} |x\rangle_X |y\rangle_Y |z\rangle_Z$ may also be written as $\sum_{x,y,z} \alpha_{xyz} |y\rangle_Y |x\rangle_X |z\rangle_Z$ or $\sum_{x,y,z} \alpha_{xyz} |z\rangle_Z |y\rangle_Y |x\rangle_X$. Similarly, when the identity operator on the remaining registers is clear from context, we may write a linear operator L in place of $L \otimes I$ or $I \otimes L$.

We sometimes denote the state defined in Eq. (i) by $|(i)\rangle$. This shorthand is used to simplify long derivations and surrounding text. If Eq. (i) has the form $|\psi\rangle \overset{\varepsilon}{\approx} \cdots \overset{\varepsilon'}{\approx} |\phi\rangle$ (or $|\psi\rangle = \cdots = |\phi\rangle$), then $|(i)\rangle$ refers to the explicit ket vector $|\phi\rangle$ appearing in that equation (typically the rightmost ket expression in that equation). We use this shorthand only when the intended vector expression is unambiguous.

Merkle-Damgård Construction. Let m and n be integers that are powers of two. For a compression function $h : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$ and an initialization vector $IV \in \{0,1\}^n$ the plain (unpadded) Merkle-Damgård construction $\mathsf{pMD}_{IV}[h] : (\{0,1\}^m)^+ \to \{0,1\}^n$ is defined by

$$\mathsf{pMD}_{IV}[h] := h(x_i, h(x_{i-1}, \ldots, h(x_1, IV))) \tag{1}$$

for $x_1, \ldots, x_i \in \{0,1\}^m$.

For a positive integer ℓ, let $\mathsf{pad} : \{0,1\}^{\leq \ell} \to (\{0,1\}^m)^+$ be an injective padding function such that

$$\mathsf{pad}(x) = x || \mathsf{enc}(|x|) \tag{2}$$

holds, where $\mathsf{enc}(|x|)$ is a bit string that encodes the length $|x|$ (as in the usual Merkle-Damgård strengthening). The strengthened Merkle-Damgård construction $\mathsf{MD}_{IV}[h] : \{0,1\}^{\leq \ell} \to \{0,1\}^n$ is defined by

$$\mathsf{MD}_{IV}[h](x) := \mathsf{pMD}_{IV}[h]\big(\mathsf{pad}(x)\big). \tag{3}$$

Since most widely deployed Merkle-Damgård hash functions including the SHA-2 family [28] (especially, SHA-256 and SHA-512) and RIPEMD-160 [19] adopt the strengthened Merkle-Damgård construction, we refer to this strengthened variant simply as Merkle-Damgård. When the concrete value of IV is not important, we write $\mathsf{MD}[h]$ instead of $\mathsf{MD}_{IV}[h]$. We assume that the security parameter is n, ℓ is sufficiently large, and m is in $\Omega(n)$.

Example: SHA-256 and SHA-512. For SHA-256 (resp., SHA-512), we have $(n, m) = (256, 512)$ (resp., $(512, 1024)$). The function $\mathsf{enc}(|x|)$ of SHA-256 (resp., SHA-512) is defined as $\mathsf{enc}(|x|) = 10^k || \langle x \rangle_{\mathsf{bin}}$, where k is the smallest non-negative integer satisfying $|x| + 1 + k \equiv 448 \pmod{512}$ (resp., $\equiv 896 \pmod{1024}$), and $\langle x \rangle_{\mathsf{bin}}$ is a 64-bit (resp., 128-bit) binary representation of $|x|$.

Remark 1. Zhandry [35] showed that the Merkle-Damgård construction is quantumly indifferentiable from a random oracle, assuming that a prefix-free encoding is involved. This result is a quantum version of the work by Coron et al [10][2]. Assuming prefix-free encoding essentially means that the padding function of Eq. (2) is replaced by an encoding whose range is prefix-free[3], or the length of the inputs is fixed. Meanwhile, various standards, including SHA-256, SHA-512, and RIPEMD-160, are typically used without such modifications, and many applications hash variable-length inputs. Our motivation is to study the post-quantum security of the most practically common (strengthened) version of Merkle-Damgård, in alignment with real-world implementations.

2.1 Compressed Oracle Technique

Recording quantum queries to an oracle is generally difficult due to the no-cloning theorem. For quantum random oracles, however, Zhandry's compressed-oracle technique [35] partially overcomes this difficulty.

Let RO be a (VIL) random oracle with domain $\mathcal{X} := \{0,1\}^{\leq \ell}$ and range $\{0,1\}^n$. The quantum random oracle is primarily modeled by (1) sampling a function $f : \mathcal{X} \to \{0,1\}^n$ uniformly at random and then (2) providing oracle access to the unitary $O_f : |x\rangle_X |y\rangle_Y \mapsto |x\rangle_X |y \oplus f(x)\rangle_Y$. As oracles perfectly indistinguishable from this, the *standard oracle* and the *compressed oracle* are defined.

Standard Oracle. The standard oracle is obtained by purifying the randomness of the choice of f into a uniform superposition. Specifically:

1. The oracle sets its initial state as $|\mathsf{Init}_{\mathsf{StO}}\rangle_F := \frac{1}{\sqrt{N}} \sum_{f \in \mathsf{Funcs}} |f\rangle_F$, where Funcs is the set of all functions from $\mathcal{X}$ to $\{0,1\}^n$, $N := |\mathsf{Funcs}|$, and F is a register to store the truth table of $f \in \mathsf{Funcs}$.

[2] The recent work by Alagic et al. [1] pointed out a gap in the proof of [35] and explained how it could be fixed. Similarly, [7] also discussed a gap in the classical proof of [10] and gave a careful treatment addressing it.

[3] Note that the padding function of Eq. (2) is not prefix-free because, e.g., $\mathsf{pad}(x)$ is a prefix of $\mathsf{pad}(\mathsf{pad}(x))$.

2. Queries are responded by applying the unitary operator $\mathsf{StO} : |x\rangle_X |y\rangle_Y |f\rangle_F \mapsto |x\rangle_X |y \oplus f(x)\rangle_Y |f\rangle_F$.

The standard oracle is perfectly indistinguishable from the original quantum random oracle.

Fix an arbitrary ordering of the elements of $\mathcal{X}$, and denote the i-th element by $x^{(i)}$. We represent the truth table of f as the $n|\mathcal{X}|$-bit string $f(x^{(0)})||f(x^{(1)})||\cdots||f(x^{(|\mathcal{X}|-1)})$ and view register F as consisting of subregisters $F[x]$ that store $f(x)$ for $x \in \mathcal{X}$. In particular, $|f\rangle_F = \bigotimes_{x\in\mathcal{X}} |f(x)\rangle_{F[x]}$.

Database. A core concept in the compressed oracle technique is the *database*. Consider partial functions $f : \mathcal{X} \to \{0,1\}^n \cup \{\perp\}$, where $f(x) = \perp$ indicates that the value at x is undefined. The truth table of such a partial function can be identified with the set of pairs $\{(x, f(x)) : x \in \mathcal{X},\ f(x) \neq \perp\}$.

Conversely, any set $D := \{(x_1, y_1), \ldots, (x_t, y_t)\} \subset \mathcal{X} \times \{0,1\}^n$ with distinct x_i's can be identified with the partial function f_D such that $f_D(x_i) = y_i$ for $i = 1, \ldots, t$ and $f_D(x) = \perp$ for $x \in \mathcal{X} \setminus \{x_1, \ldots, x_t\}$. We call such a set D a *database*. The empty database is denoted by $\varnothing$.

We use partial functions and databases interchangeably: for a database D, we also write $D(x_i) = y_i$ and $D(x) = \perp$ for $x \notin \{x_1, \ldots, x_t\}$. For any database D and any $(x, y) \in \mathcal{X} \times \{0,1\}^n$, let $D[x \mapsto y]$ be the database such that $D[x \mapsto y](x) = y$ and $D[x \mapsto y](x') = D(x')$ for $x' \neq x$. Similarly, $D[x_1 \mapsto y_1, x_2 \mapsto y_2]$ denotes the database that assigns $x_i \mapsto y_i$ for $i \in \{1, 2\}$ and agrees with D elsewhere.

Compressed Oracle. From now on, we extend the register F so that it can store databases (i.e., partial functions) as well as fully defined functions. Equivalently, for each x, the Hilbert space of $F[x]$ is spanned by $\{|y\rangle : y \in \{0,1\}^n \cup \{\perp\}\}$. We also extend StO so that $\mathsf{StO}\,|x\rangle_X |y\rangle_Y |D\rangle_F = |x\rangle_X |y\rangle_Y |D\rangle_F$ if $D(x) = \perp$.

For $y \in \{0,1\}^n$, let $|\hat{y}\rangle := \sum_{w\in\{0,1\}^n} \frac{(-1)^{w\cdot y}}{2^{n/2}} |w\rangle$, where $w \cdot y$ denotes the dot product over $\mathbb{F}_2$. Define the unitary $\mathsf{Comp}_{F[x]}$ by $\mathsf{Comp}_{F[x]} := |\perp\rangle\langle\hat{0}|_{F[x]} + |\hat{0}\rangle\langle\perp|_{F[x]} + \sum_{y\in\{0,1\}^n\setminus\{0^n\}} |\hat{y}\rangle\langle\hat{y}|_{F[x]}$, and (with a slight abuse of notation) set $\mathsf{Comp}_F := \bigotimes_{x\in\mathcal{X}} \mathsf{Comp}_{F[x]}$. Note that $\mathsf{Comp}^\dagger_{F[x]} = \mathsf{Comp}_{F[x]}$ and hence $\mathsf{Comp}^\dagger_F = \mathsf{Comp}_F$.

Definition 1 (Compressed Oracle). *The compressed oracle initializes register F in the state $|\varnothing\rangle_F$ and answers quantum queries by applying the unitary*

$$\mathsf{cO}_{XYF} := \mathsf{Comp}_F \cdot \mathsf{StO}_{XYF} \cdot \mathsf{Comp}_F. \tag{4}$$

This oracle is perfectly indistinguishable from the standard oracle (and thus from the original quantum random oracle), since $\mathsf{Comp}_F |\varnothing\rangle_F = |\mathsf{Init}_{\mathsf{StO}}\rangle_F$ and Comp_F acts only on the oracle register F.

Let $L_{XF} := \sum_{x\in\mathcal{X}} |x\rangle\langle x|_X \otimes \mathsf{Comp}_{F[x]}$. As shown in [1], we also have

$$\mathsf{cO}_{XYF} = L_{XF} \cdot \mathsf{StO}_{XYF} \cdot L_{XF}. \tag{5}$$

The following lemma is an immediate consequence of (5).

Lemma 1. *For any x, y and any database D, every database that appears in the superposition $\mathsf{cO}\,|x\rangle_X\,|y\rangle_Y\,|D\rangle_F$ has size at most $|D|+1$. In particular, after an adversary makes q queries to the compressed oracle, the oracle register F is supported on databases of size at most q.*

The following lemma follows by a straightforward calculation from (5).

Lemma 2. *If $D(x)=\bot$, then $\mathsf{cO}\,|x\rangle_X\,|y\rangle_Y\,|D\rangle_F$ is equal to*

$$\sum_{\alpha\in\{0,1\}^n}\frac{1}{2^{n/2}}\,|x\rangle_X\,|y\oplus\alpha\rangle_Y\,|D[x\mapsto\alpha]\rangle_F+\frac{1}{2^{n/2}}\,|x\rangle_X\,|\hat{0}\rangle_Y\,|D\rangle_F \\ -\sum_{\beta\in\{0,1\}^n}\frac{1}{2^n}\,|x\rangle_X\,|\hat{0}\rangle_Y\,|D[x\mapsto\beta]\rangle_F\,.$$

Remark 2. The compressed oracle can be implemented by an efficient quantum algorithm [35]. This fact is crucial for proving some security notions, but we do not rely on it and thus omit the details.

3 Original Indifferentiability in the Classical Setting

This section reviews the original indifferentiability in the classical setting [25] because the security notions of interest to us, i.e., public indifferentiability [13] and sequential indifferentiability [24], are defined as its variants.

In what follows, prim denotes an idealized primitive (e.g., an ideally random compression function for a Merkle-Damgård hash, or a random permutation for a sponge function), and $\mathsf{C}[\mathsf{prim}]$ denotes a construction built from prim (e.g., a Merkle-Damgård or sponge hash). Moreover, R denotes an ideally random object that has the same input/output interface as $\mathsf{C}[\mathsf{prim}]$ (e.g., a VIL random oracle).

Indifferentiability of $\mathsf{C}[\mathsf{prim}]$ from R is defined via indistinguishability between a real world and an ideal world equipped with a simulator. In the real world, a distinguisher $\mathcal{D}$ is given oracle access to $(\mathsf{C}[\mathsf{prim}],\mathsf{prim})$. In the ideal world, $\mathcal{D}$ is given oracle access to $(\mathsf{R},\mathcal{S}[\mathsf{R}])$, where the simulator $\mathcal{S}$ has oracle access to R. Intuitively, by querying R, $\mathcal{S}$ tries to mimic the behavior of prim consistently with R. (See also Fig. 1 in the appendix.) The construction $\mathsf{C}[\mathsf{prim}]$ is said to be indifferentiable from R if there exists an efficient $\mathcal{S}$ such that the distinguishing advantage is negligible for any efficient $\mathcal{D}$. The formal definition is as follows.

Definition 2 (Indifferentiability). *The construction $\mathsf{C}[\mathsf{prim}]$ is* indifferentiable *from R if there exists an efficient simulator $\mathcal{S}$ such that, for any efficient distinguisher $\mathcal{D}$, the indifferentiability advantage*

$$\mathsf{Adv}^{\mathsf{indiff}}_{\mathsf{R},\mathsf{C}[\mathsf{prim}],\mathcal{S}}(\mathcal{D}):=\left|\Pr\left[1\leftarrow\mathcal{D}\big[\mathsf{C}[\mathsf{prim}],\mathsf{prim}\big]\right]-\Pr\left[1\leftarrow\mathcal{D}\big[\mathsf{R},\mathcal{S}[\mathsf{R}]\big]\right]\right| \quad (6)$$

is negligibly small.

We call the oracle of $\mathsf{C}[\mathsf{prim}]$ in the real world (resp., R in the ideal world) the *construction oracle*, and the oracle of prim (resp., the oracle interface of $\mathcal{S}[\mathsf{R}]$ given to $\mathcal{D}$) the *primitive oracle*. Accordingly, we refer to queries to $\mathsf{C}[\mathsf{prim}]$ or R as *construction queries*, and those to prim or $\mathcal{S}[\mathsf{R}]$ as *primitive queries*.

Remark 3 (On the Order of Quantifiers). In this paper, indifferentiability and related notions are defined in their strong form: A construction is considered secure only if there exists a universal simulator $\mathcal{S}$ such that the distinguishing advantage is small for any distinguisher $\mathcal{D}$ (i.e., $\exists\,\mathcal{S}\ \forall\,\mathcal{D}$). Meanwhile, one can also consider a weaker variant that allows the simulator to depend on the distinguisher. That is, a construction is defined as secure if, for every $\mathcal{D}$, there exists a suitable simulator $\mathcal{S}_{\mathcal{D}}$ (and $\mathcal{S}_{\mathcal{D}}$ may change depending on $\mathcal{D}$), corresponding to the quantifier order $\forall\,\mathcal{D}\ \exists\,\mathcal{S}_{\mathcal{D}}$. The original definition by Maurer et al. [25] follows the weaker quantifier order (i.e., the latter version). We adopt the stronger one to remain consistent with the works that introduced public indifferentiability [13] and sequential indifferentiability [24]. Sometimes, whether a universal simulator exists really affects the (im)possibility of achieving a given indifferentiability-related notion [36]. However, all the results of this paper hold regardless of the order of the quantifiers for $\mathcal{S}$ and $\mathcal{D}$ (i.e., regardless of whether $\exists\,\mathcal{S}\ \forall\,\mathcal{D}$ or $\forall\,\mathcal{D}\ \exists\,\mathcal{S}_{\mathcal{D}}$).

Composition Theorem. Indifferentiability means that $\mathsf{C}[\mathsf{prim}]$ behaves as random as R, so that R can be safely replaced with $\mathsf{C}[\mathsf{prim}]$ in various cryptosystems. This property is formalized as the composition theorem [25,31][4]. More specifically, suppose that a cryptosystem $\Pi[\mathsf{R}]$ is secure (in the model where an ideally random R is available to all parties), the security of $\Pi[\mathsf{R}]$ is defined by a single-stage game, and $\mathsf{C}[\mathsf{prim}]$ is indifferentiable from R. Then, $\Pi[\mathsf{C}[\mathsf{prim}]]$ is also secure (in the model where an ideally random prim is available to all parties). We refer readers to [31] for the formal statement of the composition theorem.

Concrete Examples of (In)differentiable Schemes. A notable example of indifferentiable constructions is the sponge hash [5], where prim is a public invertible random permutation and R is a (variable input/output length) random oracle. If the security of a cryptosystem $\Pi[\mathsf{R}]$ is defined with a single-stage game and proven in the ROM (where R is a VIL random oracle), then it remains secure even if R is replaced with the sponge hash (in the model that the underlying public permutation is ideally random).

Meanwhile, the (strengthened) Merkle-Damgård construction is not indifferentiable from a VIL random oracle because of the length extension attack [10]. In fact, if some modifications are allowed (e.g., by using a prefix-free padding instead of Eq. (2)) or input lengths are fixed [10], Merkle-Damgård can also

[4] At first, the theorem was believed to be true for any cryptosystem [25], but later found true only if G is single-stage [31].

be made indifferentiable. However, various standards, especially SHA-256, SHA-512, and RIPEMD-160, are typically used without such modifications, and inputs to hash functions are variable-length in many applications including some signatures.

4 (Post-Quantum) Public Indifferentiability

This section reviews public indifferentiability in the classical setting [13], and then formalizes its post-quantum version.

4.1 Public Indifferentiability in the Classical Setting

Public indifferentiability [13] is defined in the same way as the original indifferentiability (Definition 2) except that the simulator is additionally allowed to see the entire set of query-response pairs for the ideal construction oracle R. More specifically, the simulator can access not only R but also a "reveal" oracle $\mathsf{rvl}_{\mathsf{R}}$, which takes no input and returns the set of all queries made so far to R together with their responses. (Only the simulator can access $\mathsf{rvl}_{\mathsf{R}}$; the adversary cannot access to it. See also Fig. 1 in the appendix.)

Dodis et al. proved that Merkle-Damgård is publicly indifferentiable from a VIL random oracle [13] when prim is an ideally random compression function[5].

Composition Theorem. A composition theorem also holds for public indifferentiability [13]. Suppose that the security of a cryptosystem $\Pi[\mathsf{R}]$ is defined by a single-stage game G as before, and assume additionally that the set of query-response pairs for R is visible to adversaries (i.e., the reveal oracle $\mathsf{rvl}_{\mathsf{R}}$ is available). If $\Pi[\mathsf{R}]$ is secure (in the model that the oracles of an ideally random R and $\mathsf{rvl}_{\mathsf{R}}$ are available to all the parties) and $\mathsf{C}[\mathsf{prim}]$ is public indifferentiable from R, then $\Pi[\mathsf{C}[\mathsf{prim}]]$ is also secure (in the model that the oracle of an ideally random prim is available to all the parties). See [13] for more details on the composition theorem for the classical public indifferentiability.

ROM-based security proofs for various cryptosystems remain valid when inputs to the random oracle are disclosed to adversaries. For example, in Fiat-Shamir signatures and FDH signatures, inputs to a VIL random oracle are messages to sign and some public random values. Therefore, the composition theorem guarantees that those cryptosystems remain secure even if a VIL random oracle is replaced with a Merkle-Damgård hash.

4.2 Formalization of Post-Quantum Public Indifferentiability

A fully quantum analogue of (original) indifferentiability is conceptually simple: one lets all parties be quantum and allows superposition queries to all oracles. Such notions have been studied extensively in [9,36,8,37,1].

[5] In fact, they showed a stronger result that the public indifferentiability holds when the compression function is public-use guarded random oracle.

In contrast, defining a quantum analogue of public indifferentiability is not quite straightforward. Classically, the reveal oracle exposes the set of queries made so far to R together with their responses. However, recording quantum queries is, in general, highly non-trivial due to the no-cloning theorem. While Zhandry's compressed-oracle technique (or related tools) may help in specific settings, formalizing a fully quantum reveal interface would require sophisticated additional machinery.

To keep the discussion focused on the essential issues and avoid unnecessary technical complications, we adopt a hybrid post-quantum setting tailored to typical QROM proofs for *classical* cryptosystems, where honest parties remain classical. Our goal is a notion such that a composition theorem holds for a cryptosystem $\Pi[\mathsf{R}]$ in the following situation:

(i) $\Pi[\mathsf{R}]$ is a classical cryptosystem, i.e., it is executable on classical computers and classical communication channel,
(ii) $\Pi[\mathsf{R}]$ is post-quantum secure when R is ideally random and quantum adversaries may query R in superposition, and
(iii) all inputs to R that are generated within $\Pi[\mathsf{R}]$ can be public.

When R is a VIL random oracle, (i)-(iii) are indeed satisfied by many QROM security proofs, including those for Fiat-Shamir and FDH signatures in various previous works [6,34,32,20,15,21,14,33,17,22]. We also want that, after R is replaced with $\mathsf{C}[\mathsf{prim}]$, the resulting system $\Pi[\mathsf{C}[\mathsf{prim}]]$ remains secure even if adversaries are allowed superposition queries to prim.

Trying to formalize a post-quantum analogue of public indifferentiability like above, we naturally arrive at the following observations[6].

- Distinguishers must be allowed superposition queries to the primitive oracle (prim in the real world, or a simulator in the ideal world). To enable the simulator to effectively simulate the quantum oracle for prim, the simulator should also be allowed to make superposition queries to R.
- Distinguishers' queries to the construction oracle (R or $\mathsf{C}[\mathsf{prim}]$) can be kept classical: In composition theorems, such construction queries correspond to oracle calls to R made by the security game (e.g., queries to a random oracle made in a signing oracle), which are classical under our assumptions.

The next issue to consider is how to define the reveal oracle in the post-quantum setting: In the classical definition, the reveal oracle returns the full set of construction queries together with the corresponding responses. In our setting, since the simulator may perform quantum queries, how to efficiently record its own queries is unclear. Meanwhile, the distinguisher's construction queries are classical and can be recorded straightforwardly. We therefore let the reveal oracle expose only the set of the distinguisher's construction queries and the corresponding responses. (Indeed, even classically, revealing to the simulator the set of its own queries is redundant.) In the end, we obtain the following definition.

[6] This hybrid access pattern is not new. See, e.g., the analogous observation for (post-)quantum indifferentiability in [9].

Definition 3 (Post-Quantum Public indifferentiability). *The construction* $\mathsf{C}[\mathsf{prim}]$ *is* post-quantum public indifferentiable *from* R *if there exists an efficient simulator* $\mathcal{S}$ *such that, for any efficient distinguisher* $\mathcal{D}$*, the post-quantum public indifferentiability advantage*

$$\mathsf{Adv}^{\mathsf{pqpub\text{-}indiff}}_{\mathsf{R},\mathsf{C}[\mathsf{prim}],\mathcal{S}}(\mathcal{D}) := \left|\Pr\left[1 \leftarrow \mathcal{D}\big[\mathsf{C}[\mathsf{prim}],\mathsf{prim}\big]\right] - \Pr\left[1 \leftarrow \mathcal{D}\big[\mathsf{R},\mathcal{S}[\mathsf{R},\mathsf{rvl}_{\mathsf{R}}]\big]\right]\right|$$

is negligibly small. Here, both $\mathcal{S}$ *and* $\mathcal{D}$ *are quantum algorithms.* $\mathcal{S}$ *may make quantum superposition queries to* R*. The distinguisher's construction queries are classical, while its primitive queries may be in quantum superposition (in particular,* $\mathcal{S}$ *has to simulate the quantum oracle of* prim*). The oracle* $\mathsf{rvl}_{\mathsf{R}}$ *takes no input and returns the set of (classical) construction queries to* R *made so far by* $\mathcal{D}$ *together with the corresponding responses.*

Composition Theorem. Definition 3 indeed admits a composition theorem analogous to the classical one. Let $\Pi[\mathsf{R}]$ be a classical cryptosystem with classical oracle access to R. Assume its security is defined by a single-stage game G with a quantum adversary. The adversary may query R in superposition, whereas all other calls to R made within G (e.g., queries to a random oracle inside a signing oracle) are classical. The game provides an interface $\mathsf{rvl}_{\mathsf{R}}^{(G)}$ that returns the set of all such classical queries to R (i.e., the queries other than those made directly by the adversary), together with their responses. At the end, G outputs a bit b. Then, the following theorem holds.

Theorem 1 (Composition theorem for post-quantum public indifferentiability). *For any quantum adversary* $\mathcal{A}$ *interacting with* G *and the quantum oracle of* prim*, and any quantum simulator* $\mathcal{S}$ *for the post-quantum public indifferentiability of* $\mathsf{C}[\mathsf{prim}]$ *from* R*, there exist a quantum distinguisher* $\mathcal{D}$ *(for post-quantum public indifferentiability) and a quantum adversary* $\mathcal{B}$ *interacting with* G*, the quantum oracle of* R*, and the reveal oracle* $\mathsf{rvl}_{\mathsf{R}}^{(G)}$ *such that*

$$\Pr\left[1 \leftarrow G\big[\mathsf{C}[\mathsf{prim}],\mathcal{A}[\mathsf{prim}]\big]\right] \leq \Pr\left[1 \leftarrow G\big[\mathsf{R},\mathcal{B}[\mathsf{R},\mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]\right] + \mathsf{Adv}^{\mathsf{pqpub\text{-}indiff}}_{\mathsf{R},\mathsf{C}[\mathsf{prim}],\mathcal{S}}(\mathcal{D})\,,$$

where $\mathcal{D}$ *is efficient if* $\mathcal{A}$ *is (resp.,* $\mathcal{B}$ *is efficient if* $\mathcal{A}$ *and* $\mathcal{S}$ *are), and* $\mathcal{D}$ *does not depend on* $\mathcal{S}$*.*

The proof is analogous to [31]. See Section A and Fig. 2 for more details.

Remarks on Definition 3 and Theorem 1. Our reveal oracle returns only the distinguisher's classical construction queries to R (and responses), and ignores the simulator's own queries. If all algorithms and queries are classical, this coincides with the "weak public indifferentiability" of [18][7]. For stateless and deterministic constructions, it coincides with the usual public indifferentiability[8].

[7] Here "weak" refers to the reveal interface (excluding simulator queries), not to the quantifier order discussed in Remark 3.

[8] As discussed in [18], public indifferentiability implies weak public indifferentiability, and weak public indifferentiability trivially implies sequential indifferentiability ex-

In Section 6, we show that Merkle-Damgård does not satisfy Definition 3, and thereby we claim that it does not satisfy public indifferentiability in the post-quantum setting. One might wonder whether a different (possibly fully quantum) notion could coincide with the classical public indifferentiability while still being achievable by Merkle-Damgård. However, such notions seemingly have to be at least as strong as Definition 3 (or, Definition 4 explained later, which is weaker than Definition 3), and our result rules out this possibility as well.

Finally, although the reveal interface in Theorem 1 differs slightly from the composition theorem for classical public indifferentiability [13] (because $\mathsf{rvl}_{\mathsf{R}}^{(G)}$ in our Theorem 1 does not return the queries made directly by an adversary while the reveal oracle in [13] does), it still captures the intended setting where all inputs to R generated within $\Pi[\mathsf{R}]$ are public and visible to adversaries. Accordingly, the theorem is applicable to QROM proofs that remain valid even if all classical inputs to R are disclosed, such as those for Fiat-Shamir and FDH signatures in [6,34,32,20,15,21,14,33,17,22].

5 (Post-Quantum) Sequential Indifferentiability

In this section, we will review sequential indifferentiability in the classical setting [24], and then formalize its post-quantum version.

5.1 Sequential Indifferentiability in the Classical Setting

In the classical setting, sequential indifferentiability [24] is defined in the same way as the usual indifferentiability, except that the distinguisher $\mathcal{D}$ is restricted to make construction queries only after it finishes all primitive queries. Equivalently, once $\mathcal{D}$ makes a construction query, it may no longer make primitive queries. See also Fig. 1 in the appendix.

The following sequence of implications immediately follows from the definitions:

$$\text{Indifferentiability} \Rightarrow \text{Public indifferentiability} \Rightarrow \text{Sequential indifferentiability}.$$

Moreover, Mandal et al. [24] showed that sequential indifferentiability implies public indifferentiability (and thus the two notions are equivalent) if the construction oracles $\mathsf{C}[\mathsf{prim}]$ and R are stateless and deterministic[9].

5.2 Formalization of Post-Quantum Sequential Indifferentiability

By restricting the distinguisher's query order in post-quantum public indifferentiability, we obtain a post-quantum analogue of sequential indifferentiability. Specifically, we consider the following setup:

plained later. Moreover, sequential indifferentiability implies public indifferentiability for stateless and deterministic constructions [24]. Thus, the three notions coincide.

[9] The equivalence does not hold unconditionally. They also show an example of a stateful construction that is sequential indifferentiable but not public indifferentiable.

– As in the classical notion, the distinguisher queries the construction oracle only after finishing all primitive queries[10],
– primitive queries may be in quantum superposition, whereas construction queries remain classical, and
– the simulator may make quantum superposition queries to the ideally random construction oracle.

Definition 4 (Post-Quantum Sequential indifferentiability). *The construction* $\mathsf{C}[\mathsf{prim}]$ *is* post-quantum sequentially indifferentiable *from* R *if there exists an efficient simulator* $\mathcal{S}$ *such that, for any efficient distinguisher* $\mathcal{D}$ *that makes construction queries only after finishing primitive queries, the advantage*

$$\mathsf{Adv}^{\mathsf{pqseq\text{-}indiff}}_{\mathsf{R},\mathsf{C}[\mathsf{prim}],\mathcal{S}}(\mathcal{D}) := \left|\Pr\left[1 \leftarrow \mathcal{D}\big[\mathsf{C}[\mathsf{prim}],\mathsf{prim}\big]\right] - \Pr\left[1 \leftarrow \mathcal{D}\big[\mathsf{R},\mathcal{S}[\mathsf{R}]\big]\right]\right|$$

is negligible. Here, both $\mathcal{S}$ *and* $\mathcal{D}$ *are quantum algorithms.* $\mathcal{S}$ *makes quantum superposition queries to* R*. The distinguisher's construction queries are classical, while its primitive queries may be in quantum superposition (in particular,* $\mathcal{S}$ *has to simulate the quantum oracle of* prim*).*

As in the classical case, post-quantum public indifferentiability implies post-quantum sequential indifferentiability. For stateless and deterministic constructions, the converse might also hold as in [24]. Still, proving it in our post-quantum setting would require substantial additional analysis and is beyond our scope.

6 Attack Breaking the Post-Quantum Sequential/Public Indifferentiability of Merkle-Damgård

This section shows an efficient distinguisher $\mathcal{D}$ that breaks the post-quantum sequential indifferentiability of the Merkle-Damgård construction $\mathsf{MD}[h]$. In what follows, the ideal construction oracle R is a VIL random oracle (denoted by RO) with range $\{0,1\}^n$, and the primitive oracle prim is the FIL random oracle $h : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^n$. The distinguisher $\mathcal{D}$ also automatically breaks the post-quantum public indifferentiability of Merkle-Damgård[11].

Assumptions. We assume that the domain of RO and $\mathsf{MD}[h]$ is $\{0,1\}^{\leq \ell}$ for a sufficiently large ℓ (in particular, $\ell \geq 3m/2$). To keep the presentation simple while retaining generality, we assume that the padding function pad (see Eq. (2)) satisfies the following condition:

– Messages of length $m/2$ (resp., $3m/2$) pad to exactly one block (resp., two blocks). Namely, $|\mathsf{pad}(x)| = m$ if $|x| = m/2$, and $|\mathsf{pad}(x)| = 2m$ if $|x| = 3m/2$. Equivalently, $|\mathsf{enc}(|x|)| = m/2$ for $|x| \in \{m/2,\, 3m/2\}$.

[10] Under this restriction, the reveal oracle $\mathsf{rvl}_{\mathsf{R}}$ of Definition 3 becomes meaningless because it always returns the empty set.

[11] $\mathcal{D}$ makes construction queries only after finishing primitive queries. For such a distinguisher, the reveal oracle is meaningless for simulators because it always returns the empty set.

This setting covers many standard paddings used in practice, including those of SHA-2 family (especially, SHA-256 and SHA-512) [28] and RIPEMD-160 [12]. (See also Section B for further remarks on the padding assumption.)

6.1 Distinguisher to Break the Security of Merkle-Damgård

Below, we define three distinguishers $\mathcal{D}_0$, $\mathcal{D}_1$, and $\mathcal{D}$ against the post-quantum sequential indifferentiability of the (strengthened) Merkle-Damgård construction $\mathsf{MD}[h]$. Before describing the details, we provide some remarks on notations.

Notational Remarks. The distinguishers use six quantum registers: X, Y, Z, W, T_1, and T_2. Their roles are as follows.

X**:** An $(m+n)$-qubit register to send a query to the primitive oracle.
Y**:** An n-qubit register to receive the response from the primitive oracle.
Z**:** An $m/2$-qubit local-memory register.
W**:** An n-qubit local-memory register.
T_1, T_2**:** Single-qubit auxiliary flag registers.

In the real world, the quantum oracle for h acts on register X, Y as $O_h \left|x\right\rangle_X \left|y\right\rangle_Y = \left|x\right\rangle_X \left|y \oplus h(x)\right\rangle_Y$. In the ideal world, whenever a distinguisher makes a primitive query, it hands (X, Y) to the simulator $\mathcal{S}$, which applies an arbitrary quantum computation on (X, Y) and its local memory (including quantum queries to RO), and then returns (X, Y) to the distinguisher.

The registers X, Y, Z, and W are prepared at the beginning of the distinguishers, whereas T_1 and T_2 are introduced only when required during execution. Each of T_1 and T_2 is initialized to $|0\rangle$ when introduced, and flipped if a consistency check fails (e.g., if $\mathcal{S}$ unreasonably modifies the query register X or answers inconsistently with a previous response).

Shorthand for Variables and Reversible Operations. By abuse of notation, we sometimes use a register symbol to denote the classical bit string stored in that register (possibly in superposition). We also use the following shorthand for standard reversible operations:

- "Add Δ to a register A" means applying the unitary $\left|x\right\rangle_A \mapsto \left|x \oplus \Delta\right\rangle_A$ (and acting as identity on all other registers), where Δ may depend on the contents of other registers. For instance, "Add $\mathsf{pad}(Z)||IV$ to X" denotes applying the unitary $\left|x\right\rangle_X \left|z\right\rangle_Z \mapsto \left|x \oplus (\mathsf{pad}(z)||IV)\right\rangle_X \left|z\right\rangle_Z$.
- "Flip T_1 if P" means applying a controlled-NOT on T_1 conditioned on a predicate P evaluated (coherently) on the computational-basis contents of the relevant registers, namely, $\left|u\right\rangle \left|b\right\rangle_{T_1} \mapsto \left|u\right\rangle \left|b \oplus P(u)\right\rangle_{T_1}$, where u ranges over computational-basis states of the relevant registers, and $P(u) \in \{0, 1\}$. For example, "Flip T_1 if $X \neq \mathsf{pad}(Z)||IV$" corresponds to the predicate $P(x, z) := [x \neq \mathsf{pad}(z)||IV]$. The same convention applies to T_2.

We define the distinguishers $\mathcal{D}_0$, $\mathcal{D}_1$, and $\mathcal{D}$ below. Register W is not used in $\mathcal{D}_0$, but we write it in the description for the consistency with $\mathcal{D}_1$. $\mathcal{D}_1$ is identical to $\mathcal{D}_0$ up to Step 4, except that $\mathcal{D}_1$ additionally adds Y to W in Step 1. (The intuition behind the distinguishers is given later on p. 18.)

Distinguisher $\mathcal{D}_0$.

1. Prepare the quantum state $\frac{1}{2^{m/4}} \sum_{a \in \{0,1\}^{m/2}} |\mathsf{pad}(a)||IV\rangle_X |0^n\rangle_Y |a\rangle_Z |0^n\rangle_W$ (namely, put Z in the uniform superposition over $\{0,1\}^{m/2}$ and set $X = \mathsf{pad}(Z)||IV$). Then, make a query to the primitive oracle (using register X, Y).
2. Introduce an auxiliary 1-qubit register T_1, and set it as $|0\rangle_{T_1}$. Flip T_1 if $X \neq \mathsf{pad}(Z)||IV$.
3. Make a query to the primitive oracle again (using register X, Y).
4. Add $\mathsf{pad}(Z)||IV$ to X. Then, introduce another auxiliary 1-qubit register T_2, and set it as $|0\rangle_{T_2}$. Flip T_2 if $Y \neq 0^n$ or $X \neq 0^{m+n}$.
5. Apply the Hadamard operator $H^{\otimes m/2}$ to register Z, and then measure all the registers in the computational basis. If the outcome is

$$|0^{m+n}\rangle_X |0^n\rangle_Y |0^{m/2}\rangle_Z |0^n\rangle_W |0\rangle_{T_1} |0\rangle_{T_2},$$

   then return 0. Otherwise, return 1.

Distinguisher $\mathcal{D}_1$.

1. Prepare the quantum state $\frac{1}{2^{m/4}} \sum_{a \in \{0,1\}^{m/2}} |\mathsf{pad}(a)||IV\rangle_X |0^n\rangle_Y |a\rangle_Z |0^n\rangle_W$ (namely, put Z in the uniform superposition over $\{0,1\}^{m/2}$ and set $X = \mathsf{pad}(Z)||IV$). Then, make a query to the primitive oracle (using register X, Y). Finally, add Y to W.
2. Introduce an auxiliary 1-qubit register T_1, and set it as $|0\rangle_{T_1}$. Flip T_1 if $X \neq \mathsf{pad}(Z)||IV$.
3. Make a query to the primitive oracle again (using register X, Y).
4. Add $\mathsf{pad}(Z)||IV$ to X. Then, introduce another auxiliary 1-qubit register T_2, and set it as $|0\rangle_{T_2}$. Flip T_2 if $Y \neq 0^n$ or $X \neq 0^{m+n}$.
5. Add $\left(0^{m/2}||\mathsf{enc}(\frac{3m}{2})||W\right)$ to X.[12]
6. Make a query to the primitive oracle (using register X, Y).
7. (a) Measure all the registers in the computational basis, and let the outcome be $|x\rangle_X |d\rangle_Y |a\rangle_Z |c\rangle_W |\mathsf{flag}_1\rangle_{T_1} |\mathsf{flag}_2\rangle_{T_2}$, where $x, d, a, c, \mathsf{flag}_1, \mathsf{flag}_2$ are some classical bit strings[13].
   (b) Query the (classical values) a and $\mathsf{pad}(a)||0^{m/2}$ to the construction oracle (denoted by $\mathcal{O}_\mathsf{c}$, which is either $\mathsf{MD}[h]$ or RO), and let $\gamma = \mathcal{O}_\mathsf{c}(a)$ and $\delta = \mathcal{O}_\mathsf{c}(\mathsf{pad}(a)||0^{m/2})$ be the responses.
   (c) If $\mathsf{flag}_1 = \mathsf{flag}_2 = 0$ and $c = \gamma$ and $d = \delta$, then return 0. Otherwise, return 1.

[12] Note that $|\mathsf{enc}(\frac{3m}{2})| = m/2$ by the assumption on padding.
[13] Note that register Z has not been modified since Step 1.

Distinguisher $\mathcal{D}$.

1. Choose a bit r uniformly at random.
2. Run $\mathcal{D}_r$, and return what $\mathcal{D}_r$ outputs as $\mathcal{D}$'s own output.

Our main result is that the above distinguisher $\mathcal{D}$ breaks the post-quantum sequential indifferentiability of $\mathsf{MD}[h]$. Formally, the following theorem holds.

Theorem 2. *For any simulator $\mathcal{S}$ making at most q queries, where q is in $o(2^{m/4})$, the advantage $\mathsf{Adv}^{\mathsf{pqseq\text{-}indiff}}_{\mathsf{RO},\mathsf{MD}[h],\mathcal{S}}(\mathcal{D})$ is in $\Omega(1)$. More precisely, it holds that $\mathsf{Adv}^{\mathsf{pqseq\text{-}indiff}}_{\mathsf{RO},\mathsf{MD}[h],\mathcal{S}}(\mathcal{D}) \geq 2^{-23} - O\left(q/2^{m/4} + 1/2^{n/2}\right)$.*

A formal proof of the theorem will be given in Section 7. Here, we explain the intuition behind the theorem and the distinguishers $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}$.

Intuition behind the Theorem and the Distinguishers. Both $\mathcal{D}_0$ and $\mathcal{D}_1$ are constructed to always output 0 in the real world. Hence $\mathcal{D}$ also outputs 0 there. Thus the distinguishing advantage comes solely from their behavior in the ideal world.

The flag registers T_1 and T_2 are used only to detect apparent inconsistencies in the ideal world (e.g., $\mathcal{S}$ unreasonably modifying X or answering inconsistently). Whenever either flag is set, the distinguisher outputs 1. For the intuition below, we condition on the event $T_1 = T_2 = 0$ (the cases $T_1 = 1$ or $T_2 = 1$ can only increase the probability that the distinguishers output 1).

Intuitively, $\mathcal{D}_0$ targets simulators that try to learn the information of (i.e., make non-negligible (quantum) correlation with) the superposition variable a prepared in Step 1. If $\mathcal{S}$ stores information correlated with a, then its private workspace becomes entangled with Z. In that case, Step 5 detects the disturbance: The probability of measuring $0^{m/2}$ in register Z after applying $H^{\otimes m/2}$ drops noticeably, and $\mathcal{D}_0$ returns 1.

Conversely, $\mathcal{D}_1$ targets simulators that do *not* record information about a. In the ideal world, the response to the primitive query in Step 6 (which will be d in Step 7(a)) corresponds to the value that should match $\mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$. In Steps 7(b) and 7(c), $\mathcal{D}_1$ recomputes $\delta = \mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$ and checks whether $d = \delta$. To pass the check in Step 7(c), $\mathcal{S}$ has to return the correct value $\mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$ in Step 6 (as d). For this, $\mathcal{S}$ must have already learned (i.e., made quantum correlation with) a while answering the earlier primitive query in Step 1 (or Step 3). In other words, if $\mathcal{S}$ did not learn the value a, then d will be different from $\mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$, and thus $\mathcal{D}_1$ will output 1.

In summary, $\mathcal{D}_0$ catches the case where $\mathcal{S}$ records information about a, whereas $\mathcal{D}_1$ catches the case where it does not. Moreover, $\mathcal{D}_0$ and $\mathcal{D}_1$ are identical up to Step 4 (except for adding Y to W in Step 1 of $\mathcal{D}_1$), so $\mathcal{S}$ must decide whether to learn a before it can tell which distinguisher it is facing. Therefore, when $\mathcal{D}$ runs $\mathcal{D}_0$ or $\mathcal{D}_1$ with probability 1/2, it outputs 1 with non-negligible probability in the ideal world.

Comparison with Prefix-Free Merkle-Damgård. In contrast to our result, the prefix-free Merkle-Damgård is known to achieve full indifferentiability, even in the quantum setting [10,35,1][14]. The intuition for the difference is as follows.

For the prefix-free variant, along a compression-function chain corresponding to one complete input to Merkle-Damgård, at most one compression-function call—namely, the terminal call of a valid prefix-free chain—has to be compatible with RO. The remaining compression-function outputs are internal to the simulation: they can be simulated locally as values of a random function, and in the quantum proof this local simulation can be tracked using the compressed-oracle technique, so that the simulator retains the necessary consistency information. Moreover, $\mathcal{S}$ does not have to remember the query to RO; for every primitive query, $\mathcal{S}$ can coherently check whether a distinguisher is trying to compute the last block of some valid input to Merkle-Damgård by reconstructing the chain from the current simulated compressed database, and $\mathcal{S}$ can make a query again to RO whenever necessary. Hence, $\mathcal{S}$ can simulate the primitive oracle against distinguishers like $\mathcal{D}_1$, while not creating quantum correlation with distinguishers like $\mathcal{D}_0$.

For the non-prefix-free Merkle-Damgård construction considered here, a distinguisher can instead arrange a chain of compression-function queries in which every output of the compression function in the chain must be consistent with an output of RO. When $\mathcal{S}$ simulates the primitive oracle for $\mathcal{D}_1$, the first compression output must equal $\mathsf{RO}(a)$ for a short input a, and the same value is then reused as the chaining value for a longer input $\mathsf{pad}(a)||0^{m/2}$, whose hash must equal $\mathsf{RO}(\mathsf{pad}(a)||0^{m/2})$. To answer the latter query consistently, the simulator must know which value of a gave rise to the chaining value. Creating such information produces quantum correlation with the distinguisher's register and is detected by $\mathcal{D}_0$; avoiding such correlation makes the consistency check of $\mathcal{D}_1$ fail with high probability.

7 Proof of Theorem 2

This section provides a proof of Theorem 2. First, it is easy to check that both $\mathcal{D}_0$ and $\mathcal{D}_1$ always output 0 in the real world. Therefore, $\Pr_h\left[1 \leftarrow \mathcal{D}\left[\mathsf{MD}[h], h\right]\right] = \sum_{r\in\{0,1\}} \frac{1}{2} \Pr_h\left[1 \leftarrow \mathcal{D}_r\left[\mathsf{MD}[h], h\right]\right] = 0$ holds. Let $p := \Pr_{\mathsf{RO}}\left[1 \leftarrow \mathcal{D}\left[\mathsf{RO}, \mathcal{S}[\mathsf{RO}]\right]\right]$ and $p_r := \Pr_{\mathsf{RO}}\left[1 \leftarrow \mathcal{D}_r\left[\mathsf{RO}, \mathcal{S}[\mathsf{RO}]\right]\right]$ for $r \in \{0,1\}$. Then, it follows that

$$\begin{aligned}\mathsf{Adv}^{\mathsf{pqseq\text{-}indiff}}_{\mathsf{RO},\mathsf{MD}[h],\mathcal{S}}(\mathcal{D}) &= \left|\Pr_{\mathsf{RO}}\left[1 \leftarrow \mathcal{D}\left[\mathsf{RO}, \mathcal{S}[\mathsf{RO}]\right]\right] - \Pr_h\left[1 \leftarrow \mathcal{D}\left[\mathsf{MD}[h], h\right]\right]\right| \\ &= p = \frac{p_0 + p_1}{2}.\end{aligned}$$

Below, we will show that p is non-negligible for any efficient simulator $\mathcal{S}$. Without loss of generality, we assume that $\mathcal{S}$ does not perform intermediate measurements.

[14] To be more precise, [35] showed a quantum proof, and [1] pointed out a gap and how it could be fixed.

When $\mathcal{D}_1$ runs relative to RO and $\mathcal{S}[\mathsf{RO}]$ in the ideal world, define

$$p_1^{T_1} := \Pr_{\mathsf{RO}}\left[\mathsf{flag}_1 = 1 \text{ in Step 7 of } \mathcal{D}_1\right], \quad p_1^{T_2} := \Pr_{\mathsf{RO}}\left[\mathsf{flag}_2 = 1 \text{ in Step 7 of } \mathcal{D}_1\right],$$
$$p_1^{c\neq\gamma} := \Pr_{\mathsf{RO}}\left[c \neq \gamma \text{ in Step 7 of } \mathcal{D}_1\right].$$

Our technical core is the proposition below, which roughly says that p_1 becomes non-negligible if p_0, $p_1^{T_1}$, $p_1^{T_2}$, and $p_1^{c\neq\gamma}$ are negligible.

Proposition 1. *Suppose that $p_0, p_1^{T_1}, p_1^{T_2}$, and $p_1^{c\neq\gamma}$ are upper bounded by a positive value $\varepsilon \leq 2^{-20}$. If $\mathcal{S}$ makes at most q queries in total and $q \leq 2^{m/2-4}$, then $p_1 \geq 1 - \varepsilon_* - O\left(q/2^{m/4} + 1/2^{n/2}\right)$ holds, where $\varepsilon_* := 8\sqrt[4]{\varepsilon} + 64\sqrt[4]{\varepsilon^3} + 22\sqrt{\varepsilon} + 64\varepsilon$.*

Before proving Proposition 1, we show that Theorem 2 indeed follows from the proposition.

Proof (of Theorem 2). Let ε be the maximum of $p_0, p_1^{T_1}, p_1^{T_2}$, and $p_1^{c\neq\gamma}$. Since $\mathcal{D}_1$ outputs 1 whenever $\mathsf{flag}_1 = 1$, $\mathsf{flag}_2 = 1$, or $c \neq \gamma$ holds in Step 7, we have $p_1 \geq \max\{p_1^{T_1},\ p_1^{T_2},\ p_1^{c\neq\gamma}\}$. Hence, $\mathsf{Adv}^{\mathsf{pqseq\text{-}indiff}}_{\mathsf{RO},\mathsf{MD}[h],\mathcal{S}}(\mathcal{D}) = \frac{p_0+p_1}{2} = \frac{p_0+(p_1+p_1+p_1+p_1)/4}{2} \geq \frac{p_0/4+(p_1+p_1^{T_1}+p_1^{T_2}+p_1^{c\neq\gamma})/4}{2} \geq \frac{1}{8}\max\left\{\varepsilon, p_1\right\}$.

If $\varepsilon \geq 2^{-20}$, then the statement of the theorem immediately follows from the above inequality. If $\varepsilon \leq 2^{-20}$, then ε_* of Proposition 1 is upper bounded by $1/2$. Thus, $p_1 \geq 1/2 - O\left(q/2^{m/4} + 1/2^{n/2}\right)$ holds. Therefore, the claim of the theorem also holds when $\varepsilon \leq 2^{-20}$. □

Next, we prove Proposition 1.

7.1 Proof of Proposition 1

Recall that the domain of RO and $\mathsf{MD}[h]$ is assumed to be $\{0,1\}^{\leq\ell}$ for a sufficiently large ℓ. Let Funcs be the set of all functions from $\{0,1\}^{\leq\ell}$ to $\{0,1\}^n$, and let DBs be the set of all databases for the compressed oracle of RO. We write $N := |\mathsf{Funcs}|$ (so $N = 2^{n(2^{\ell+1}-1)}$). The quantum register for $\mathcal{S}$'s offline computation and local memory is denoted by S.

Definition of Experiments. We prove the proposition by using four experiments $\mathsf{E}_0, \ldots, \mathsf{E}_3$, which are defined as follows.

- E_0 **and** E_1**.** E_0 (resp., E_1) corresponds to the ideal world for post-quantum sequential indifferentiability of $\mathsf{MD}[h]$ from RO when the distinguisher is $\mathcal{D}_0$ (resp., $\mathcal{D}_1$) and the simulator is $\mathcal{S}$. That is, $\mathcal{D}_0$ (resp., $\mathcal{D}_1$) runs relative to $(\mathsf{RO}, \mathcal{S}[\mathsf{RO}])$, and E_0 (resp., E_1) outputs the final output of $\mathcal{D}_0$ (resp., $\mathcal{D}_1$).
- E_2 **and** E_3**.** E_2 (resp., E_3) runs a distinguisher $\mathcal{A}$ (resp., $\mathcal{B}$) defined below, relative to $(\mathsf{RO}, \mathcal{S}[\mathsf{RO}])$ and outputs its final output. Unlike E_0 and E_1, in E_2 and E_3 the distinguisher is allowed to make a direct *quantum* query to the construction oracle in the middle of its execution. Hence, $\mathcal{A}$ and $\mathcal{B}$ are no longer distinguishers for post-quantum sequential indifferentiability.

Distinguisher $\mathcal{A}$.

- Perform Steps 1–4 of $\mathcal{D}_0$.
- Query the construction oracle on register Z in quantum superposition so that the answer is written into register W. We refer to this as Step 4.5.
- Perform Steps 5–7 of $\mathcal{D}_1$.

Distinguisher $\mathcal{B}$.

- Steps 1–6 are the same as for $\mathcal{A}$.
- As Step 7, do the following. Copy[15] $\mathsf{pad}(Z)||0^{m/2}$ into an auxiliary register $\mathsf{aux1}$, and query the construction oracle on $\mathsf{aux1}$ (in quantum superposition) so that the result is written into another auxiliary register $\mathsf{aux2}$. Measure registers Y and $\mathsf{aux2}$ in the computational basis and let d and δ be the outcomes. Output 1 if $d \neq \delta$, and output 0 otherwise.

For each $r = 0, \dots, 3$, we consider two variants of E_r depending on how the quantum oracle for RO is implemented: The standard oracle or the compressed oracle. We denote them by $\mathsf{E}_r^{\mathsf{StO}}$ and $\mathsf{E}_r^{\mathsf{cO}}$, respectively. We also define E_r^f to be the experiment in which the random oracle is replaced by a *fixed* VIL function $f \in \mathsf{Funcs}$. For $r = 0$ and $r = 1$, we have $p_r := \Pr[1 \leftarrow \mathsf{E}_r^{\mathsf{StO}}] = \Pr[1 \leftarrow \mathsf{E}_r^{\mathsf{cO}}] = \mathbf{E}_f\big[\Pr[1 \leftarrow \mathsf{E}_r^f]\big]$, where $\mathbf{E}_f$ denotes the expected value taken over the random choice of f.

For each $r = 0, \dots, 3$, each $i = 1, 2, \dots$, and a fixed $f \in \mathsf{Funcs}$, let $|\mathsf{E}_r^f; i\rangle$ be the joint quantum state of the distinguisher ($\mathcal{D}_0$, $\mathcal{D}_1$, $\mathcal{A}$, or $\mathcal{B}$) and the simulator $\mathcal{S}$ just after Step i of the distinguisher in E_r^f. In addition, let[16] $|\mathsf{E}_r^{\mathsf{StO}}; i\rangle$ (resp., $|\mathsf{E}_r^{\mathsf{cO}}; i\rangle$) be the whole quantum state of the distinguisher, the simulator, and the oracle just after Step i in $\mathsf{E}_r^{\mathsf{StO}}$ (resp., $\mathsf{E}_r^{\mathsf{cO}}$). Then,

$$|\mathsf{E}_r^{\mathsf{StO}}; i\rangle = \frac{1}{\sqrt{N}} \sum_{f \in \mathsf{Funcs}} |\mathsf{E}_r^f; i\rangle \otimes |f\rangle_F\,, \ |\mathsf{E}_r^{\mathsf{cO}}; i\rangle = \mathsf{Comp}_F\, |\mathsf{E}_r^{\mathsf{StO}}; i\rangle\,. \tag{7}$$

Outline of the Proof. Here, we briefly outline the proof.

Our goal is to lower bound p_1, i.e., the probability that E_1 returns 1. To this end, we analyze the quantum states of the experiments step by step (just after Step i for $i = 1, 2, \dots$). To exploit the assumption that p_0 is upper bounded, we also analyze E_0.

We introduce E_2 to bridge the analyses of E_0 and E_1: Informally, $\mathcal{A}$ behaves like $\mathcal{D}_0$ up to Step 4 and like $\mathcal{D}_1$ afterwards, with an additional construction-oracle query inserted as Step 4.5. We introduce E_3 for a technical reason to simplify the analysis of Step 7 in E_2.

Depending on the situation, it is more convenient to analyze either $\mathsf{E}_r^{\mathsf{cO}}$, $\mathsf{E}_r^{\mathsf{StO}}$, or E_r^f than the others, and we will switch between them as appropriate.

[15] That is, set $\mathsf{aux1}$ as $|0^{3m/2}\rangle_{\mathsf{aux1}}$ and then add $\mathsf{pad}(Z)||0^{m/2}$ to $\mathsf{aux1}$.

[16] We define $|\mathsf{E}_r^f; i\rangle$, $|\mathsf{E}_r^{\mathsf{StO}}; i\rangle$, and $|\mathsf{E}_r^{\mathsf{cO}}; i\rangle$ only for steps i that are not the final step of the distinguisher in E_r, since the final step involves a measurement.

By Eq. (7), it suffices to analyze all fixed functions $f \in \mathsf{Funcs}$ for the StO case, and the StO and cO cases are equivalent.

We now proceed to the formal proof. Recall that we use the shorthand like $|(i)\rangle$ as explained in Section 2.

Analysis of Steps 1-2 in $\mathsf{E_0}$ and $\mathsf{E_1}$. We first analyze E_0 and E_1 through Step 2 of the distinguishers $\mathcal{D}_0$ and $\mathcal{D}_1$. By construction of $\mathcal{D}_1$, there exist unit vectors $|\Phi_{f,a}\rangle_{XYWT_1S}$ such that

$$|\mathsf{E}_1^{\mathsf{StO}}; 2\rangle = \frac{1}{2^{m/4}\sqrt{N}} \sum_{a\in\{0,1\}^{m/2}, f\in\mathsf{Funcs}} |\Phi_{f,a}\rangle_{XYWT_1S} |a\rangle_Z |f\rangle_F \tag{8}$$

(recall that S denotes the simulator's private workspace). The following lemma holds (see Section C in the appendix for a proof).

Lemma 3. *If we measure registers X, Z, W, T_1, F of $|\mathsf{E}_1^{\mathsf{StO}}; 2\rangle$ in the computational basis, then with probability at least $1 - 2\varepsilon$ the outcome is of the form $|\mathsf{pad}(a)\|IV\rangle_X |a\rangle_Z |f(a)\rangle_W |f\rangle_F |0\rangle_{T_1}$ for some a and f.*

Since Eq. (8) and Lemma 3 hold (and the value in register Y is equal to that in W at the end of Step 2 of $\mathcal{D}_1$), there exist some subunit vectors[17] $|\phi_{f,a}\rangle_S$ satisfying

$$|\mathsf{E}_1^{\mathsf{StO}}; 2\rangle \overset{\sqrt{2\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2} \\ f\in\mathsf{Funcs}}} \begin{array}{r} |\mathsf{pad}(a)||IV\rangle_X |f(a)\rangle_Y |a\rangle_Z \\ \otimes |f(a)\rangle_W |\phi_{f,a}\rangle_S |f\rangle_F |0\rangle_{T_1} . \end{array} \tag{9}$$

Furthermore, the difference between $\mathcal{D}_0$ and $\mathcal{D}_1$ in Steps 1-2 is only whether the value of Y is added to W in Step 1. Since this addition commutes with Step 2, we also have

$$|\mathsf{E}_0^{\mathsf{StO}}; 2\rangle \overset{\sqrt{2\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2} \\ f\in\mathsf{Funcs}}} \begin{array}{r} |\mathsf{pad}(a)||IV\rangle_X |f(a)\rangle_Y |a\rangle_Z \\ \otimes |0^n\rangle_W |\phi_{f,a}\rangle_S |f\rangle_F |0\rangle_{T_1} . \end{array} \tag{10}$$

Analysis of Steps 3-4 in $\mathsf{E_0}$ and $\mathsf{E_1}$. Steps 3 and 4 are common between $\mathcal{D}_0$ and $\mathcal{D}_1$, and only the registers X, Y, S are modified (and T_2 is introduced) in those steps. Since Eq. (9)-(10) hold, there exist subunit vectors[18] $|\xi_{f,a}\rangle_{XYST_2}$

[17] More explicitly, $|\phi_{f,a}\rangle_S$ can be taken so that $\Pi^{f,a} |\Phi_{f,a}\rangle_{XYWT_1S} = |\mathsf{pad}(a)||IV\rangle_X |f(a)\rangle_Y |f(a)\rangle_W |0\rangle_{T_1} |\phi_{f,a}\rangle_S$ holds, where $\Pi^{f,a}$ is the projector onto the space such that the registers X, Y, W, and T_1 are of the form $|\mathsf{pad}(a)||IV\rangle_X |f(a)\rangle_Y |f(a)\rangle_W |0\rangle_{T_1}$.

[18] Specifically, $|\xi_{f,a}\rangle_{XYST_2} = U |\mathsf{pad}(a)||IV\rangle_X |f(a)\rangle_Y |\phi_{f,a}\rangle_S |0\rangle_{T_2}$ holds, where U is the unitary operator corresponding to Steps 3-4 of $\mathcal{D}_0$ (or $\mathcal{D}_1$) and the procedure of $\mathcal{S}$ to respond to the query in Step 3.

that satisfy

$$|\mathsf{E}_0^{\mathsf{StO}};4\rangle \overset{\sqrt{2\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2}\\ f\in\mathsf{Funcs}}} |a\rangle_Z |0^n\rangle_W |\xi_{f,a}\rangle_{XYST_2} |f\rangle_F |0\rangle_{T_1}, \tag{11}$$

$$|\mathsf{E}_1^{\mathsf{StO}};4\rangle \overset{\sqrt{2\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2}\\ f\in\mathsf{Funcs}}} |a\rangle_Z |f(a)\rangle_W |\xi_{f,a}\rangle_{XYST_2} |f\rangle_F |0\rangle_{T_1}. \tag{12}$$

Register T_2 is not modified after Step 4. Thus, the probability that measuring T_2 of $|\mathsf{E}_0^{\mathsf{StO}};4\rangle$ (resp., $|\mathsf{E}_1^{\mathsf{StO}};4\rangle$) yields 1 is equal to the probability that the measurement of T_2 in Step 5 of $\mathcal{D}_0$ (resp., Step 7 of $\mathcal{D}_1$) yields 1, which is upper bounded by $p_0 \le \varepsilon$ (resp., $p_1^{T_2} \le \varepsilon$). Hence,

$$|\mathsf{E}_r^{\mathsf{StO}};4\rangle \overset{\sqrt{\varepsilon}}{\approx} \Pi_{X,Y,T_2=0} |\mathsf{E}_r^{\mathsf{StO}};4\rangle \tag{13}$$

holds for both $r = 0$ and $r = 1$, where $\Pi_{X,Y,T_2=0}$ is the projector onto the space spanned by vectors such that registers X, Y, and T_2 are of the form $|0^{m+n}\rangle_X |0^n\rangle_Y |0\rangle_{T_2}$. Since Eq. (11)-(13) holds, there exist[19] subunit vectors $|\zeta_{f,a}\rangle_S$ that satisfy

$$|\mathsf{E}_0^{\mathsf{StO}};4\rangle \overset{3\sqrt{\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2}\\ f\in\mathsf{Funcs}}} |0^{m+2n}\rangle_{XY} |a\rangle_Z |0^n\rangle_W \otimes |\zeta_{f,a}\rangle_S |f\rangle_F |0,0\rangle_{T_1T_2}, \tag{14}$$

$$|\mathsf{E}_1^{\mathsf{StO}};4\rangle \overset{3\sqrt{\varepsilon}}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2}\\ f\in\mathsf{Funcs}}} |0^{m+2n}\rangle_{XY} |a\rangle_Z |f(a)\rangle_W \otimes |\zeta_{f,a}\rangle_S |f\rangle_F |0,0\rangle_{T_1T_2}. \tag{15}$$

The following lemma holds. (We provide a proof in Section D in the appendix.)

Lemma 4. *Let* $\varepsilon_4' := 4\sqrt[4]{\varepsilon} + 3\sqrt{\varepsilon}$ *and* $|\zeta_f\rangle_S := \sum_{a\in\{0,1\}^{m/2}} \frac{1}{2^{m/2}} |\zeta_{f,a}\rangle_S$ *for each* f*. Then, it holds that*

$$|\mathsf{E}_0^{\mathsf{StO}};4\rangle \overset{\varepsilon_4'}{\approx} \left(\frac{1}{2^{m/4}} \sum_{a\in\{0,1\}^{m/2}} |0^{m+2n}\rangle_{XY} |a\rangle_Z |0^n\rangle_W\right) \otimes \left(\frac{1}{\sqrt{N}} \sum_{f\in\mathsf{Funcs}} |\zeta_f\rangle_S |f\rangle_F |0,0\rangle_{T_1T_2}\right), \tag{16}$$

$$|\mathsf{E}_1^{\mathsf{StO}};4\rangle \overset{\varepsilon_4'}{\approx} \frac{1}{2^{m/4}\sqrt{N}} \sum_{\substack{a\in\{0,1\}^{m/2}\\ f\in\mathsf{Funcs}}} |0^{m+2n}\rangle_{XY} |a\rangle_Z |f(a)\rangle_W \otimes |\zeta_f\rangle_S |f\rangle_F |0,0\rangle_{T_1T_2}. \tag{17}$$

[19] $|\zeta_{f,a}\rangle_S$ can be taken so that $|0^{m+n}\rangle_X |0^n\rangle_Y |\zeta_{f,a}\rangle_S |0\rangle_{T_2} = \Pi_{X,Y,T_2=0} |\xi_{f,a}\rangle_{XYST_2}$.

Remark 4. Roughly speaking, Eq. (16) reflects the intuition that if p_0 is small ($\mathcal{D}_0$ returns 1 only with negligible probability), then the quantum states of $\mathcal{D}_0$ and the simulator $\mathcal{S}$ are not much entangled. Eq. (17) shows that the state of E_1 at Step 4 can be created by starting from the corresponding state of E_0 and making a construction query from the Z register in a way that the response will be written to the W register (this corresponds to Step 4.5 of $\mathcal{A}$, and it means that $\mathcal{A}$ of E_2 bridges the analyses of E_0 and E_1 well).

Analysis of Steps 4-5 in E_2. E_2 is identical to E_0 until Step 4 of the distinguishers ($\mathcal{A}$ and $\mathcal{D}_0$). Thus, we have

$$|\mathsf{E}_2^{\mathsf{StO}}; 4\rangle = |\mathsf{E}_0^{\mathsf{StO}}; 4\rangle\,. \tag{18}$$

If a direct construction query is made from Z register of the state $|(16)\rangle$ and the response is written into W register, the state changes to $|(17)\rangle$. Therefore, we have

$$\begin{aligned} |\mathsf{E}_2^{\mathsf{StO}}; 4.5\rangle &= \mathsf{StO}_{ZWF}\,|\mathsf{E}_2^{\mathsf{StO}}; 4\rangle = \mathsf{StO}_{ZWF}\,|\mathsf{E}_0^{\mathsf{StO}}; 4\rangle \\ &\overset{\varepsilon_4'}{\approx} \mathsf{StO}_{ZWF}\,|(16)\rangle = |(17)\rangle \overset{\varepsilon_4'}{\approx} |\mathsf{E}_1^{\mathsf{StO}}; 4\rangle\,. \end{aligned} \tag{19}$$

In addition,

$$|\mathsf{E}_2^{\mathsf{cO}}; 4\rangle = |\mathsf{E}_0^{\mathsf{cO}}; 4\rangle \tag{20}$$

follows from Eq. (7) and (18).

For any j, let $\Pi_{|D|\le j}$ be the projector onto the space spanned by the vectors with databases of size at most j. Since $\mathcal{S}$ makes at most q queries in total,

$$|\mathsf{E}_2^{\mathsf{cO}}; 4\rangle = |\mathsf{E}_0^{\mathsf{cO}}; 4\rangle = \Pi_{|D|\le q}\,|\mathsf{E}_0^{\mathsf{cO}}; 4\rangle \tag{21}$$

follows from Lemma 1 and Eq. (20). Since Eq. (7) and (16) hold, we have

$$\begin{aligned} |(21)\rangle &= \Pi_{|D|\le q}\mathsf{Comp}_F\,|\mathsf{E}_0^{\mathsf{StO}}; 4\rangle \overset{\varepsilon_4'}{\approx} \Pi_{|D|\le q}\mathsf{Comp}_F\,|(16)\rangle \\ &= \left(\frac{1}{2^{m/4}}\sum_{a\in\{0,1\}^{m/2}} |0^{m+2n}\rangle_{XY}\,|a\rangle_Z\,|0^n\rangle_W\right)\left(\sum_{\substack{D\in\mathsf{DBs}\\ |D|\le q}} |\zeta_D\rangle_S\,|D\rangle_F\,|0,0\rangle_{T_1T_2}\right), \end{aligned} \tag{22}$$

where $|\zeta_D\rangle_S := \frac{1}{\sqrt{N}}\sum_{f\in\mathsf{Funcs}}\langle D|\mathsf{Comp}|f\rangle\,|\zeta_f\rangle_S$.

Let $\Pi_{ZF}^{\perp}$ be the projector onto the space spanned by the vectors $|a\rangle_Z\,|D\rangle_F$ such that $D(a)=\perp$ and $D(\mathsf{pad}(a)||0^{m/2})=\perp$. Then, we have

$$\left\|\left(I-\Pi_{ZF}^{\perp}\right)|(22)\rangle\right\|^2 \le \frac{2q}{2^{m/2}} \tag{23}$$

because register Z in $|(22)\rangle$ is a uniform superposition of $a \in \{0,1\}^{m/2}$ and unentangled from F, and the size of the databases are at most q. Thus,

$$\begin{aligned}|\mathsf{E}_2^{\mathsf{cO}};4\rangle &\overset{\text{Eq. (21)}}{=} |\mathsf{E}_0^{\mathsf{cO}};4\rangle \overset{\varepsilon_4'}{\approx} |(22)\rangle \overset{\sqrt{2q/2^{m/2}}}{\approx} \Pi_{ZF}^{\perp}|(22)\rangle \\ &= \frac{1}{2^{m/4}} \sum_{\substack{a\in\{0,1\}^{m/2},D\in\mathsf{DBs}\\ |D|\le q, D(a)=\perp,\\ D(\mathsf{pad}(a)||0^{m/2})=\perp}} |0^{m+2n}\rangle_{XY}|a\rangle_Z|0^n\rangle_W \\ &\qquad\qquad |\zeta_D\rangle_S|D\rangle_F|0,0\rangle_{T_1T_2}\end{aligned} \tag{24}$$

holds. Note that

$$\||(24)\rangle\| \le 1 \tag{25}$$

because $\| |(24)\rangle \| = \left\|\Pi_{ZF}^{\perp}|(22)\rangle\right\| \le \||(22)\rangle\| = \left\|\Pi_{|D|\le q}\mathsf{Comp}_F|(16)\rangle\right\| \le \||(16)\rangle\| \le 1$. In addition, the following lemma holds (see Section E in the appendix for a proof, which is a straightforward application of Lemma 2).

Lemma 5. $\mathsf{cO}_{ZWF}|(24)\rangle$ *is* $(2/2^{n/2})$*-close to*

$$\frac{1}{2^{m/4+n/2}} \sum_{\substack{a\in\{0,1\}^{m/2},c\in\{0,1\}^n,D\in\mathsf{DBs}\\ |D|\le q, D(a)=\perp,\\ D(\mathsf{pad}(a)||0^{m/2})=\perp}} |0^{m+2n}\rangle_{XY}|a\rangle_Z|c\rangle_W \\ |\zeta_D\rangle_S|D[a\mapsto c]\rangle_F|0,0\rangle_{T_1T_2}. \tag{26}$$

By using Lemma 5 and the definition of Step 4.5 of $\mathcal{A}$, we obtain

$$|\mathsf{E}_2^{\mathsf{cO}};4.5\rangle = \mathsf{cO}_{ZWF}|\mathsf{E}_2^{\mathsf{cO}};4\rangle \overset{\varepsilon_4'+\sqrt{2q/2^{m/2}}}{\approx} \mathsf{cO}_{ZWF}|(24)\rangle \overset{2/2^{n/2}}{\approx} |(26)\rangle.$$

This implies that

$$|\mathsf{E}_2^{\mathsf{cO}};4.5\rangle \overset{\varepsilon_5'}{\approx} |(26)\rangle \tag{27}$$

holds, where

$$\varepsilon_5' := \varepsilon_4' + \sqrt{2q/2^{m/2}} + 2/2^{n/2} = 4\sqrt[4]{\varepsilon} + 3\sqrt{\varepsilon} + \sqrt{2q/2^{m/2}} + 2/2^{n/2}. \tag{28}$$

Step 5 of $\mathcal{A}$ is the same as Step 5 of $\mathcal{D}_1$, which adds $\left(0^{m/2}||\mathsf{enc}\left(\frac{3m}{2}\right)||W\right)$ to X. By denoting this operation by $U_{\mathcal{A},5}$, we have

$$\begin{aligned}|\mathsf{E}_2^{\mathsf{cO}};5\rangle &= U_{\mathcal{A},5}|\mathsf{E}_2^{\mathsf{cO}};4.5\rangle \overset{\varepsilon_5'}{\approx} U_{\mathcal{A},5}|(26)\rangle \\ &= \frac{1}{2^{\frac{m}{4}+\frac{n}{2}}} \sum_{\substack{a\in\{0,1\}^{m/2},c\in\{0,1\}^n,D\in\mathsf{DBs}\\ |D|\le q, D(a)=\perp,\\ D(\mathsf{pad}(a)||0^{m/2})=\perp}} |0^{m/2}||\mathsf{enc}(3m/2)||c\rangle_X|0^n\rangle_Y|a\rangle_Z|c\rangle_W \\ &\qquad\qquad |\zeta_D\rangle_S|D[a\mapsto c]\rangle_F|0,0\rangle_{T_1T_2}.\end{aligned} \tag{29}$$

Moreover, since $\| \, |(29)\rangle \, \|^2 = \|U_{\mathcal{A},5} \, |(26)\rangle \, \|^2 = \| \, |(26)\rangle \, \|^2 = \frac{1}{2^n} \sum_c \| \, |(24)\rangle \, \|^2 \leq 1$ (by Eq. (25)), we have

$$\| \, |(29)\rangle \, \| \leq 1. \tag{30}$$

Remark 5. Intuitively, Eq. (29) indicates that, at Step 5 of $\mathcal{A}$ in E_2, the simulator's private register S carries (almost) no information about the value a in Z (nor about $\mathsf{pad}(a)$): The ket vector on register S is of the form $|\zeta_D\rangle$ and contains no explicit dependence on a. Any correlation between S and a can only arise indirectly through the conditioning on D in the superposition (e.g., $D(a) = \perp$), and we will bound this effect in the subsequent analysis. We will show that this near-independence of S from $(a, \mathsf{pad}(a))$ persists through Steps 6–7 in E_2 and E_3, which eventually implies that $\mathcal{S}$ fails to simulate the primitive oracle and that $\mathcal{D}_1$ in E_1 outputs 1 with non-negligible probability.

Analysis of Step 6 in E_2. Step 6 is common between $\mathcal{D}_1$ and $\mathcal{A}$: The distinguishers make a primitive query from register X. To respond to the query, $\mathcal{S}$ performs some computation, making some quantum queries to RO, and finally writes the response into register Y.

Without loss of generality, the behavior of $\mathcal{S}$ in Step 6 is modeled as follows:

- The register S (for $\mathcal{S}$'s local memory) is divided into three sub-registers X', Y', and S'. Here, X' and Y' are the registers for $\mathcal{S}$ to make queries to (the compressed oracle of) the construction oracle RO and receive the response, respectively, and S' is the remaining part of S.
- $\mathcal{S}$ responds to the primitive query by applying unitary operators U_0, $\mathsf{cO}_{X'Y'F}$, U_1, …, U_{q-1}, $\mathsf{cO}_{X'Y'F}$, U_q on the register $S = X'Y'S'$ in a sequential order[20]. (Note that applying cO means that $\mathcal{S}$ makes a query to the compressed oracle of RO, and U_i corresponds to computation between $\mathcal{S}$'s $(i-1)$th and ith queries.)

In particular, we have

$$|\mathsf{E}_2^{\mathsf{cO}}; 6\rangle = U_{\mathcal{S}} \, |\mathsf{E}_2^{\mathsf{cO}}; 5\rangle \, , \tag{31}$$

where

$$U_{\mathcal{S}} := U_q \cdot \mathsf{cO}_{X'Y'F} \cdot U_{q-1} \cdots U_1 \cdot \mathsf{cO}_{X'Y'F} \cdot U_0. \tag{32}$$

Recall that $\Pi_{|D| \leq j}$ is the projector onto the space spanned by the vectors with database (in register F) of size at most j. We introduce a few additional projectors as follows (we will provide intuition later in Remark 6).

[20] This implicitly assumes that $\mathcal{S}$ makes q queries in this step, though the number of queries here may actually be less than q (because q is the upper bound of the number of total queries by $\mathcal{S}$ to RO through the experiments). In such a case, we increase the number of queries made by $\mathcal{S}$ by modifying it to make some dummy queries.

- Let Π_{hit} (resp., $\Pi_{\neg\mathsf{hit}}$) be the projector onto the space spanned by the vectors $|a\rangle_Z |x'\rangle_{X'}$ such that[21] $x' = a$ or $x' = \mathsf{pad}(a)||0^{m/2}$ (resp., $x' \neq a$ and $x' \neq \mathsf{pad}(a)||0^{m/2}$). Note that $\Pi_{\mathsf{hit}} + \Pi_{\neg\mathsf{hit}} = I$ holds.
- Let Π_{good} be the projector onto the space spanned by vectors of the form

$$\sum_{\substack{a\in\{0,1\}^{m/2}, D\in\mathsf{DBs}\\ D(a)=\bot\\ D(\mathsf{pad}(a)||0^{m/2})=\bot}} \frac{1}{2^{m/4}} |a\rangle_Z |c\rangle_W |D[a\mapsto c]\rangle_F |\eta_{c,D}\rangle_{XYS} |0,0\rangle_{T_1T_2} \tag{33}$$

  for some $c \in \{0,1\}^n$ and some vectors $|\eta_{c,D}\rangle_{XYS}$.
- Let $\Pi_{\mathsf{good}\wedge|D|\le j} := \Pi_{\mathsf{good}}\Pi_{|D|\le j}$ $\left(= \Pi_{|D|\le j}\Pi_{\mathsf{good}}\right)$.

In addition, define U_i' and $U_{\mathcal{S}}'$ by

$$U_i' := \Pi_{\neg\mathsf{hit}} U_i \Pi_{\mathsf{good}\wedge|D|\le q+i+1},$$
$$U_{\mathcal{S}}' := U_q' \cdot \mathsf{cO}_{X'Y'F} \cdot U_{q-1}' \cdots U_1' \cdot \mathsf{cO}_{X'Y'F} \cdot U_0'.$$

Then, the following two lemmas hold. The proofs are given in Section F and G in the appendix, respectively.

Lemma 6. *For any $j \ge 1$, the image of $\mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j}$ is unchanged by the projector $\Pi_{\mathsf{good}\wedge|D|\le j+1}$. That is,*

$$\mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j} = \Pi_{\mathsf{good}\wedge|D|\le j+1}\mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j}. \tag{34}$$

Moreover, $\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j} \overset{2\sqrt{2}/\sqrt{2^{m/2}}}{\approx} \Pi_{\mathsf{good}\wedge|D|\le j}$ holds for any $j \le 2^{m/2-2}$. That is,

$$\left\|\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j} - \Pi_{\mathsf{good}\wedge|D|\le j}\right\|_{\mathrm{op}} \left(= \left\|\Pi_{\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j}\right\|_{\mathrm{op}}\right) \le \frac{2\sqrt{2}}{\sqrt{2^{m/2}}}. \tag{35}$$

Lemma 7. $\Pi_{\mathsf{good}\wedge|D|\le 2q+2}U_{\mathcal{S}}' \overset{2\sqrt{2}(q+1)/2^{m/4}}{\approx} U_{\mathcal{S}}\Pi_{\mathsf{good}\wedge|D|\le q+1}$.

Remark 6. Intuitively, "good" states (vectors of the form Eq. (33)) correspond to the following situation: The registers that $\mathcal{S}$ can manipulate (i.e., X, Y, S) are only weakly correlated with Z (and hence with the values of a and $\mathsf{pad}(a)$), except possibly through the conditioning on the database register F (cf. Remark 5). Eq. (29) shows that the state after Step 5 of $\mathcal{A}$ in $\mathsf{E}_2^{\mathsf{cO}}$ is close to good. What we are trying to show here is that the state remains "good" throughout Step 6 (i.e., the computation and queries by $\mathcal{S}$ performed during Step 6 of $\mathcal{A}$, which correspond to $U_{\mathcal{S}}$ of Eq. (32), do not make the state "bad") unless the following event happens: the simulator $\mathcal{S}$ queries a or $\mathsf{pad}(a)||0^{m/2}$ to cO. This event

[21] Length of inputs to RO is variable, and the register X' contains bit strings x' of various length. As mentioned in Section 2, we are assuming that the length of each input x' is encoded with it.

corresponds to the range (support) of the projector Π_{hit}. Eq. (34) formalizes that the simulator's queries to cO preserve the "good" subspace on the $\neg\mathsf{hit}$ component, and Eq. (35) shows that the hit component is unlikely to occur if the state is good. Lemma 7 then combines these facts (over the q queries by $\mathcal{S}$ to RO in Step 6), roughly implying that if a state is good before $U_{\mathcal{S}}$, then it remains good after $U_{\mathcal{S}}$ up to negligible error.

By definition of Π_{good}, there exist some subunit vectors[22] $|\xi_{c,D,x,d}\rangle$ such that

$$\begin{aligned}&\Pi_{\mathsf{good}\wedge|D|\le 2q+2}U'_{\mathcal{S}}\,|(29)\rangle\\ &\quad=\sum_{\substack{a\in\{0,1\}^{m/2},D,c,x,d\\ D(a)=\perp,|D|\le 2q+1\\ D(\mathsf{pad}(a)||0^{m/2})=\perp}}\frac{1}{2^{m/4}}\,|x\rangle_X\,|d\rangle_Y\,|a\rangle_Z\,|c\rangle_W\\ &\qquad\qquad|D[a\mapsto c]\rangle_F\,|\xi_{c,D,x,d}\rangle_S\,|0,0\rangle_{T_1T_2}\,,\end{aligned}\tag{36}$$

and

$$\|\,|(36)\rangle\,\|=\|\Pi_{\mathsf{good}\wedge|D|\le 2q+2}U'_{\mathcal{S}}\,|(29)\rangle\,\|\le 1\tag{37}$$

follows from Eq. (30) and the definition of $U'_{\mathcal{S}}$. In addition,

$$\Pi_{\mathsf{good}\wedge|D|\le q+1}\,|(29)\rangle=|(29)\rangle\tag{38}$$

holds because $|(29)\rangle$ is a linear combination of the vectors of the form Eq. (33). Therefore, we have

$$\begin{aligned}|\mathsf{E}_2^{\mathsf{cO}};6\rangle&\overset{(*)}{=}U_{\mathcal{S}}\,|\mathsf{E}_2^{\mathsf{cO}};5\rangle\overset{\varepsilon'_5}{\approx}U_{\mathcal{S}}\,|(29)\rangle\overset{(**)}{=}U_{\mathcal{S}}\Pi_{\mathsf{good}\wedge|D|\le q+1}\,|(29)\rangle\\&\overset{2\sqrt{2}(q+1)/2^{m/4}}{\approx}\Pi_{\mathsf{good}\wedge|D|\le 2q+2}U'_{\mathcal{S}}\,|(29)\rangle=|(36)\rangle\end{aligned}$$

where $(*)$ follows from Eq. (31), $(**)$ follows from Eq. (38), and the second approximated equality follows from Lemma 7. That is,

$$|\mathsf{E}_2^{\mathsf{cO}};6\rangle\overset{\varepsilon'_6}{\approx}|(36)\rangle\tag{39}$$

with

$$\begin{aligned}\varepsilon'_6&:=\varepsilon'_5+\frac{2\sqrt{2}(q+1)}{2^{m/4}}=4\sqrt[4]{\varepsilon}+3\sqrt{\varepsilon}+\sqrt{2q/2^{m/2}}+\frac{2}{2^{n/2}}+\frac{2\sqrt{2}(q+1)}{2^{m/4}}\\&\le 4\sqrt[4]{\varepsilon}+8\sqrt{\varepsilon}+2/2^{n/2}+4\sqrt{2}(q+1)/2^{m/4}.\end{aligned}\tag{40}$$

Remark 7. Eq. (39) shows that the state of E_2 just after Step 6 of $\mathcal{A}$ is close to a "good" state, i.e., the registers controlled by the simulator $\mathcal{S}$ carry (almost) no information about the values of a and $\mathsf{pad}(a)$. This implies that, in Step 7 of $\mathcal{A}$ in E_2 (and of $\mathcal{B}$ in E_3), the check $d=\delta$ is unlikely to be satisfied (and hence the simulation by $\mathcal{S}$ is likely to fail). For Step 7, we will mainly analyze $\mathcal{B}$ in E_3 instead of $\mathcal{A}$ in E_2 because the former is easier to analyze than the latter.

[22] Here, $|\eta_{c,D}\rangle_{XYS}$ of Eq. (33) corresponds to $\sum_{x,d}|x\rangle_X\,|d\rangle_Y\,|\xi_{c,D,x,d}\rangle_S$. The vectors $|\xi_{c,D,x,d}\rangle_S$ can be taken as subunit because $\|\,|(36)\rangle\,\|\le 1$.

Analysis of Step 7 in E_2 and E_3. Since E_2 and E_3 differ only in Step 7 of $\mathcal{A}$ and $\mathcal{B}$, we have

$$|\mathsf{E}_3^{\mathsf{cO}}; i\rangle = |\mathsf{E}_2^{\mathsf{cO}}; i\rangle \text{ for } i = 1, \ldots, 6. \tag{41}$$

In addition,

$$\Pr[1 \leftarrow \mathsf{E}_3] \leq \Pr[1 \leftarrow \mathsf{E}_2] \tag{42}$$

holds by definition of the experiments[23].

Step 7 of $\mathcal{B}$ first copies $\mathsf{pad}(Z)||0^{m/2}$ into a new auxiliary register $\mathsf{aux1}$ (we denote this unitary operation by $U_{\mathcal{B},7}$), and then query the value to the construction oracle, where the response is written into another new auxiliary register $\mathsf{aux2}$ (in $\mathsf{E}_3^{\mathsf{cO}}$, this means to apply $\mathsf{cO}_{\mathsf{aux1aux2}F}$). We have

$$|\mathsf{E}_3^{\mathsf{cO}}; \text{b.m.}\rangle = \mathsf{cO}_{\mathsf{aux1aux2}F} U_{\mathcal{B},7} |\mathsf{E}_3^{\mathsf{cO}}; 6\rangle, \tag{43}$$

where $|\mathsf{E}_3^{\mathsf{cO}}; \text{b.m.}\rangle$ is the state of $\mathsf{E}_3^{\mathsf{cO}}$ just before the measurement in Step 7 of $\mathcal{B}$.

Since $|\mathsf{E}_3^{\mathsf{cO}}; 6\rangle = |\mathsf{E}_2^{\mathsf{cO}}; 6\rangle \overset{\varepsilon_6'}{\approx} |(36)\rangle$ holds by Eq. (39) and (41), we have

$$\begin{aligned}
|\mathsf{E}_3^{\mathsf{cO}}; \text{b.m.}\rangle &= \mathsf{cO}_{\mathsf{aux1aux2}F} U_{\mathcal{B},7} |\mathsf{E}_3^{\mathsf{cO}}; 6\rangle \overset{\varepsilon_6'}{\approx} \mathsf{cO}_{\mathsf{aux1aux2}F} U_{\mathcal{B},7} |(36)\rangle \\
&= \mathsf{cO}_{\mathsf{aux1aux2}F} \sum_{\substack{a\in\{0,1\}^{m/2}, D, c, x, d \\ D(a)=\perp, |D|\leq 2q+1 \\ D(\mathsf{pad}(a)||0^{m/2})=\perp}} \frac{1}{2^{m/4}} |x\rangle_X |d\rangle_Y |a\rangle_Z |c\rangle_W |D[a \mapsto c]\rangle_F |\xi_{c,D,x,d}\rangle_S \\
&\qquad\qquad |0,0\rangle_{T_1T_2} |\mathsf{pad}(a)||0^{m/2}, 0\rangle_{\mathsf{aux1aux2}}.
\end{aligned} \tag{44}$$

In addition, the following lemma holds (the proof is the same as Lemma 5; a straightforward application of Lemma 2).

Lemma 8. $|(44)\rangle$ *is* $\left(2/2^{n/2}\right)$*-close to*

$$\begin{aligned}
\sum_{\substack{a\in\{0,1\}^{m/2}, D, c, x, d \\ D(a)=\perp, |D|\leq 2q+1 \\ D(\mathsf{pad}(a)||0^{m/2})=\perp \\ \delta\in\{0,1\}^n}} &\frac{1}{2^{m/4+n/2}} |x\rangle_X |d\rangle_Y |a\rangle_Z |c\rangle_W \\
&\otimes |D[a \mapsto c, \mathsf{pad}(a)||0^{m/2} \mapsto \delta]\rangle_F |\xi_{c,D,x,d}\rangle_S \\
&\otimes |0,0\rangle_{T_1T_2} |\mathsf{pad}(a)||0^{m/2}, \delta\rangle_{\mathsf{aux1aux2}}.
\end{aligned} \tag{45}$$

Combining Lemma 8 with Eq. (44), we obtain

$$|\mathsf{E}_3^{\mathsf{cO}}; \text{b.m.}\rangle \overset{\varepsilon_7'}{\approx} |(45)\rangle, \tag{46}$$

where

$$\begin{aligned}
\varepsilon_7' &:= \varepsilon_6' + 2/2^{n/2} \leq 4\sqrt[4]{\varepsilon} + 8\sqrt{\varepsilon} + 4/2^{n/2} + 4\sqrt{2}(q+1)/2^{m/4} \\
&\leq 4\sqrt[4]{\varepsilon} + 8\sqrt{\varepsilon} + O(q/2^{m/4} + 1/2^{n/2}).
\end{aligned} \tag{47}$$

[23] Note that the probability that an experiment returns 1 does not depend on how the quantum oracle of RO is implemented.

Let $\Pi_{Y=\mathsf{aux2}}$ be the projector onto the space spanned by vectors such that the registers Y and $\mathsf{aux2}$ are of the form $|d\rangle_Y |\delta\rangle_{\mathsf{aux2}}$ with $d = \delta$. Then, some straightforward calculation shows that

$$\|\Pi_{Y=\mathsf{aux2}} |(45)\rangle\| = \frac{1}{\sqrt{2^n}} \||(36)\rangle\| \leq \frac{1}{\sqrt{2^n}} \tag{48}$$

(the last inequality follows from Eq. (37)). Therefore, we have

$$\begin{aligned}
\Pr[0 \leftarrow \mathsf{E}_3] = & \underbrace{\left\|\Pi_{Y=\mathsf{aux2}} |\mathsf{E}_3^{\mathsf{cO}}; \text{b.m.}\rangle\right\|^2}_{\leq \left(\|\Pi_{Y=\mathsf{aux2}}|(45)\rangle\| + \varepsilon_7'\right)^2 \text{ by Eq. (46)}} \\
& \leq \|\Pi_{Y=\mathsf{aux2}} |(45)\rangle\|^2 + 2\varepsilon_7' \cdot \|\Pi_{Y=\mathsf{aux2}} |(45)\rangle\| + (\varepsilon_7')^2 \\
& \leq \frac{1}{2^n} + 2\varepsilon_7' \cdot \frac{1}{\sqrt{2^n}} + (\varepsilon_7')^2,
\end{aligned} \tag{49}$$

where the last inequality follows from Eq. (48).

Finishing the Proof of Proposition 1. Steps 5-7 of $\mathcal{A}$ are identical to those of $\mathcal{D}_1$. Hence, $|\mathsf{E}_2^{\mathsf{StO}}; i\rangle \overset{2\varepsilon_4'}{\approx} |\mathsf{E}_1^{\mathsf{StO}}; i\rangle$ follows from Eq. (19) for $i = 5, 6$, and thus $\left|\Pr[1 \leftarrow \mathsf{E}_1] - \Pr[1 \leftarrow \mathsf{E}_2]\right| \leq \left\||\mathsf{E}_1^{\mathsf{cO}}; 6\rangle - |\mathsf{E}_2^{\mathsf{cO}}; 6\rangle\right\| \leq 2\varepsilon_4'$ holds. Consequently, we obtain

$$\begin{aligned}
\Pr[1 \leftarrow \mathsf{E}_1] \geq & \underbrace{\Pr[1 \leftarrow \mathsf{E}_2]}_{\geq \Pr[1 \leftarrow \mathsf{E}_3] \text{ by Eq. (42)}} - \underbrace{\left|\Pr[1 \leftarrow \mathsf{E}_1] - \Pr[1 \leftarrow \mathsf{E}_2]\right|}_{\leq 2\varepsilon_4'} \\
& \geq \Pr[1 \leftarrow \mathsf{E}_3] - 2\varepsilon_4' = 1 - \underbrace{\Pr[0 \leftarrow \mathsf{E}_3]}_{\text{bounded as in Eq. (49)}} - 2\varepsilon_4' \\
& \geq 1 - \frac{1}{2^n} - \frac{2\varepsilon_7'}{\sqrt{2^n}} - (\varepsilon_7')^2 - 2\varepsilon_4' \\
& \geq 1 - 8\sqrt[4]{\varepsilon} - 64\sqrt[4]{\varepsilon^3} - 22\sqrt{\varepsilon} - 64\varepsilon - O\left(q/2^{m/4} + 1/2^{n/2}\right),
\end{aligned}$$

which completes the proof of Proposition 1 (the last inequality is derived by using $\varepsilon_7' \leq 4\sqrt[4]{\varepsilon} + 8\sqrt{\varepsilon} + O(q/2^{m/4} + 1/2^{n/2})$).

Acknowledgment

We thank the anonymous reviewers for their helpful comments, especially the reviewer who suggested how to explain the difference between our result and prior works on the quantum indifferentiability of prefix-free Merkle-Damgård. We used Microsoft 365 Copilot and ChatGPT (accessed June 2026) mainly as a writing aid, including limited drafting of non-technical exposition from author-provided bullet points, language polishing, and suggesting candidate references for related work (all references were independently verified by the authors). We

also used it to perform high-level sanity checks of the exposition (e.g., to flag potentially unclear steps or possible gaps in proof write-up), solely to inform our own manual verification. It was not used to create any technical content (theorems, proofs, attacks, reductions), nor was it relied upon as a source of validation or correctness. All text and all technical claims were ultimately reviewed and validated by the authors, and any AI-assisted (non-technical) drafts were rewritten by the authors.

References

1. Alagic, G., Carolan, J., Majenz, C., Tokat, S.: The sponge is quantum indifferentiable. In: Proceedings of the 66th IEEE Symposium on Foundations of Computer Science (FOCS). IEEE, Sydney, Australia (2025)
2. Bellare, M., Rogaway, P.: Random oracles are practical: A paradigm for designing efficient protocols. In: Denning, D.E., Pyle, R., Ganesan, R., Sandhu, R.S., Ashby, V. (eds.) ACM CCS 93. pp. 62–73. ACM Press (Nov 1993). https://doi.org/10.1145/168588.168596
3. Benedikt, B.J., Fischlin, M.: The order of hashing in Fiat-Shamir schemes. In: Hanaoka, G., Yang, B.Y. (eds.) Advances in Cryptology – ASIACRYPT 2025, Part IV. vol. 16248, pp. 103–137. Springer Nature Singapore, Singapore (2025)
4. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Sponge functions. ECRYPT Hash Workshop (2007)
5. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: On the indifferentiability of the Sponge construction. In: Smart, N.P. (ed.) EUROCRYPT 2008. LNCS, vol. 4965, pp. 181–197. Springer, Berlin, Heidelberg (Apr 2008). https://doi.org/10.1007/978-3-540-78967-3_11
6. Boneh, D., Dagdelen, Ö., Fischlin, M., Lehmann, A., Schaffner, C., Zhandry, M.: Random oracles in a quantum world. In: Lee, D.H., Wang, X. (eds.) ASIACRYPT 2011. LNCS, vol. 7073, pp. 41–69. Springer, Berlin, Heidelberg (Dec 2011). https://doi.org/10.1007/978-3-642-25385-0_3
7. Canteaut, A., Fuhr, T., Naya-Plasencia, M., Paillier, P., Reinhard, J.R., Videau, M.: A unified indifferentiability proof for permutation- or block cipher-based hash functions. Cryptology ePrint Archive, Report 2012/363 (2012), https://eprint.iacr.org/2012/363
8. Carolan, J., Poremba, A., Zhandry, M.: (Quantum) indifferentiability and precomputation. Cryptology ePrint Archive, Report 2024/1727 (2024), https://eprint.iacr.org/2024/1727
9. Carstens, T.V., Ebrahimi, E., Tabia, G.N., Unruh, D.: On quantum indifferentiability. Cryptology ePrint Archive, Report 2018/257 (2018), https://eprint.iacr.org/2018/257
10. Coron, J.S., Dodis, Y., Malinaud, C., Puniya, P.: Merkle-Damgård revisited: How to construct a hash function. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 430–448. Springer, Berlin, Heidelberg (Aug 2005). https://doi.org/10.1007/11535218_26
11. Damgård, I.: A design principle for hash functions. In: Brassard, G. (ed.) CRYPTO'89. LNCS, vol. 435, pp. 416–427. Springer, New York (Aug 1990). https://doi.org/10.1007/0-387-34805-0_39

12. Dobbertin, H., Bosselaers, A., Preneel, B.: RIPEMD-160: A strengthened version of RIPEMD. In: Gollmann, D. (ed.) FSE'96. LNCS, vol. 1039, pp. 71–82. Springer, Berlin, Heidelberg (Feb 1996). https://doi.org/10.1007/3-540-60865-6_44
13. Dodis, Y., Ristenpart, T., Shrimpton, T.: Salvaging Merkle-Damgård for practical applications. In: Joux, A. (ed.) EUROCRYPT 2009. LNCS, vol. 5479, pp. 371–388. Springer, Berlin, Heidelberg (Apr 2009). https://doi.org/10.1007/978-3-642-01001-9_22
14. Don, J., Fehr, S., Majenz, C.: The measure-and-reprogram technique 2.0: Multi-round fiat-shamir and more. In: Micciancio, D., Ristenpart, T. (eds.) CRYPTO 2020, Part III. LNCS, vol. 12172, pp. 602–631. Springer, Cham (Aug 2020). https://doi.org/10.1007/978-3-030-56877-1_21
15. Don, J., Fehr, S., Majenz, C., Schaffner, C.: Security of the Fiat-Shamir transformation in the quantum random-oracle model. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part II. LNCS, vol. 11693, pp. 356–383. Springer, Cham (Aug 2019). https://doi.org/10.1007/978-3-030-26951-7_13
16. Fiat, A., Shamir, A.: How to prove yourself: Practical solutions to identification and signature problems. In: Odlyzko, A.M. (ed.) CRYPTO'86. LNCS, vol. 263, pp. 186–194. Springer, Berlin, Heidelberg (Aug 1987). https://doi.org/10.1007/3-540-47721-7_12
17. Grilo, A.B., Hövelmanns, K., Hülsing, A., Majenz, C.: Tight adaptive reprogramming in the QROM. In: Tibouchi, M., Wang, H. (eds.) ASIACRYPT 2021, Part I. LNCS, vol. 13090, pp. 637–667. Springer, Cham (Dec 2021). https://doi.org/10.1007/978-3-030-92062-3_22
18. Hosoyamada, A., Isobe, T., Todo, Y., Yasuda, K.: A modular approach to the incompressibility of block-cipher-based AEADs. In: Agrawal, S., Lin, D. (eds.) ASIACRYPT 2022, Part II. LNCS, vol. 13792, pp. 585–619. Springer, Cham (Dec 2022). https://doi.org/10.1007/978-3-031-22966-4_20
19. ISO/IEC: ISO/IEC 10118-3: Information technology — Security techniques — Hash-functions — Part 3: Dedicated hash-functions (2018)
20. Kiltz, E., Lyubashevsky, V., Schaffner, C.: A concrete treatment of Fiat-Shamir signatures in the quantum random-oracle model. In: Nielsen, J.B., Rijmen, V. (eds.) EUROCRYPT 2018, Part III. LNCS, vol. 10822, pp. 552–586. Springer, Cham (Apr / May 2018). https://doi.org/10.1007/978-3-319-78372-7_18
21. Liu, Q., Zhandry, M.: Revisiting post-quantum Fiat-Shamir. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part II. LNCS, vol. 11693, pp. 326–355. Springer, Cham (Aug 2019). https://doi.org/10.1007/978-3-030-26951-7_12
22. Liu, Y., Jiang, H., Zhao, Y.: Tighter post-quantum proof for plain FDH, PFDH and GPV-IBE. Cryptology ePrint Archive, Report 2022/1441 (2022), https://eprint.iacr.org/2022/1441
23. Ma, F., Huang, H.Y.: How to construct random unitaries. In: Koucký, M., Bansal, N. (eds.) 57th ACM STOC. pp. 806–809. ACM Press (Jun 2025). https://doi.org/10.1145/3717823.3718254
24. Mandal, A., Patarin, J., Seurin, Y.: On the public indifferentiability and correlation intractability of the 6-round Feistel construction. In: Cramer, R. (ed.) TCC 2012. LNCS, vol. 7194, pp. 285–302. Springer, Berlin, Heidelberg (Mar 2012). https://doi.org/10.1007/978-3-642-28914-9_16
25. Maurer, U.M., Renner, R., Holenstein, C.: Indifferentiability, impossibility results on reductions, and applications to the random oracle methodology. In: Naor, M. (ed.) TCC 2004. LNCS, vol. 2951, pp. 21–39. Springer, Berlin, Heidelberg (Feb 2004). https://doi.org/10.1007/978-3-540-24638-1_2

26. Merkle, R.C.: One way hash functions and DES. In: Brassard, G. (ed.) CRYPTO'89. LNCS, vol. 435, pp. 428–446. Springer, New York (Aug 1990). https://doi.org/10.1007/0-387-34805-0_40
27. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf (2008)
28. National Institute of Standards and Technology: Secure hash standard (shs). FIPS Publication 180-4 (Aug 2015), doi:10.6028/NIST.FIPS.180-4
29. National Institute of Standards and Technology: SHA-3 standard: Permutation-based hash and extendable-output functions. FIPS Publication 202 (Aug 2015), doi:10.6028/NIST.FIPS.202
30. National Institute of Standards and Technology: Module-Lattice-Based Digital Signature Standard. FIPS Publication 204 (Aug 2024), doi:10.6028/NIST.FIPS.204
31. Ristenpart, T., Shacham, H., Shrimpton, T.: Careful with composition: Limitations of the indifferentiability framework. In: Paterson, K.G. (ed.) EUROCRYPT 2011. LNCS, vol. 6632, pp. 487–506. Springer, Berlin, Heidelberg (May 2011). https://doi.org/10.1007/978-3-642-20465-4_27
32. Unruh, D.: Post-quantum security of Fiat-Shamir. In: Takagi, T., Peyrin, T. (eds.) ASIACRYPT 2017, Part I. LNCS, vol. 10624, pp. 65–95. Springer, Cham (Dec 2017). https://doi.org/10.1007/978-3-319-70694-8_3
33. Yamakawa, T., Zhandry, M.: Classical vs quantum random oracles. In: Canteaut, A., Standaert, F.X. (eds.) EUROCRYPT 2021, Part II. LNCS, vol. 12697, pp. 568–597. Springer, Cham (Oct 2021). https://doi.org/10.1007/978-3-030-77886-6_20
34. Zhandry, M.: Secure identity-based encryption in the quantum random oracle model. In: Safavi-Naini, R., Canetti, R. (eds.) CRYPTO 2012. LNCS, vol. 7417, pp. 758–775. Springer, Berlin, Heidelberg (Aug 2012). https://doi.org/10.1007/978-3-642-32009-5_44
35. Zhandry, M.: How to record quantum queries, and applications to quantum indifferentiability. In: Boldyreva, A., Micciancio, D. (eds.) CRYPTO 2019, Part II. LNCS, vol. 11693, pp. 239–268. Springer, Cham (Aug 2019). https://doi.org/10.1007/978-3-030-26951-7_9
36. Zhandry, M.: Redeeming reset indifferentiability and applications to post-quantum security. In: Tibouchi, M., Wang, H. (eds.) ASIACRYPT 2021, Part I. LNCS, vol. 13090, pp. 518–548. Springer, Cham (Dec 2021). https://doi.org/10.1007/978-3-030-92062-3_18
37. Zhandry, M.: How to model unitary oracles. In: Kalai, Y.T., Kamara, S.F. (eds.) CRYPTO 2025, Part II. LNCS, vol. 16001, pp. 237–268. Springer, Cham (Aug 2025). https://doi.org/10.1007/978-3-032-01878-6_8

A Proof of Theorem 1

The proof of Theorem 1 is analogous to the proof given in [31].

Proof (of Theorem 1). Distinguishers for public indifferentiability are given a construction oracle and a primitive oracle, which we denote by $\mathcal{O}_{\mathsf{c}}$ and $\mathcal{O}_{\mathsf{p}}$ (i.e., $(\mathcal{O}_{\mathsf{c}}, \mathcal{O}_{\mathsf{p}})$ is either $(\mathsf{C}[\mathsf{prim}], \mathsf{prim})$ or $(\mathsf{R}, \mathcal{S}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}])$). Define $\mathcal{D}$ as the distinguisher that runs the game G with $\mathcal{A}$, relaying the construction queries and primitive queries to $\mathcal{O}_{\mathsf{c}}$ and $\mathcal{O}_{\mathsf{p}}$ given to itself. That is, $\mathcal{D}\,[\mathcal{O}_{\mathsf{c}}, \mathcal{O}_{\mathsf{p}}] := G\big[\mathcal{O}_{\mathsf{c}}, \mathcal{A}[\mathcal{O}_{\mathsf{p}}]\big]$. In addition, let $\mathcal{B}$ be an adversary (given oracle access to R and $\mathsf{rvl}_{\mathsf{R}}^{(G)}$) that runs

$\mathcal{A}$ and returns the output of $\mathcal{A}$ as its own output, emulating the prim oracle by $\mathcal{S}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]$, i.e., $\mathcal{B}\big[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}\big] := \mathcal{A}\big[\mathcal{S}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]$. (See also Fig. 2.)

By the definition of $\mathcal{D}$ and $\mathcal{B}$,

$$\Pr\Big[1 \leftarrow G\big[\mathsf{C}[\mathsf{prim}], \mathcal{A}[\mathsf{prim}]\big]\Big] = \Pr\Big[1 \leftarrow \mathcal{D}\big[\mathsf{C}[\mathsf{prim}], \mathsf{prim}\big]\Big]$$

and

$$\begin{aligned}\Pr\Big[1 \leftarrow G\big[\mathsf{R}, \mathcal{B}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]\Big] &= \Pr\Big[1 \leftarrow G\big[\mathsf{R}, \mathcal{A}\big[\mathcal{S}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]\big]\Big] \\ &= \Pr\Big[1 \leftarrow \mathcal{D}\big[\mathsf{R}, \mathcal{S}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]\Big]\end{aligned}$$

hold. In addition, the reveal interface $\mathsf{rvl}_{\mathsf{R}}^{(G)}$ of the game G matches the reveal oracle in considering post-quantum public indifferentiability advantage of the above $\mathcal{D}$ with $\mathcal{S}$ because, in the ideal world, the queries to R by $\mathcal{D}$ correspond to the queries to R generated in G other than those by an adversary.

Hence, the inequality

$$\Pr\Big[1 \leftarrow G\big[\mathsf{C}[\mathsf{prim}], \mathcal{A}[\mathsf{prim}]\big]\Big] \leq \Pr\Big[1 \leftarrow G\big[\mathsf{R}, \mathcal{B}[\mathsf{R}, \mathsf{rvl}_{\mathsf{R}}^{(G)}]\big]\Big] + \mathsf{Adv}^{\mathsf{pqpub\text{-}indiff}}_{\mathsf{R},\mathsf{C}[\mathsf{prim}],\mathcal{S}}(\mathcal{D})$$

follows.

The claims about the efficiency also follow from the definition of $\mathcal{D}$ and $\mathcal{B}$. □

B Remarks on the Padding Assumption

In Section 6, we assume that the padding function satisfies $|\mathsf{pad}(x)| = m$ if $|x| = m/2$, and $|\mathsf{pad}(x)| = 2m$ if $|x| = 3m/2$. We believe that it may be possible to prove the impossibility result in a more general setting without this assumption (or under a somewhat weaker assumption). Even so, we impose the assumption for the following reasons (i) and (ii).

(i) If the assumption were completely absent or weaker, the description of the distinguishers and the proofs would become unnecessarily complicated in non-essential aspects, and we wanted to avoid that. This is the primary reason why we chose to impose a condition on the padding.
(ii) As for why we impose the assumption only on messages of length $m/2$ or $3m/2$, this is because we adjusted it so that it would remain as simple as possible while still covering SHA-256 and SHA-512, the most important instantiations.

More details are as follows.

First, in Step 1 of the distinguishers $\mathcal{D}_0$ and $\mathcal{D}_1$, we prepare a uniform superposition over the variable 'a'. Here, we want the number of possible values of 'a' to be exponential (if 'a' were only a few bits, the proof argument would no longer work well). Also, if padding the variable 'a' resulted in two or more compression-function calls inside Merkle-Damgård, then additional registers would be needed

for the argument, and the proof would become more complicated. To take these circumstances into account while keeping the discussion/assumption as simple as possible, we set $|a| = m/2$ and imposed the condition $|\mathsf{pad}(x)| = m$ for $|x| = m/2$.

Second, in Step 7(b) of $\mathcal{D}_1$, we query $\mathcal{O}_c$ on the value $\mathsf{pad}(a)||0^{m/2}$, and this value $\mathsf{pad}(a)||0^{m/2}$ has length $3m/2$. Here again, when evaluating this value by the Merkle-Damgård construction, we want to keep the number of compression-function calls as small as possible. This is the reason for imposing the condition $|\mathsf{pad}(x)| = 2m$ for $|x| = 3m/2$. (If the number of compression-function calls became larger here, then, in order to maintain consistency, the number of registers and oracle queries around Steps 5-6 would also have to increase, which would make the proof more cumbersome. We could use a bit string of another length (say, $m/3$) instead of $0^{m/2}$, but we chose $0^{m/2}$ because the practical padding functions that satisfy $|\mathsf{pad}(x)| = m$ for $|x| = m/2$ usually also satisfy $|\mathsf{pad}(x)| = 2m$ for $|x| = 3m/2$.)

That said, this assumption was chosen somewhat arbitrarily and is not essential, and we believe it is possible that some other assumptions could also be used. In particular, the assumption is used to keep the proof's presentation simple, rather than because the attack fundamentally relies on those exact message lengths.

C Proof of Lemma 3

Proof (of Lemma 3). Firstly, let Good_1 (resp. Good_2) be the event that measuring X, Z, T_1 (resp., Z, W, F) of the state $|\mathsf{E}_1^{\mathsf{StO}}; 2\rangle$ in the computational basis yields an outcome of the form $|\mathsf{pad}(a)\|IV\rangle_X |a\rangle_Z |0\rangle_{T_1}$ for some a (resp., $|a\rangle_Z |f(a)\rangle_W |f\rangle_F$ for some a and f).

Register T_1 is not modified in Steps 3–6 of D_1. Hence, even if we measure it immediately after Step 2 (instead of in Step 7), the probability of obtaining outcome 1 is $p_1^{T_1}$. Since T_1 is flipped iff $X \neq \mathsf{pad}(Z)\|IV$, we have $\Pr[\neg\mathsf{Good}_1] = p_1^{T_1}$. Moreover, registers Z, W, F are not modified in Steps 2–6 (under the standard-oracle implementation of RO), and therefore $\Pr[\neg\mathsf{Good}_2] = p_1^{c\neq\gamma}$. Thus, $\Pr[\mathsf{Good}_1 \wedge \mathsf{Good}_2] \geq 1 - \Pr[\neg\mathsf{Good}_1] - \Pr[\neg\mathsf{Good}_2] = 1 - p_1^{T_1} - p_1^{c\neq\gamma} \geq 1 - 2\varepsilon$. □

D Proof of Lemma 4

Proof (of Lemma 4). In Step 5 of $\mathcal{D}_0$, the Hadamard operator $H^{\otimes m/2}$ is applied to register Z, and then all the registers are measured. Let $|\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle$ be the state of $\mathsf{E}_0^{\mathsf{StO}}$ just before the measurement in Step 5 of $\mathcal{D}_0$. Then, since Eq. (14)

holds, we have

$$|\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle = H_Z^{\otimes m/2} |\mathsf{E}_0^{\mathsf{StO}}; 4\rangle$$

$$\overset{3\sqrt{\varepsilon}}{\approx} \sum_{\substack{a,z\in\{0,1\}^{m/2} \\ f\in\mathsf{Funcs}}} \frac{(-1)^{a\cdot z}}{2^{m/2}\sqrt{N}} |0^{m+2n}\rangle_{XY} |z\rangle_Z |0^n\rangle_W |\zeta_{f,a}\rangle_S \otimes |f\rangle_F |0,0\rangle_{T_1T_2}. \tag{50}$$

In addition, when measuring register Z of $|\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle$, the probability of obtaining $z = 0^{m/2}$ is at least $1 - p_0 \geq 1 - \varepsilon$. Hence,

$$\begin{aligned} 1 - \varepsilon &\leq \Pr\left[0^{m/2} \text{ is obtained when measuring } Z \text{ of } |\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle\right] \\ &= \left\| \Pi_{Z=0} |\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle \right\|^2 \end{aligned} \tag{51}$$

follows, where $\Pi_{Z=0}$ is the projector onto the space such that register Z is $|0^{m/2}\rangle_Z$. This implies that

$$\begin{aligned} \|\Pi_{Z=0} |(50)\rangle\| &\geq \underbrace{\left\|\Pi_{Z=0} |\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle\right\|}_{\geq\sqrt{1-\varepsilon} \text{ by Eq. (51)}} - \underbrace{\left\|\Pi_{Z=0}\left(|(50)\rangle - |\mathsf{E}_0^{\mathsf{StO}}; \text{b.m.}\rangle\right)\right\|}_{\leq 3\sqrt{\varepsilon}} \\ &\geq \sqrt{1-\varepsilon} - 3\sqrt{\varepsilon} \end{aligned} \tag{52}$$

In addition, some straightforward calculation shows

$$\langle(16)|(14)\rangle = \langle(17)|(15)\rangle = \sum_{\substack{a,b\in\{0,1\}^{m/2} \\ f\in\mathsf{Funcs}}} \frac{1}{N}\frac{1}{2^m} \langle\zeta_{f,b}|\zeta_{f,a}\rangle_S = \|\Pi_{Z=0} |(50)\rangle\|^2. \tag{53}$$

From Eq. (52) and (53), it follows that both $\langle(16)|(14)\rangle$ and $\langle(17)|(15)\rangle$ are non-negative real numbers, and

$$\langle(16)|(14)\rangle = \langle(17)|(15)\rangle \geq \left(\sqrt{1-\varepsilon} - 3\sqrt{\varepsilon}\right)^2 \geq 1 - 6\sqrt{\varepsilon} \tag{54}$$

(note that $\sqrt{1-\varepsilon} - 3\sqrt{\varepsilon} > 0$ due to the assumption that $\varepsilon \leq 2^{-20}$). Therefore, we have

$$\begin{aligned} \left\||\mathsf{E}_0^{\mathsf{StO}}; 4\rangle - |(16)\rangle\right\| &\leq \left\||\mathsf{E}_0^{\mathsf{StO}}; 4\rangle - |(14)\rangle\right\| + \||(14)\rangle - |(16)\rangle\| \\ &\leq 3\sqrt{\varepsilon} + \sqrt{\| \, |(14)\rangle \, \|^2 + \| \, |(16)\rangle \, \|^2 - 2\mathrm{Re}\, \langle(16)|(14)\rangle} \\ &\overset{(*)}{\leq} 3\sqrt{\varepsilon} + \sqrt{1 + 1 - 2(1 - 6\sqrt{\varepsilon})} \\ &\leq 4\sqrt[4]{\varepsilon} + 3\sqrt{\varepsilon}, \end{aligned} \tag{55}$$

where $(*)$ follows from Eq. (54). We can show $\left\||\mathsf{E}_1^{\mathsf{StO}}; 4\rangle - |(17)\rangle\right\| \leq 4\sqrt[4]{\varepsilon} + 3\sqrt{\varepsilon}$ in the same way. □

E Proof of Lemma 5

Proof (of Lemma 5). Applying Lemma 2 to $|(24)\rangle$, we obtain

$$\begin{aligned}
&\mathsf{cO}_{ZWF}\,|(24)\rangle \\
&\quad = |(26)\rangle \\
&\quad + \frac{1}{2^{m/4+n/2}} \sum_{\substack{a\in\{0,1\}^{m/2}, D\in\mathsf{DBs},\\ |D|\le q, D(a)=\bot,\\ D(\mathsf{pad}(a)||0^{m/2})=\bot}} |0^{m+2n}\rangle_{XY}\,|a\rangle_Z\,|\hat{0}\rangle_W \\
&\qquad\qquad\qquad\qquad |\zeta_D\rangle_S\,|D\rangle_F\,|0,0\rangle_{T_1T_2} \qquad (56) \\
&\quad - \frac{1}{2^{m/4+n}} \sum_{\substack{a\in\{0,1\}^{m/2}, c\in\{0,1\}^n,\\ D\in\mathsf{DBs},\\ |D|\le q, D(a)=\bot,\\ D(\mathsf{pad}(a)||0^{m/2})=\bot}} |0^{m+2n}\rangle_{XY}\,|a\rangle_Z\,|\hat{0}\rangle_W \\
&\qquad\qquad\qquad\qquad |\zeta_D\rangle_S\,|D[a\mapsto c]\rangle_F\,|0,0\rangle_{T_1T_2}\,. \qquad (57)
\end{aligned}$$

Since $\|\,|(24)\rangle\,\| \le 1$ holds (by Eq. (25)), some straightforward calculation shows that

$$\|\,|(56)\rangle\,\| = \frac{1}{2^{n/2}}\|\,|(24)\rangle\,\| \le \frac{1}{2^{n/2}}$$

and

$$\|\,|(57)\rangle\,\|^2 = \frac{1}{2^{2n}} \sum_{c\in\{0,1\}^n} \|\,|(24)\rangle\,\|^2 \le \frac{1}{2^n}$$

hold. Therefore, we have $\|\mathsf{cO}_{ZWF}\,|(24)\rangle - |(26)\rangle\| \le 2/2^{n/2}$. □

F Proof of Lemma 6

Proof (of Lemma 6). Eq. (34) immediately follows from the definitions of $\Pi_{\neg\mathsf{hit}}$ and Π_{good}. (If a query is made from $|x'\rangle_{X'}$ to cO with $x' \neq a$ and $x' \neq \mathsf{pad}(a)||0^{m/2}$, then cO does not change the registers $F[a]$ and $F[\mathsf{pad}(a)||0^{m/2}]$. In addition, a query to cO increases the size of database at most 1.)

Next, we show Eq. (35). In what follows, we omit to write the T_1 and T_2 registers. They are always $|0,0\rangle_{T_1T_2}$ in this proof.

Let c be an arbitrary n-bit string, and $|\phi_c\rangle$ be a (non-zero) vector of the form of Eq. (33) with databases of size at most j. Then, there exist some vectors[24] $|\eta_{c,D,x'}\rangle_{Y'XYS'}$ such that

$$|\phi_c\rangle = \sum_{\substack{a\in\{0,1\}^{m/2}, D\in\mathsf{DBs},\\ x'\in\{0,1\}^{\le\ell}\\ |D|<j, D(a)=\bot\\ D(\mathsf{pad}(a)||0^{m/2})=\bot}} \frac{1}{2^{m/4}}\,|a\rangle_Z\,|c\rangle_W\,|D[a\mapsto c]\rangle_F\,|x'\rangle_{X'}\,|\eta_{c,D,x'}\rangle_{Y'XYS'}\,. \qquad (58)$$

[24] Here, $|\eta_{c,D}\rangle_{XYS}$ in Eq. (33) corresponds to $\sum_{x'}|x'\rangle_{X'}\,|\eta_{c,D,x'}\rangle_{Y'XYS'}$.

We have

$$\left\||\phi_c\rangle\right\|^2 = \sum_{\substack{a\in\{0,1\}^{m/2}, D\in\mathsf{DBs}\\ x'\in\{0,1\}^{\le\ell}\\ |D|<j, D(a)=\bot\\ D(\mathsf{pad}(a)||0^{m/2})=\bot}} \frac{1}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2$$

$$= \sum_{\substack{x'\in\{0,1\}^{\le\ell}, D\in\mathsf{DBs}\\ |D|<j}} \frac{1}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2 \times \#\left\{a : D(a)=\bot \text{ and } D(\mathsf{pad}(a)||0^{m/2})=\bot\right\}$$

$$\ge (2^{m/2}-2j)\cdot \sum_{\substack{x'\in\{0,1\}^{\le\ell}, D\in\mathsf{DBs}\\ |D|<j}} \frac{1}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2 \tag{59}$$

and

$$\begin{aligned}
&\left\|\Pi_{\mathsf{hit}}|\phi_c\rangle\right\|^2\\
&= \sum_{\substack{a\in\{0,1\}^{m/2}, D\in\mathsf{DBs},\\ x'\in\{0,1\}^{\le\ell}\\ |D|<j, D(a)=\bot\\ D(\mathsf{pad}(a)||0^{m/2})=\bot\\ x'=a \text{ or } x'=\mathsf{pad}(a)||0^{m/2}}} \frac{1}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2\\
&= \sum_{\substack{x'\in\{0,1\}^{\le\ell}, D\in\mathsf{DBs}\\ |D|<j}} \frac{1}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2\\
&\qquad \times \#\left\{a : D(a)=D(\mathsf{pad}(a)||0^{m/2})=\bot \text{ and } (x'=a \text{ or } x'=\mathsf{pad}(a)||0^{m/2})\right\}\\
&\le \sum_{\substack{x'\in\{0,1\}^{\le\ell}, D\in\mathsf{DBs}\\ |D|<j}} \frac{2}{2^{m/2}} \left\||\eta_{c,D,x'}\rangle_{Y'XYS'}\right\|^2,
\end{aligned} \tag{60}$$

and thus

$$\frac{\left\|\Pi_{\mathsf{hit}}|\phi_c\rangle\right\|}{\left\||\phi_c\rangle\right\|} \le \frac{2}{\sqrt{2^{m/2}-2j}} \le \frac{2\sqrt{2}}{\sqrt{2^{m/2}}} \tag{61}$$

(recall that $j \le 2^{m/2-2}$ is assumed).

Any non-zero vector $|\phi\rangle$ in the image of $\Pi_{\mathsf{good}\wedge|D|\le j}$ is represented as a linear combination $|\phi\rangle = \sum_c |\phi_c\rangle$ for some $|\phi_c\rangle$ like above. For such $|\phi\rangle$,

$$\frac{\left\|\Pi_{\mathsf{hit}}|\phi\rangle\right\|}{\left\||\phi\rangle\right\|} \le \frac{2\sqrt{2}}{\sqrt{2^{m/2}}} \tag{62}$$

follows from Eq. (61), because $|\phi_c\rangle$ and $\Pi_{\mathsf{hit}}|\phi_c\rangle$ are orthogonal to $|\phi_{c'}\rangle$ and $\Pi_{\mathsf{hit}}|\phi_{c'}\rangle$, respectively, for $c \neq c'$. Therefore,

$$\begin{aligned}\left\|\Pi_{\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j}\right\|_{\mathrm{op}} &= \max_{|\phi_*\rangle\neq 0}\frac{\left\|\Pi_{\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le j}|\phi_*\rangle\right\|}{\left\|\Pi_{\mathsf{good}\wedge|D|\le j}|\phi_*\rangle\right\|}\\ &= \max_{|\phi\rangle\in\mathrm{Im}(\Pi_{\mathsf{good}\wedge|D|\le j})\setminus\{0\}}\frac{\|\Pi_{\mathsf{hit}}|\phi\rangle\|}{\||\phi\rangle\|} \le \frac{2\sqrt{2}}{\sqrt{2^{m/2}}}\end{aligned}$$

holds, and thus Eq. (35) follows. □

G Proof of Lemma 7

Proof (of Lemma 7). For each $i = 0, \ldots, q$, we have

$$\begin{aligned}&\mathsf{cO}_{X'Y'F}U_i\Pi_{\mathsf{good}\wedge|D|\le q+i+1}\\ &\overset{(*)}{=} \mathsf{cO}_{X'Y'F}\Pi_{\mathsf{good}\wedge|D|\le q+i+1}U_i\Pi_{\mathsf{good}\wedge|D|\le q+i+1}\\ &\overset{2\sqrt{2}/\sqrt{2^{m/2}}}{\approx} \mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le q+i+1}U_i\Pi_{\mathsf{good}\wedge|D|\le q+i+1}\\ &\overset{(**)}{=} \Pi_{\mathsf{good}\wedge|D|\le q+i+2}\mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}\Pi_{\mathsf{good}\wedge|D|\le q+i+1}U_i\Pi_{\mathsf{good}\wedge|D|\le q+i+1}\\ &\overset{(*)}{=} \Pi_{\mathsf{good}\wedge|D|\le q+i+2}\mathsf{cO}_{X'Y'F}\Pi_{\neg\mathsf{hit}}U_i\Pi_{\mathsf{good}\wedge|D|\le q+i+1}\\ &= \Pi_{\mathsf{good}\wedge|D|\le q+i+2}\mathsf{cO}_{X'Y'F}U_i', \qquad (63)\end{aligned}$$

where $(*)$ holds because[25] U_i acts only on register S, the approximation follows from Eq. (35) of Lemma 6, and $(**)$ follows from Eq. (34) of Lemma 6 (with $j := q+i+1$)[26], and the last equality follows from the definition of U_i'. Thus,

$$\begin{aligned}&U_{\mathcal{S}}\Pi_{\mathsf{good}\wedge|D|\le q+1}\\ &= U_q\mathsf{cO}_{X'Y'F}\cdots\mathsf{cO}_{X'Y'F}U_1\underbrace{\mathsf{cO}_{X'Y'F}U_0\Pi_{\mathsf{good}\wedge|D|\le q+1}}_{\overset{2\sqrt{2}/2^{m/4}}{\approx}\Pi_{\mathsf{good}\wedge|D|\le q+2}\mathsf{cO}_{X'Y'F}U_0' \text{ by Eq. (63)}}\\ &\overset{2\sqrt{2}/2^{m/4}}{\approx} U_q\mathsf{cO}_{X'Y'F}\cdots\underbrace{\mathsf{cO}_{X'Y'F}U_1\Pi_{\mathsf{good}\wedge|D|\le q+2}}_{\overset{2\sqrt{2}/2^{m/4}}{\approx}\Pi_{\mathsf{good}\wedge|D|\le q+3}\mathsf{cO}_{X'Y'F}U_1' \text{ by Eq. (63)}}\mathsf{cO}_{X'Y'F}U_0'\\ &\overset{2\sqrt{2}/2^{m/4}}{\approx} \cdots\\ &\overset{2\sqrt{2}/2^{m/4}}{\approx} \Pi_{\mathsf{good}\wedge|D|\le 2q+2}\underbrace{U_q'\cdot\mathsf{cO}_{X'Y'F}U_{q-1}'\cdots\mathsf{cO}_{X'Y'F}\cdot U_0'}_{=U_{\mathcal{S}}'}\\ &= \Pi_{\mathsf{good}\wedge|D|\le 2q+2}U_{\mathcal{S}}'.\end{aligned}$$

[25] More precisely: Since Π_{good} leaves S register unrestricted, the subspace $\mathrm{Im}(\Pi_{\mathsf{good}})$ is not changed by any unitary U_i acting solely on S. Therefore $U_i\Pi_{\mathsf{good}} = \Pi_{\mathsf{good}}U_i\Pi_{\mathsf{good}}$.

[26] Note that $2q+1 \le 2^{m/2-2}$ holds because $q \le 2^{m/2-4}$ is assumed.

Hence, the claim of the lemma holds. □

H Figures

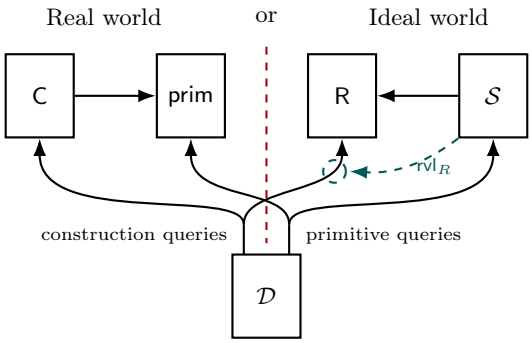


Fig. 1. Illustration of the definition of indifferentiability notions. The goal of $\mathcal{D}$ is to distinguish the real world and the ideal world. In public indifferentiability, the simulator can see the queries to R through the reveal oracle rvl_R (this does not appear in the original indifferentiability and sequential indifferentiability). In sequential indifferentiability, $\mathcal{D}$ can make construction queries only after finishing all primitive queries.

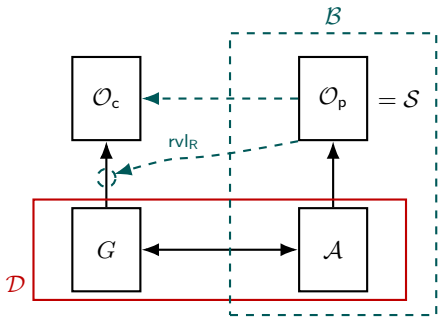


Fig. 2. Illustration of the composition theorem. When the construction oracle $\mathcal{O}_\mathsf{c}$ is the ideally random object R, the primitive oracle $\mathcal{O}_\mathsf{p}$ is replaced with the simulator $\mathcal{S}$, which can see the construction queries made by $\mathcal{D}$ or in the game G, through the reveal oracle.