Quantum Physical Unclonable Functions: Possibilities and Impossibilities

Quantum Physical Unclonable Functions: Possibilities and Impossibilities

Myrto Arapinis[1], Mahshid Delavar[1], Mina Doosti[1], and Elham Kashefi[1,2]

[1]School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK
[2]Departement Informatique et Reseaux, CNRS, Sorbonne Université, 4 Place Jussieu 75252 Paris CEDEX 05, France

A Physical Unclonable Function (PUF) is a device with unique behaviour that is hard to clone hence providing a secure fingerprint. A variety of PUF structures and PUF-based applications have been explored theoretically as well as being implemented in practical settings. Recently, the inherent unclonability of quantum states has been exploited to derive the quantum analogue of PUF as well as new proposals for the implementation of PUF. We present the first comprehensive study of quantum Physical Unclonable Functions (qPUFs) with quantum cryptographic tools. We formally define qPUFs, encapsulating all requirements of classical PUFs as well as introducing a new testability feature inherent to the quantum setting only. We use a quantum game-based framework to define different levels of security for qPUFs: quantum exponential unforgeability, quantum existential unforgeability and quantum selective unforgeability. We introduce a new quantum attack technique based on the universal quantum emulator algorithm of Marvin and Lloyd to prove no qPUF can provide quantum existential unforgeability. On the other hand, we prove that a large family of qPUFs (called unitary PUFs) can provide quantum selective unforgeability which is the desired level of security for most PUF-based applications.

1 Introduction

Canetti and Fischlin's result on the impossibility of achieving secure cryptographic protocols without any setup assumptions [9] has motivated a rich line of research investigating the advantages of making hardware assumptions in protocol design. The idea was first introduced by Katz in [27], and attracted the attention of researchers and developers as it adopts physical assumptions and eliminates the need to trust a designated party or to rely on computational assumptions. Among different hardware assumptions, Physical Unclonable Functions (PUFs) have greatly impacted the field [4].

PUFs are hardware structures designed to utilize the random physical disorder which appear in any physical device during the manufacturing process. Because of the uncontrollable nature of these random disorders, building a clone of the device is considered impractical. The behaviour of a PUF is usually equivalent to a set of Challenge-Response

Mina Doosti: m.doosti@sms.ed.ac.uk, This work has been presented at QCrypt 2019 (9th International Conference on Quantum Cryptography)

Pairs (CRPs) which are extracted through physically querying the PUF and measuring its responses. The PUF's responses depend on its physical features and are assumed to be unpredictable, i.e. even the manufacturer of the PUF, with access to many CRPs, cannot predict the response to a new challenge [42]. This property makes PUFs different from other hardware tokens in the sense that the manufacturer of a hardware token is completely aware of the behaviour of the token they have built [7].

So far, the cryptographic literature has mainly considered what we will call classical PUFs (or cPUFs) restricted to classical CRPs. Most cPUFs generate only a finite, albeit possibly exponential (in some desired security parameters), number of CRPs [11]. However, most of them remain vulnerable against different attacks like side-channel [50, 11] and machine-learning [19, 44, 43, 28]. Thus, considering the importance of cPUFs as a hardware security primitive in several real-world applications, on one hand, [11, 26, 15, 2, 33, 30, 36][1] and the recent advances in quantum technology, on the other hand, it is worth investigating whether quantum technologies could boost the security of cPUFs or if they, on the contrary, threaten their security. In the current work, we address the general and formal treatment of PUFs in a quantum world for the first time by defining quantum PUFs (qPUFs) as a quantum token that can be challenged with quantum states and respond with quantum states. We identify the requirements a qPUF needs to meet to provide the main security property required for most of the qPUF-based applications, that is *unforgeability*[2]. All prior similar works [45, 46, 39, 54] (see related work paragraph below) considered the special case of qPUFs where the encoding of the responses is known to the manufacturer and in fact, the evaluation of the qPUF is public information. We provide a general and formal mathematical framework for the study of qPUFs as a new quantum primitive inspired from the theoretical literature of classical PUF while taking into account full capabilities of a quantum adversary. However, it is worth mentioning that designing and implementing concrete qPUFs satisfying our proposed level of security set up remains a challenging task that we are exploring separately as a follow up of this work.

1.0.1 Our Contributions.

We first define qPUFs as quantum channels and formalize the standard requirements of robustness, uniqueness and collision-resistance for qPUFs guided by the classical counterparts to establish the requirements that qPUFs should satisfy to enable their usage as a cryptographic primitive. We then use the game-based framework to define three security notions for qPUFs: quantum exponential unforgeability, quantum existential unforgeability and quantum selective unforgeability capturing the strongest type of attack models where the adversary has access to the qPUF and can query it with his chosen quantum states. In this new model, we demonstrate how quantum learning techniques, such as the universal quantum emulator algorithm of [34], can lead to successful attacks. In doing so we establish several possibility and impossibility results.

- *No qPUF provides Quantum Exponential Unforgeability.* The presented attack is the correct analogue of the brute-force attack for classical PUFs.
- *No qPUF provides Quantum Existential Unforgeability.* We show how the universal

[1] Recently SAMSUNG announced that in their new processor Exynox 9820 they have integrated SRAM based PUF to store and manage personal data in perfect isolation. Also, a UK company, Quantum Base, has started to mass-produce its patented optical quantum PUFs.

[2] *Unpredictability* and *unclonability* are other equivalent terms for this notion used often in the literature.

quantum emulator algorithm (which is polynomial in the size of the qPUF's dimension) can break this security property of any qPUFs.

- *Any qPUF provides Quantum Selective Unforgeability.* In other words, no QPT adversary can, on average, generate the response of a qPUF to random challenges.

1.0.2 Other Related Works

The concept of Physical Unclonable Functions was first introduced by Pappu *et al.* [41] in 2001, devising the first implementation of an Optical PUF. Optical PUFs were subsequently improved as to generating an independent number of CRPs [35]. Several structures of Physical Unclonable Functions were further introduced including Arbiter PUFs [20], Ring-Oscillator based PUFs [49, 16] and SRAM PUFs [24]. For a comprehensive overview of existing PUF structures, we refer the reader to [32, 25].

Recently, the concept of "quantum read-out of PUF (QR-PUF)" was introduced in [45] to exploit the no-cloning feature of quantum states to potentially solve the spoofing problem in the remote device identification. The QR-PUF-based identification protocol has been implemented in [22]. In addition to the security analysis of this protocol against intercept-resend attack in [45], its security has also been analysed against other special types of attacks targeting extracting information from an unknown challenge state [47, 53]. In another work, [39], the continuous variable encoding is exploited to implement another practical QR-PUF based identification protocol. The security of this protocol has also been analysed only against an attacker who aims to efficiently estimate or clone an unknown challenge quantum state [38, 18]. Moreover, some other applications of QR-PUFs have been introduced in [48] and [51].

In another independent recent work, Gianfelici et al. have presented a common theoretical framework for both cPUFs and QR-PUFs [21]. They quantitatively characterize the PUF properties, particularly robustness and unclonability. They also introduce a generic PUF-based identification scheme and parameterize its security based on the values obtained from the experimental implementation of PUF.

2 Quantum Emulation Algorithm

In this section, we describe the Quantum Emulation (QE) algorithm presented in [34] as a quantum process learning tool that can outperform the existing approaches based on quantum tomography [14]. The main idea behind quantum emulation comes from the question on the possibility of emulating the action of an unknown unitary transformation on an unknown input quantum state by having some of the input-output samples of the unitary. An emulator is not trying to completely recreate the transformation or simulate the same dynamics. Instead, it outputs the action of the transformation on a quantum state. The original algorithm was developed and proposed in the context of quantum process tomography, thus the analysis did not consider any adversarial behaviour. For our cryptanalysis purposes, we need to provide a new fidelity analysis for challenges not fully lying within the subspace of the learning phase. We further optimise the success probability of our attack by optimising the choice of the reference state.

2.1 The Circuit and Description

The circuit of the quantum emulation algorithm is depicted in Figure 1 also in [34] and works as follows: Let U be a unitary transformation on a D-dimensional Hilbert space $\mathcal{H}^D$,

$S_{in} = \{|\phi_i\rangle ; i = 1, ..., K\}$ be a sample of input states and $S_{out} = \{|\phi_i^{out}\rangle ; i = 1, ..., K\}$ the set of corresponding outputs, i.e $|\phi_i^{out}\rangle = \mathrm{U}\,|\phi_i\rangle$. Also, let d be the dimension of the Hilbert space $\mathcal{H}^d$ spanned by S_{in} and $|\psi\rangle$, a challenge state. The goal of the algorithm is to find the output of U on $|\psi\rangle$, that is $\mathrm{U}\,|\psi\rangle$.

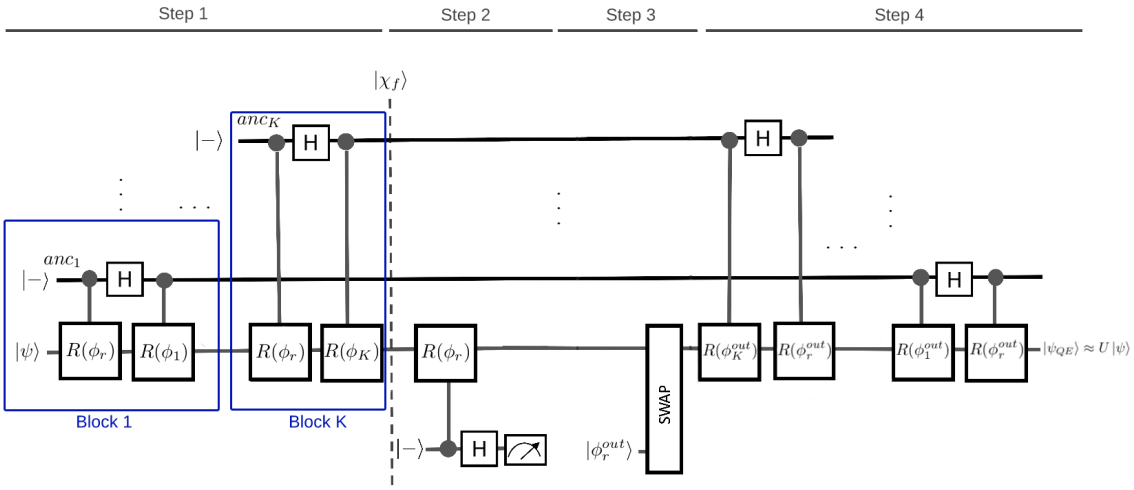

Figure 1: The circuit of the quantum emulation algorithm. $|\phi_r\rangle$ is the reference state and $|\phi_r^{out}\rangle$ is the output of the reference state. $R(*)$ gates are controlled-reflection gates. In each block of Step 1, a reflection around the reference and another sample state is being performed.

The main building blocks of the algorithm are controlled-reflection gates described as:

$$R_c(\phi) = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes e^{i\pi|\phi\rangle\langle\phi|} \tag{1}$$

A controlled-reflection gate acts as the identity ($\mathbb{I}$) if the control qubit is $|0\rangle$, and as $R(\phi) = e^{i\pi|\phi\rangle\langle\phi|} = \mathbb{I} - 2\,|\phi\rangle\langle\phi|$ if the control qubit is $|1\rangle$. The circuit also uses Hadamard and SWAP gates and consists of four stages.

Stage 1. K number of sample states and a specific number of ancillary qubits are chosen and used through the algorithm. We assume the algorithm uses all of the states in S_{in}. The ancillary systems are all qubits prepared at $|-\rangle$. Let $|\phi_r\rangle \in S_{in}$ be considered as the reference state. This state can be chosen at random or according to a special distribution. The first step consists of $K-1$ blocks wherein each block, the following gates run on the state of the system and an ancilla:

$$W(i) = R_c(\phi_i)HR_c(\phi_r). \tag{2}$$

In each block represented by equation (2), a controlled-reflection around the reference state $|\phi_r\rangle$ is performed on $|\psi\rangle$ with the control qubit being on the $|-\rangle$ ancillary state. Then a Hadamard gate (H) runs on the ancilla followed by another controlled-reflection around the sample state $|\phi_i\rangle$. This is repeated for each of the K states in S_{in} such that the input state is being entangled with the ancillas and also it is being projected into the subspace $\mathcal{H}^d$ in a way that the information of $|\psi\rangle$ is encoded in the coefficients of the general entangled state. This information is the overlap of $|\psi\rangle$ with all the sample inputs. By reflecting around the reference state in each block, the main state is pushed to $|\phi_r\rangle$ and the probability of finding the system at the reference state increases. The overall state of the circuit after Stage 1 is:

$$[W(K)...W(1)]\,|\psi\rangle\,|-\rangle^{\otimes K} \approx |\phi_r\rangle\,|\Omega(anc)\rangle \tag{3}$$

where $|\Omega(anc)\rangle$ is the entangled state of K ancillary qubits. The approximation comes from the fact that the state is not only projected on the reference quantum state but it is also projected on other sample quantum states with some probability. We present a more precise formula in the next subsection.

Stage 2. In this stage, first a reflection around $|\phi_r\rangle$ is performed and after applying a Hadamard gate on an extra ancilla, that ancilla is measured in the computational basis $\{|0\rangle, |1\rangle\}$. Based on the output of the measurement, one can decide whether the first step was successful (i.e. the output of the measurement is 0) or not. If the first step is successful, the main state has been pushed to the reference state. In this case, the algorithm proceeds with Stage 3. If the output is 1, the projection was unsuccessful and the input state remains almost unchanged. In this case, either the algorithm aborts or it goes back to the first stage and picks a new state as the reference. This stage has a post-selection role which can be skipped to output a mixed state of two possible outputs.

Stage 3. The main state is swapped with $|\phi_r^{out}\rangle = \mathrm{U}|\phi_r\rangle$ that is the output of the reference state. This is done by means of a SWAP gate. At this point, the overall state of the system is:

$$(\mathrm{SWAP} \otimes I^{\otimes K})|\phi_r^{out}\rangle|\phi_r\rangle|\Omega(anc)\rangle = |\phi_r\rangle|\phi_r^{out}\rangle|\Omega(anc)\rangle. \tag{4}$$

By tracing out the first qubit, the state of the system becomes $|\phi_r^{out}\rangle|\Omega(anc)\rangle$.

Stage 4. The last stage is very similar to the first one except that all blocks are run in reverse order and the reflection gates are made from corresponding output quantum states. The action of stage 4 is equivalent to:

$$W^{out}(i) = R_c(\phi_i^{out})HR_c(\phi_r^{out}) = (\mathrm{U} \otimes I)W(i)(\mathrm{U}^\dagger \otimes \mathbb{I}). \tag{5}$$

After repeating this gate for all the output samples, U is acted on the projected components of $|\psi\rangle$ and by restoring back the information of $|\psi\rangle$ from the ancilla, the input state approaches $\mathrm{U}|\psi\rangle$. The overall output state of the circuit at the end of this stage is:

$$[W^{out}(1)...W^{out}(K)]|\phi_r^{out}\rangle|\Omega(anc)\rangle \approx \mathrm{U}|\psi\rangle|-\rangle^{\otimes K} \tag{6}$$

where equality is obtained whenever the success probability of Stage 2 is equal to 1.

2.2 Output fidelity analysis

We are interested in the fidelity of the output state $|\psi_{QE}\rangle$ of the algorithm and the intended output $\mathrm{U}|\psi\rangle$ to estimate the success. In the original paper, the fidelity analysis is first provided for ideal controlled-reflection gates and later a protocol is presented to implement them efficiently. In this paper, as we are more interested in the theoretical bounds for the fidelity, all the gates including the controlled-reflection gates are assumed to be ideal keeping in mind that the implementation is possible [34, 31]. We recall the main theorem of [34]:

Theorem 1 *[34] Let $\mathcal{E}_\mathrm{U}$ be the quantum channel that describes the overall effect of the algorithm presented above. Then for any input state ρ, the Uhlmann fidelity of $\mathcal{E}_\mathrm{U}(\rho)$ and the desired state $\mathrm{U}\rho\mathrm{U}^\dagger$ satisfies:*

$$F(\rho_{QE}, \mathrm{U}\rho\mathrm{U}^\dagger) \geq F(\mathcal{E}_\mathrm{U}(\rho), \mathrm{U}\rho\mathrm{U}^\dagger) \geq \sqrt{P_{succ-stage1}} \tag{7}$$

where $\rho_{QE} = |\psi_{QE}\rangle\langle\psi_{QE}|$ is the main output state(tracing out the ancillas) when the post-selection in Stage 2 has been performed. $\mathcal{E}_{\mathrm{U}}(\rho)$ is the output of the whole circuit without the post-selection measurement in Stage 2 and $P_{succ-stage1}$ is the success probability of Stage 1.

For the purpose of this paper, we need a more precise and concrete expression for the output fidelity not covered in [34]. From the proof of Theorem 1 in [34], it can be seen that the success probability of Stage 1 is calculated as follows:

$$P_{succ-stage1} = |\langle\phi_r| Tr_{anc}(|\chi_f\rangle\langle\chi_f|)|\phi_r\rangle|^2 \tag{8}$$

where $|\chi_f\rangle$ is the final state of the circuit after Stage 1 and $Tr_{anc}(\cdot)$ computes the reduced density matrix by tracing out the ancillas. The overlap of the resulting state and the reference state equals the success probability of Stage 1. Now relying on Theorem 1, we only use equation (8) for our analysis henceforward.

The fidelity of the output state of the circuit highly depends on the choice of the reference state (equation (8)) such that it may increase or decrease the success probability of the adversary in different security models as we will discuss in the Section 3. We establish the following recursive relation for the state of the circuit after the i-th block of Stage 1, in terms of the previous state:

$$|\chi_i\rangle = \frac{1}{2}[(I - R(\phi_r))|\chi_{i-1}\rangle|0\rangle + R(\phi_i)(\mathbb{I} + R(\phi_r))|\chi_{i-1}\rangle|1\rangle]. \tag{9}$$

Now by using this relation, we can prove the following theorem. The proof can be found in Appendix B

Theorem 2 *Let $|\chi_K\rangle$ be the output state of K-th block of the circuit (Figure 1). Let $|\psi\rangle$ be the input state of the circuit, $|\phi_r\rangle$ the reference state and $|\phi_i\rangle$ other sample states. We have:*

$$\begin{aligned}|\chi_K\rangle = {} & \langle\phi_r|\psi\rangle|\phi_r\rangle|0\rangle^{\otimes K} + |\psi\rangle|1\rangle^{\otimes K} - \langle\phi_r|\psi\rangle|\phi_r\rangle|1\rangle^{\otimes K} \\ & + \sum_{i=1}^{K}\sum_{j=0}^{i}[f_{ij}2^{l_{ij}}|\langle\phi_r|\psi\rangle|^{x_{ij}}|\langle\phi_i|\psi\rangle|^{y_{ij}}|\langle\phi_r|\phi_i\rangle|^{z_{ij}}]|\phi_r\rangle|q_{anc}(i,j)\rangle \\ & + \sum_{i=1}^{K}\sum_{j=0}^{i}[g_{ij}2^{l'_{ij}}|\langle\phi_r|\psi\rangle|^{x'_{ij}}|\langle\phi_i|\psi\rangle|^{y'_{ij}}|\langle\phi_r|\phi_i\rangle|^{z'_{ij}}]|\phi_i\rangle|q'_{anc}(i,j)\rangle\end{aligned} \tag{10}$$

where l_{ij}, x_{ij}, y_{ij}, z_{ij}, l'_{ij}, x'_{ij}, y'_{ij} and z'_{ij} are integer values indicating the power of the terms of the coefficient. Note that f_{ij} and g_{ij} can be 0, 1 or -1 and $q_{anc}(i,j)$ and $q'_{anc}(i,j)$ output a computational basis of K qubits (other than $|0\rangle^{\otimes K}$).

Having a precise expression for $|\chi_f\rangle$ from Theorem 2, one can calculate $P_{succ-step1}$ of equation (8) by tracing out all the ancillary systems from the density matrix of $|\chi_f\rangle\langle\chi_f|$. Also, now it is clear that if $|\psi\rangle$ is orthogonal to the $\mathcal{H}^d$, the only term remaining in equation (10) is $|\psi\rangle|1\rangle^{\otimes K}$. So, the input state remains unchanged after the first stage and $P_{succ-step1} = 0$.

For states projected in the subspace spanned by S_{in}, the overall channel describing the quantum emulation algorithm has always a fixed point inside the subspace [34]. Hence, Stage 1 is successful with probability close to 1 by assuming the gates to be ideal.

3 Quantum Physical Unclonable Functions

We consider a set of quantum devices that have been created through the same manufacturing process. These devices respond with a general quantum state when challenged with a quantum state. Similar to the classical setting (see Appendix A), we formalize the manufacturing process of qPUFs by defining a QGen algorithm:

$$\text{qPUF}_{\mathbf{id}} \leftarrow \text{QGen}(\lambda) \tag{11}$$

where **id** is the identifier of $\text{qPUF}_{\mathbf{id}}$ and λ the security parameter.

We also need to define the QEval algorithm mapping any input quantum state $\rho_{in} \in \mathcal{H}^{d_{in}}$ to an output quantum state $\rho_{out} \in \mathcal{H}^{d_{out}}$ where $\mathcal{H}^{d_{in}}$ and $\mathcal{H}^{d_{out}}$ are the domain and range Hilbert spaces of $\text{qPUF}_{\mathbf{id}}$, denoted as:

$$\rho_{out} \leftarrow \text{QEval}(\text{qPUF}_{\mathbf{id}}, \rho_{in}). \tag{12}$$

For now, we allow for the most general form of trace-preserving quantum maps, i.e. CPT maps for QEval. So, we have:

$$\rho_{out} = \Lambda_{\mathbf{id}}(\rho_{in}) \tag{13}$$

Apart from these common algorithms (that are analogue to the classical setting), we also require qPUFs to include an efficient test algorithm $\mathcal{T}$ as we will formally define in Definition 4 to test the equality between two unknown quantum states. We will also need the concept of quantum state distinguishability, which can be defined with different quantum distance measures such as trace distance or fidelity. Here we use the fidelity-based definition as follows: Let $F(\cdot,\cdot)$ denote the fidelity, and μ and ν the distinguishability and indistinguishability threshold parameters respectively such that $0 \leq \mu, \nu \leq 1$. We say two quantum states ρ and σ are μ-distinguishable if $0 \leq F(\rho,\sigma) \leq 1-\mu$ and ν-indistinguishable if $\nu \leq F(\rho,\sigma) \leq 1$. Finally, we can define a Quantum Physical Unclonable Function as follows.

Definition 1 (Quantum Physical Unclonable Function) *Let λ be the security parameter, and $\delta_r, \delta_u, \delta_c \in [0,1]$ the robustness, uniqueness and collision resistance thresholds. A $(\lambda, \delta_r, \delta_u, \delta_c)$-qPUF includes the algorithms:* QGen, QEval *and $\mathcal{T}$ satisfying Requirements 1, 2, and 3 defined below:*

Requirement 1 (δ_r-Robustness) *For any* $\text{qPUF}_{\mathbf{id}}$ *generated through* $\text{QGen}(\lambda)$ *and evaluated using* QEval *on any two input states ρ_{in} and σ_{in} that are δ_r-indistinguishable, the corresponding output quantum states ρ_{out} and σ_{out} are also δ_r-indistinguishable with overwhelming probability,*

$$\Pr[\delta_r \leq F(\rho_{out}, \sigma_{out}) \leq 1] = 1 - negl(\lambda). \tag{14}$$

Requirement 2 (δ_u-Uniqueness) *For any two* qPUF*s generated by the* QGen *algorithm, i.e.* $\text{qPUF}_{\mathbf{id}_i}$ *and* $\text{qPUF}_{\mathbf{id}_j}$*, the corresponding CPT map models, i.e. Λ_{id_i} and Λ_{id_j} are δ_u-distinguishable with overwhelming probability,*

$$\Pr[\ \| \ (\Lambda_{id_i} - \Lambda_{id_j})_{i \neq j} \ \|_{\diamond} \geq \delta_u \ ] = 1 - negl(\lambda). \tag{15}$$

Requirement 3 (δ_c-Collision-Resistance (Strong)) *For any* $\text{qPUF}_{\mathbf{id}}$ *generated by* $\text{QGen}(\lambda)$ *and evaluated by* QEval *on any two input states ρ_{in} and σ_{in} that are δ_c-distinguishable,*

the corresponding output states ρ_{out} and σ_{out} are also δ_c-distinguishable with overwhelming probability,[3]

$$\Pr[0 \leq F(\rho_{out}, \sigma_{out}) \leq 1 - \delta_c] = 1 - negl(\lambda). \tag{16}$$

In qPUF-based applications such as device authentication (or identification), it is necessary that there be a clear distinction between different qPUF instances generated by the same QGen algorithm running on the same parameters λ [3]. To this end, the following conditions should be satisfied: $\delta_c \leq 1 - \delta_r$ and $\delta_u \leq 1 - \delta_r$. So, we can drop δ_u and δ_c from the notation and characterize the qPUF as $(\lambda, \delta_r) - \text{qPUF}$.

We also need to mention that, δ_r and δ_c parameters can allow for some specific noise models for each PUF device. More specifically, the collision resistance parameter i.e. δ_c or the ratio of δ_c^o/δ_c^i is directly related to the channel parameters of the qPUF evaluation. Although, as the collision-resistance is an important requirement for achieving a secure PUF, similar to classical PUFs, we choose the strong collision-resistance as the main requirement for the quantum PUF. We specify that the strong collision-resistance parameter can allow for noisy PUF evaluation under the coherent noise models. Such noise models preserve distances between the input and output states of the qPUF and this property makes them suitable candidates for quantum PUF. Also, it has been shown in [23] that a general noise can be modelled as a combination of coherent and incoherent noises. Hence only the class of noise model with an almost close to zero incoherent factor can be considered to satisfy the δ_c (strong) collision resistance. Hence for the rest of this work, aiming to formalise the first general security framework, we consider a noiseless setting and leave further investigation that would be linked to particular construction to future works.

We have initially allowed for any CPT map as QEval algorithm. Now, we let the QEval algorithm be a CPT map with the same dimension of domain and range Hilbert space, i.e. $d_{in} = d_{out}$. We show that under this assumption, only unitary transformations and CPT maps that are negligibly close to unitary, can simultaneously provide the (strong)collision-resistance and robustness requirements of qPUFs.

Theorem 3 *Let $\mathcal{E}(\rho)$ be a completely positive and trace-preserving (CPT) map described as follows:*

$$\mathcal{E}(\rho) = (1 - \epsilon) U \rho U^\dagger + \epsilon \tilde{\mathcal{E}}(\rho) \tag{17}$$

where U is a unitary transformation, $\tilde{\mathcal{E}}$ is an arbitrary (non-negligibly) contractive channel and $0 \leq \epsilon \leq 1$. Then $\mathcal{E}(\rho)$ is a $(\lambda, \delta_r, \delta_c)$-qPUF for any λ, δ_r, and δ_c and with the same dimension of domain and range Hilbert space, if and only if $\epsilon = negl(\lambda)$.

Proof: First, we note that The contractive property of trace-preserving operations [37] states that CPT maps on the same Hilbert space, can only preserve or decrease distances thus we have:

$$F(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \geq F(\rho, \sigma) \tag{18}$$

Thus the robustness is generally satisfied. As a result, the proof of the theorem reduces to proving for collision-resistance. Let ρ and δ be two δ_c-distinguishable challenge with fidelity $F(\rho, \sigma) \leq 1 - \delta_c$. Again with the above argument the fidelity of the outputs cannot

[3] A weaker variant of Collision-Resistance, with separate input/output bound can be also defined in a similar fashion where the responses generated by QEval on any two δ_c^i-distinguishable input states ρ_{in} and σ_{in}, should be at least δ_c^o-distinguishable. In fact, if $\delta_c^i = \delta_c^o = \delta_c$ we call the requirement a strong collision-resistance. Note that this equality holds up to a negligible value in the security parameter, i.e. if $\delta_c^i = \delta_c^o \pm negl(\lambda)$, the strong collision-resistance requirement has still been satisfied. If $\delta_c^o < \delta_c^i$ (the difference is non-negligible) then this is referred to as weak collision-resistance.

be smaller than $F(\rho,\sigma)$. Thus the δ_c requirement is satisfied if the fidelity of the response density matrices are equal up to a negligible value.

Now let $\rho_1 = U\rho U^\dagger$, $\sigma_1 = U\sigma U^\dagger$, $\rho_2 = \tilde{\mathcal{E}}(\rho)$, and $\sigma_2 = \tilde{\mathcal{E}}(\sigma)$. We use the joint concavity of the fidelity [37] to obtain the following relation for the channel's output fidelity:

$$\begin{aligned} F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) &= F((1-\epsilon)\rho_1 + \epsilon\rho_2, (1-\epsilon)\sigma_1 + \epsilon\sigma_2) \\ &\geq (1-\epsilon)F(\rho_1,\sigma_1) + \epsilon F(\rho_2,\sigma_2) \end{aligned} \tag{19}$$

Since the first part of the channel is unitary which is distance preserving, we have $F(\rho_1,\sigma_1) = F(\rho,\sigma)$. Also due to contractive property of trace-preserving operations we know that $F(\rho_2,\sigma_2) \geq F(\rho,\sigma)$. We have

$$F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) - F(\rho,\sigma) \geq \epsilon(F(\rho_2,\sigma_2) - F(\rho,\sigma)) \tag{20}$$

Now since the channel $\tilde{\mathcal{E}}$ is non-negligibly contractive, the value $F(\rho_2,\sigma_2) - F(\rho,\sigma)$ is not necessarily negligible and in order for the LHS of 19 to be always negligible, ϵ has to be negligible. So we have proved that CPT maps of the form 17 can be δ_c collision resistance qPUFs only if $\epsilon = negl(\lambda)$.

Now we show that all channels of the form of Equation 17 where ϵ is negligible satisfy the strong collision resistance property up to a negligible value. To show that we use the relation between fidelity and trace distance which we denote as $\mathcal{D}_{tr}$, which is $\mathcal{D}_{tr}(\rho,\sigma) \leq \sqrt{1-F(\rho,\sigma)}$. We use this inequality to relate the distance between the states $\mathcal{E}(\rho)$ and $\mathcal{E}(\sigma)$ and the original distance between ρ and σ and we subtract both sides to get the following inequality:

$$\begin{aligned} F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) - F(\rho,\sigma) &\leq \mathcal{D}_{tr}^2(\rho,\sigma) - \mathcal{D}_{tr}^2(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \\ &\leq (\mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma)))(\mathcal{D}_{tr}(\rho,\sigma) + \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma))) \\ &\leq 2(\mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma))) \end{aligned} \tag{21}$$

In Appendix C, Lemma 2 we show that the difference between the trace distance of the input and output for channels described as Equation 17, is bounded by $\epsilon\mathcal{D}_{tr}(\rho,\sigma)$. Thus we have:

$$F(\mathcal{E}(\rho),\mathcal{E}(\sigma)) - F(\rho,\sigma) \leq 2\epsilon\mathcal{D}_{tr}(\rho,\sigma) \tag{22}$$

Now since $\epsilon = negl(\lambda)$ and $0 \leq \mathcal{D}_{tr}(\rho,\sigma) \leq 1$, we can conclude that the difference between the fidelity is also negligible and hence the δ_c collision-resistance is satisfied up to a negligible value, and the proof is complete. □

The above theorem shows that only unitary or more generally, *ϵ-disturbed unitary maps* where ϵ is small, are suitable candidates for qPUF, especially when strong collision resistance is required. Thus, in the rest of the paper, we choose the QEval algorithm to be a unitary map, and also for simplicity, we establish some of our theorems with pure quantum states, noting that considering the mixed states would not affect the main results. We call this type of qPUFs, Unitary qPUFs (or simply UqPUFs) and formally define them in Definition 3. Nevertheless, we believe studying more general non-unitary qPUFs will be interesting future research directions in this field.

Moreover, we require UqPUF transformations to be initially unknown (or exponentially hard to recover) as we will formally define in Definition 2. This is a hardware assumption that is also considered in the classical setting where the PUF behaviour is unknown even for the manufacturer [42]. Although from a construction point of view, this may not seem an easily achievable requirement, from a practical point of view this assumption is reasonable

considering limited fabrication capabilities or the fact that simulating the same unitary on a quantum computer is not technologically easy due to noise or accumulated errors in each gate, even when the structure of the unitary is known. Moreover, there are promising constructions such as the family of optical schemes implemented using crystals or optical scattering media [39], where usually even the manufacturer does not know the underlying unitary unless querying it. On the other hand, in gate-based construction, one cannot avoid the fact that the manufacturer knows the underlying unitary. Hence this type of constructions cannot provide security against an adversarial manufacturer. Nevertheless, if predicting the evolution of a quantum state is difficult this is enough for security under the usual PUF assumptions. Hence such devices are still useful and practical for many applications as they can still provide security against any malicious adversary other than the manufacturer. We also note that from the theoretical point of view, this requirement is a minimal and pre-challenge requirement that can be achieved by sampling a family of unitaries indistinguishable from the Haar family of unitary transformations in single-shot, and we believe there are efficient ways to do this sampling [13, 1]. Finally, our framework and results cover both adversarial models where the manufacturer could be trusted or not.

Definition 2 (Unknown Unitary Transformation) *We say a family of unitary transformations U^u, over a D-dimensional Hilbert space $\mathcal{H}^D$ is called Unknown Unitaries, if for all QPT adversaries $\mathcal{A}$ the probability of estimating the output of U^u on any randomly picked state $|\psi\rangle \in \mathcal{H}^D$ is at most negligibly higher than the probability of estimating the output of a Haar random unitary operator on that state:*

$$|\Pr_{U\leftarrow U^u}[F(\mathcal{A}(|\psi\rangle), U|\psi\rangle) \geq \textit{non-negl}(\lambda)] - \Pr_{U_\mu\leftarrow\mu}[F(\mathcal{A}(|\psi\rangle), U_\mu|\psi\rangle) \geq \textit{non-negl}(\lambda)]| = \textit{negl}(\lambda). \tag{23}$$

where μ denotes the Haar measure and the average probability has been taken over al; the states $|\psi\rangle$.

Note that UqPUFs also satisfy a natural notion of unclonability, known as no-cloning of unitary transformation [12] which states that two black-box unitary transformations $\mathcal{O}_1$ and $\mathcal{O}_2$ cannot be perfectly cloned by a single-use apart from the trivial cases of perfect distinguishability or when $\mathcal{O}_1 = \mathcal{O}_2$. Thus, two UqPUFs, as long as they correspond to different unitaries, which is satisfied by the uniqueness requirement, are unclonable by quantum mechanics through a single-use. In the following section, we then show how this unclonability property can be extended to the case where the transformation has been used multiple times by formally introducing the notion of unforgeability. Thus, we define the unitary qPUFs as follows.

Definition 3 (Unitary qPUF (UqPUF)) *A Unitary* qPUF $((\lambda, \delta_r) - \text{UqPUF})$ *is a* $(\lambda, \delta_r) -$ qPUF *where the* QEval *algorithm is modelled by an unknown unitary transformation* $\mathrm{U}_{\mathbf{id}}$ *over a D-dimensional Hilbert space, $\mathcal{H}^D$ operating on pure input quantum states $|\psi_{in}\rangle \in \mathcal{H}^D$ and returning pure output quantum states $|\psi_{out}\rangle \in \mathcal{H}^D$,*

$$|\psi_{out}\rangle = \text{QEval}(\text{UqPUF}_{\mathbf{id}}, |\psi_{in}\rangle) = \mathrm{U}_{\mathbf{id}}|\psi_{in}\rangle\,. \tag{24}$$

As a result of the distance-preserving property of UqPUFs, we drop δ_r from the notation and simply characterise UqPUF as λ-UqPUFs.

3.1 Security notion for qPUFs

The security of most PUF-based applications such as PUF-based identification protocols relies on the unforgeability of PUFs [3]. Informally, unforgeability means that given a

subset of challenge-response pairs of the target PUF, the probability of correctly guessing a new challenge-response pair shall be negligible in terms of the security parameter. In this section, we formally define this security notion for qPUFs in a game-based framework which is a standard framework for defining security of cryptographic primitives and analysing their security [3, 5, 15].

Accordingly, we define unforgeability as a game between an adversary who represents the malicious party and a challenger who plays the role of the honest party. The game is run in four steps: Setup, Learning, Challenge and Guess.

In the *setup phase*, the necessary public and private parameters and functions are shared between the adversary and the challenger.

The *learning phase* models the amount of knowledge that the adversary can get from the challenger. Similar to [3], we consider chosen-input attacks modelling an adversary that has access to the qPUF and can query it with his own chosen inputs from the domain Hilbert space. Because of the quantum nature of the adversary's queries, the adversary has to prepare two copies of each query, keep one in his database and send the other one to the challenger.

The *challenge phase* captures the intended security notion. We consider here two types of challenge phase: Existential and Selective. In an existential challenge phase, the adversary chooses the challenge state while in a selective one, the challenge state is chosen by the challenger. We characterize a "new" existential challenge by imposing the adversary to choose a state that is μ-distinguishable from all the inputs queries in the learning phase. In the selective case, to ensure the adversary has no knowledge about the challenge, we impose the challenger to choose the challenge uniformly at random from the domain Hilbert space.

Finally, in the *guess phase*, the adversary outputs his guess of the response corresponding to the challenge chosen in the challenge phase. The challenger checks the equality between the adversary's guess and the correct response with a test algorithm. The adversary wins the game if the output of the test algorithm is 1. Due to the impossibility of perfectly distinguishing all quantum states, checking equality of two completely unknown states is a non-trivial task. This is one of the major differences between classical and quantum PUFs. Nevertheless, a probabilistic comparison of unknown quantum states can be achieved through the simple quantum SWAP test algorithm [8], and its generalisation to multiple copies introduced recently in [10]. Here we abstract from specific tests and define necessary conditions for a general quantum test.

Definition 4 (Quantum Testing Algorithm) *Let $\rho^{\otimes\kappa_1}$ and $\sigma^{\otimes\kappa_2}$ be κ_1 and κ_2 copies of two quantum states ρ and σ, respectively. A Quantum Testing algorithm $\mathcal{T}$ is a quantum algorithm that takes as input the tuple ($\rho^{\otimes\kappa_1}$,$\sigma^{\otimes\kappa_2}$) and accepts ρ and σ as equal (outputs 1) with the following probability*

$$\Pr[1 \leftarrow \mathcal{T}(\rho^{\otimes\kappa_1}, \sigma^{\otimes\kappa_2})] = 1 - \Pr[0 \leftarrow \mathcal{T}(\rho^{\otimes\kappa_1}, \sigma^{\otimes\kappa_2})] = f(\kappa_1, \kappa_2, F(\rho, \sigma))$$

where $F(\rho, \sigma)$ is the fidelity of the two states and $f(\kappa_1, \kappa_2, F(\rho, \sigma))$ satisfies the following limits:

$$\begin{cases} \lim_{F(\rho,\sigma)\to 1} f(\kappa_1, \kappa_2, F(\rho, \sigma)) = 1 \quad \forall\, (\kappa_1, \kappa_2) \\ \lim_{\kappa_1,\kappa_2\to\infty} f(\kappa_1, \kappa_2, F(\rho, \sigma)) = F(\rho, \sigma) \\ \lim_{F(\rho,\sigma)\to 0} f(\kappa_1, \kappa_2, F(\rho, \sigma)) = Err(\kappa_1, \kappa_2) \end{cases} \tag{25}$$

with $Err(\kappa_1, \kappa_2)$ characterising the error of the test algorithm and $F(\rho, \sigma)$ the fidelity of the states.

We also define another abstraction of the test algorithm in an ideal case which later helps us to demonstrate the security of the UqPUF. We formalize the ideal test $\mathcal{T}_\delta^{ideal}$ as follows:

Definition 5 ($\mathcal{T}_\delta^{ideal}$ Test Algorithm) *We call a test algorithm according to Definition 4, a $\mathcal{T}_\delta^{ideal}$ Test Algorithm when for any two state $|\psi\rangle$ and $|\phi\rangle$ the test responds as follows:*

$$\mathcal{T}_\delta^{ideal} = \begin{cases} 1 & F(|\psi\rangle, |\phi\rangle) \geq \delta \\ 0 & otherwise \end{cases} \tag{26}$$

Now we are ready to formalize unforgeability through a formal security game.

Game 1 (Formal game-based security of qPUF) *Let* $\mathrm{qPUF} = (\mathrm{QGen}, \mathrm{QEval}, \mathcal{T})$ *and* $\mathcal{T}$ *be defined as Definition 1 and 4, respectively. We define the following game* $\mathcal{G}_{c,\mu}^{\mathrm{qPUF}}(\mathcal{A}, \lambda)$ *running between an adversary* $\mathcal{A}$ *and a challenger* $\mathcal{C}$*:*

Setup. *The challenger* $\mathcal{C}$ *runs* $\mathrm{QGen}(\lambda)$ *to build an instance of the qPUF family,* $\mathrm{qPUF}_{\mathbf{id}}$*. Then,* $\mathcal{C}$ *reveals to the adversary* $\mathcal{A}$*, the domain and range Hilbert space of* $\mathrm{qPUF}_{\mathbf{id}}$ *respectively denoted by* $\mathcal{H}_{in}$ *and* $\mathcal{H}_{out}$ *as well as the identifier of* $\mathrm{qPUF}_{\mathbf{id}}$*,* $\mathbf{id}$*. The challenger initialises two empty databases,* S_{in} *and* S_{out} *and shares them with the adversary* $\mathcal{A}$*. Also* $\mathcal{H}_{in}^d$ *denotes adversary's input subspace.*

Learning. *For* $i = 1 : k$

- $\mathcal{A}$ *prepares two copies of a quantum state* $\rho_i \in \mathcal{H}_{in}^d$*, appends one to* S_{in} *and sends the other to* $\mathcal{C}$*;*
- $\mathcal{C}$ *runs* $\mathrm{QEval}(\mathrm{qPUF}_{\mathbf{id}}, \rho_i)$ *and sends* ρ_i^{out}*, to* $\mathcal{A}$*;*
- $\mathcal{A}$ *appends* ρ_i^{out} *to* S_{out}*.*

Challenge.[4]

- *If* $c = \mathsf{qEx}$*:* $\mathcal{A}$ *picks a quantum state* $\rho^* \in \mathcal{H}_{in}^d$ *at least* μ*-distinguishable from all the states in* S_{in} *and sends* κ_1 *copies of it to* $\mathcal{C}$*;*
- *If* $c = \mathsf{qSel}$*:* $\mathcal{C}$ *chooses a quantum state* ρ^* *at random from the uniform distribution over the Hilbert space* $\mathcal{H}_{in}^d$*. The challenger keeps* κ_1 *copies of* ρ^* *and sends an extra copy of* ρ^* *to* $\mathcal{A}$*.*

Guess.

- $\mathcal{A}$ *sends* κ_2 *copies of his guess* ρ' *to* $\mathcal{C}$*;*
- $\mathcal{C}$ *runs* $\mathrm{QEval}(\mathrm{qPUF}_{\mathbf{id}}, \rho^*)^{\otimes\kappa_1}$*, and gets* $\rho_{out}^{*\otimes\kappa_1}$*;*
- $\mathcal{C}$ *runs the test algorithm* $b \leftarrow \mathcal{T}(\rho_{out}^{*\otimes\kappa_1}, \rho'^{\otimes\kappa_2})$ *where* $b \in \{0, 1\}$ *and outputs* b*. The adversary wins the game if* $b = 1$*.*[5]

[4]The parameter c specifies the type of the challenge phase.

[5]Note that all the learning phase queries and the challenges represented with $\rho, \rho', |\phi\rangle$, etc. are considered to be any general separable or entangled state of a D-dimensional Hilbert space. Moreover, κ_1 and κ_2 are a choice of notation that enables us to include any desired quantum test algorithm according to Definition 4 and are independent of the number of the copies that the adversary uses in the learning phase.

Based on the above game, we define the security notions, *quantum exponential unforgeability*, *quantum existential unforgeability* and *quantum selective unforgeability* for qPUFs; where the first one, models unforgeability of qPUFs against exponential adversaries with unlimited access to the qPUF in the learning phase; the second one is the most common and strongest type of unforgeability against Quantum Polynomial-Time (QPT) adversaries; finally the third one is a weaker notion of unforgeability that is sufficient for most qPUF-based applications like qPUF-based identification protocols.

Definition 6 (Quantum Exponential Unforgeability) *A* qPUF *provides quantum exponential unforgeability if the success probability of any* exponential *adversary $\mathcal{A}$ in winning the game $\mathcal{G}^{\text{qPUF}}_{\mathsf{qEx},\mu}(\mathcal{A},\lambda)$ is negligible in λ*

$$Pr[1 \leftarrow \mathcal{G}^{\text{qPUF}}_{\mathsf{qEx},\mu}(\mathcal{A},\lambda)] = negl(\lambda) \tag{27}$$

Definition 7 (μ-Quantum Existential Unforgeability) *A* qPUF *provides μ-quantum existential unforgeability if the success probability of any Quantum Polynomial-Time (QPT) adversary $\mathcal{A}$ in winning the game $\mathcal{G}^{\text{qPUF}}_{\mathsf{qEx},\mu}(\mathcal{A},\lambda)$ is negligible in λ*

$$Pr[1 \leftarrow \mathcal{G}^{\text{qPUF}}_{\mathsf{qEx},\mu}(\mathcal{A},\lambda)] = negl(\lambda) \tag{28}$$

Definition 8 (Quantum Selective Unforgeability) *A* qPUF *provides quantum selective unforgeability if the success probability of any Quantum Polynomial-Time (QPT) $\mathcal{A}$ in winning the game $\mathcal{G}^{\text{qPUF}}_{\mathsf{qSel}}(\lambda,\mathcal{A})$ is negligible in λ*

$$Pr[1 \leftarrow \mathcal{G}^{\text{qPUF}}_{\mathsf{qSel}}(\lambda,\mathcal{A})] = negl(\lambda) \tag{29}$$

3.2 Security analysis of Unitary qPUFs

Here, we show which security notions defined in Section 4.1 can be achieved by unitary qPUFs (UqPUFs) over a D-dimensional Hilbert space operating on pure quantum states.

In the classical setting, cPUFs can be fully described by the finite set of CRPs, and this suffices for breaking unforgeability. More precisely, an unbounded or exponential adversary can extract the entire set of CRPs by querying the target cPUF with all possible challenges [11]. If the challenges are n-bit strings, the number of possible challenges is 2^n. However, in the quantum setting, a UqPUF can generate an infinite number of quantum challenge-response pairs such that extracting all of them is hard, even for exponential adversaries. This, combined with limitations imposed by quantum mechanics such as no-cloning [52] and the limits on state estimation [6], raise the question if UqPUFs could satisfy unforgeability against exponential adversaries. We now prove that no UqPUF provides quantum exponential unforgeability as defined in Definition 6.

Theorem 4 ***(No UqPUF provides quantum exponential unforgeability)*** *For any λ-UqPUF and any $0 \leq \mu \leq 1$, there exists an exponential quantum adversary $\mathcal{A}$ such that*

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda,\mathcal{A})] = non\text{-}negl(\lambda) \tag{30}$$

Proof: The key idea of the proof is based on complexity analysis of unitary tomography and implementation of a general unitary by single and double qubit gates, since for an exponential quantum adversary, it will be feasible to extract the unitary matrix by tomography and then build the extracted unitary by general gate decomposition method. By using the Solovay-Kitaev theorem [37], we then show that the adversary can build the

unitary matrix of the UqPUF performing on n-qubits, within an arbitrarily small distance ϵ using $O(n^2 4^n \log^c(n^2 4^n))$ gates and hence win the game with any test algorithm $\mathcal{T}$. Let $\text{UqPUF}_{\mathbf{id}}$ operate on n-qubit input-output pairs where $n = \log(D)$. In the learning phase, $\mathcal{A}$ selects a complete set of orthonormal basis of $\mathcal{H}^D$ denoted as $\{|b_i\rangle\}_{i=1}^{2^n}$ and queries $\text{UqPUF}_{\mathbf{id}}$ with each base 2^n times. So, the total number of queries in the learning phase is $k_1 = 2^{2n}$.

Then, $\mathcal{A}$ runs a *unitary tomography* algorithm to extract the mathematical description of the unknown unitary transformation corresponding to the $\text{UqPUF}_{\mathbf{id}}$, say $\text{U}_{\mathbf{id}}$. It has been shown in [37] that the complexity of this algorithm is $\mathcal{O}(2^{2n})$ for n-qubit input-output pairs. This is feasible for an exponential adversary. It is clear that once the mathematical description of the unitary is extracted, $\mathcal{A}$ can simply calculate the response of the unitary to a known challenge quantum state and wins the game $\mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})$ for any value of μ. So, we have:

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})] = 1. \tag{31}$$

We can also show the exponential adversary wins even the weaker notion of the security, i.e. quantum selective unforgeability, where he has only one copy of the challenge quantum state. To win the game with the selective challenge phase, the adversary needs to implement the unitary.

It is known that any unitary transformation over $\mathcal{H}^{2^n}$ requires $O(2^{2n})$ two-level unitary operations or $O(n^2 2^{2n})$ single qubit and CNOT gates [37] to be implemented. However, according to Solovay-Kitaev theorem [37], to implement a unitary with an accuracy ϵ using any circuit consisting of m single qubit and CNOT gates, $O(m \log^c(m/c))$ gates from the discrete set are required where c is a constant approximately equal to 2. Thus, an arbitrary unitary performing on n-qubit can be approximately implemented within an arbitrarily small distance ϵ using $O(n^2 4^n \log^c(n^2 4^n))$ gates.

So, $\mathcal{A}$ implements the unitary $\text{U}'_{\mathbf{id}}$ with error ϵ. Let $\mathcal{A}$ get the challenge state $|\psi\rangle$ in the qSel Challenge phase. The adversary queries $\text{U}'_{\mathbf{id}}$ with $|\psi\rangle$ and gets $|\omega\rangle = \text{U}'_{\mathbf{id}}|\psi\rangle$ as output. Since the ϵ can be arbitrary small, then $F(\text{U}_{\mathbf{id}}|\psi\rangle, \text{U}'_{\mathbf{id}}|\psi\rangle) \geq 1 - negl(\lambda)$. So, $\mathcal{A}$'s output $|\omega\rangle$ passes any test algorithm $\mathcal{T}(|\psi^{out}\rangle^{\otimes\kappa_1}, |\omega\rangle^{\otimes\kappa_2})$ with probability close to 1. Again, an unbounded adversary wins the game $\mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel},\mu}(\lambda, \mathcal{A})$ with probability 1. □

We note that this result is expected as any qPUF (same as a classical PUF), can in principle, be simulated with enough computational resources. That is why the reasonable and achievable security model is usually against a qPUF in hands of the adversary for a limited time or limited query such as QPT adversaries. It is also worth mentioning that from an engineering point of view, limiting the adversary to a certain number of queries on a hardware level, can depend on the construction and it might be possible in some qPUF implementations, while might not be feasible with some others. While this is an interesting problem to be considered in qPUF implementations, from a cryptanalysis point, our security analysis against a quantum adversary who is given polynomial time in the security parameter, is independent of the construction.

Exploiting the quantum emulation algorithm introduced in Section 2 we now turn to quantum existential unforgeability, and show that no UqPUF provides quantum existential unforgeability for any $\mu \neq 1$ as defined in Definition 7. Note that the case $\mu = 1$ corresponds to the existential challenge state being orthogonal to all the queried states in the learning phase. With $\mu = 1$, the adversary is prevented from taking advantage of its quantum access to the qPUF to win the game.

Theorem 5 ***(No UqPUF provides quantum existential unforgeability)*** *For any* λ-

UqPUF, and $0 \leq \mu \leq 1 - non\text{-}negl(\lambda)$*, there exits a QPT adversary* $\mathcal{A}$ *such that*

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})] = non\text{-}negl(\lambda). \tag{32}$$

Proof: We show there is a QPT adversary $\mathcal{A}$ who wins the game $\mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})$ with non-negligible probability in λ. The adversary $\mathcal{A}$ runs the learning phase of the game $\mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})$ with $|\phi_1\rangle$ and $|\phi_2\rangle$ such that $|\phi_1\rangle$ can be any quantum state in $\mathcal{H}^D$ and

$$|\phi_2\rangle = \begin{cases} \frac{1}{\sqrt{2}}(|\phi_1\rangle + |\phi_3\rangle) & if\ 0 \leq \mu \leq \frac{1}{2} \\ \sqrt{\mu}\,|\phi_1\rangle + \sqrt{1-\mu}\,|\phi_3\rangle & if\ \frac{1}{2} < \mu \leq 1 - non\text{-}negl(\lambda) \end{cases} \tag{33}$$

Without loss of the generality, we assume $\mathcal{A}$ chooses one of the computational basis of $\mathcal{H}^D$ as $|\phi_1\rangle$. Then, $\mathcal{A}$ chooses an orthogonal state to $|\phi_1\rangle$ as $|\phi_3\rangle$ and sets $|\phi_2\rangle$ the superposition of these two states. In the existential challenge phase, $\mathcal{A}$ sets $|\phi_3\rangle$ as his chosen challenge. Note that $|\phi_3\rangle$ satisfies the μ-distinguishability of the challenge state with both $|\phi_1\rangle$ and $|\phi_2\rangle$. In the guess phase, to estimate the output of UqPUF to $|\phi_3\rangle$, the adversary $\mathcal{A}$ runs the quantum emulation (QE) algorithm defined in Section 2 with the reference state $|\phi_r\rangle = |\phi_2\rangle$.

Relying on Theorem 2, the output state of Stage 1 of the QE algorithm is:

$$\begin{aligned} |\chi_f\rangle = &\langle\phi_2|\,\phi_3\rangle\,|\phi_2\rangle\,|0\rangle + |\phi_3\rangle\,|1\rangle - \langle\phi_2|\,\phi_3\rangle\,|\phi_2\rangle\,|1\rangle \\ &- 2\,\langle\phi_1|\,\phi_3\rangle\,|\phi_1\rangle\,|1\rangle + 2\,\langle\phi_2|\,\phi_3\rangle\,\langle\phi_2|\,\phi_1\rangle\,|\phi_1\rangle\,|1\rangle\,. \end{aligned} \tag{34}$$

Note that $\langle\phi_1|\,\phi_3\rangle = 0$ and we set $\langle\phi_2|\,\phi_3\rangle = \alpha$ and $\langle\phi_2|\,\phi_1\rangle = \beta$ based on the choice of $|\phi_2\rangle$, the above equation can be simplified as:

$$|\chi_f\rangle = \alpha\,|\phi_2\rangle\,|0\rangle + |\phi_3\rangle\,|1\rangle - \alpha\,|\phi_2\rangle\,|1\rangle + 2\alpha\beta\,|\phi_1\rangle\,|1\rangle\,. \tag{35}$$

Now, according to Theorem 1, the final fidelity in terms of the success probability of Stage 1 can be obtained by calculating the density matrix of $|\chi_f\rangle$ and tracing out the ancillas:

$$\begin{aligned} P_{succ-stage1} &= |\,\langle\phi_2|\,Tr_{anc}(|\chi_f\rangle\,\langle\chi_f|)\,|\phi_2\rangle\,|^2 \\ &= |\alpha^2(1 + 4\alpha^2\beta^2)|^2. \end{aligned} \tag{36}$$

We have different choices for the reference state depending on the distinguishability parameter μ. For cases where the adversary is allowed to produce a new state with at least overlap half with all the states in the learning phase, by choosing the uniform superposition of the states where $\alpha = \beta = \frac{1}{\sqrt{2}}$, the output fidelity will be:

$$F(|\phi_3^{out'}\rangle\,\langle\phi_3^{out'}|\,, |\phi_3^{out}\rangle\,\langle\phi_3^{out}|) \geq \sqrt{P_{succ-stage1}} = 1. \tag{37}$$

where $|\phi_3^{out'}\rangle$ and $|\phi_3^{out}\rangle$ are the output of the QE algorithm and UqPUF to $|\phi_3\rangle$, respectively.

As can be seen, these two states are completely indistinguishable So, the success probability of $\mathcal{A}$ for any test according to Definition 4 is:

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})] = Pr[1 \leftarrow \mathcal{T}(|\psi^{out}\rangle^{\otimes\kappa_1}, |\omega\rangle^{\otimes\kappa_2})] = 1 \tag{38}$$

which is the optimal choice of the reference. On the other hand, for the cases where the adversary is restricted to produce a challenge more than half distinguishable, we can still

create a superposed state with $\alpha = \sqrt{1-\mu}$ and $\beta = \sqrt{\mu}$ and end up with the following fidelity of the emulation by setting $\mu = 1 - non\text{-}negl(\lambda)$

$$\begin{aligned} F(|\phi_3^{out'}\rangle\langle\phi_3^{out'}|, |\phi_3^{out}\rangle\langle\phi_3^{out}|) &\geq |\alpha^2(1+4\alpha^2\beta^2)| \\ &= |(1-\mu)(1+4\mu(1-\mu))| \\ &= non\text{-}negl(\lambda). \end{aligned} \tag{39}$$

Recall that the security parameter λ includes the number of copies used in the test algorithm (κ_1, κ_2), by increasing them the probability of accepting will converge to the above fidelity thus for any $\frac{1}{2} < \mu \leq 1 - non\text{-}negl(\lambda)$:

$$Pr[1 \leftarrow \mathcal{G}^{\mathrm{UqPUF}}_{\mathsf{qEx},\mu}(\lambda, \mathcal{A})] = Pr[1 \leftarrow \mathcal{T}(|\phi_3^{out}\rangle^{\otimes\kappa_1}, |\phi_3^{out'}\rangle^{\otimes\kappa_2})] = non\text{-}negl(\lambda) \tag{40}$$

And the proof is complete. □

This theorem implies that the adversary can always generate the correct response to his chosen challenge provided that he can query it in superposition with other quantum states during the learning phase in terms of the parameter μ. Note that since output quantum states in the learning phase are unknown to the adversary, the more straightforward strategy of superposing the learnt output quantum states cannot be efficiently performed. More precisely, the adversary cannot prepare the precise target superposition of the output states that are completely unknown [40, 17]. Hence the proposed attack is general but non-trivial.

We now further relax the level of security and consider quantum selective unforgeability. We show that any UqPUF can provide this weaker notion of security. Note that in most of the PUF-based applications such as PUF-based identification protocols, selective unforgeability is sufficient.

We need the following lemma to prove the quantum selective unforgeability feature of UqPUFs. The lemma implies the average probability of any state in $\mathcal{H}^D$ to be projected in a subspace $\mathcal{H}^d$ where $d \leq D$. Based on this lemma, we calculate the probability of a state chosen uniformly at random from $\mathcal{H}^D$ to be projected in the orthogonal subspace of the adversary's database where the quantum emulation or similar attacks does not work.

Lemma 1 *Let $\mathcal{H}^D$ be a D-dimensional Hilbert space and $\mathcal{H}^d$ a subspace of $\mathcal{H}^D$ with dimension d. Also, let Π_d be a projector for any quantum state in $\mathcal{H}^D$ into $\mathcal{H}^d$. The average probability that any state, chosen uniformly at random from $\mathcal{H}^D$, $|\psi\rangle \underset{R}{\in} \mathcal{H}^D$ to be projected into $\mathcal{H}^d$ is equal to $\frac{d}{D}$*

$$\Pr_{|\psi\rangle,\Pi_d}[|\langle\psi|\Pi_d|\psi\rangle| = 1] = \frac{d}{D} \tag{41}$$

Proof: The proof is mainly based on the symmetry of the Hilbert space and the fact that the probability of falling into each subspace is equal for any state uniformly picked at random.

Note that Any state $|\psi\rangle \in \mathcal{H}^D$ can be written in terms of the orthonormal bases of $\mathcal{H}^D$ denoted by $|b_i\rangle$, as follows:

$$|\psi\rangle = \sum_{i=0}^{D-1} \alpha_i |b_i\rangle \quad \text{with} \quad \sum_{i=0}^{D-1} |\alpha_i|^2 = 1 \tag{42}$$

where α_i are complex coefficients. A projection into a smaller subspace consists of choosing d bases of $\mathcal{H}^D$ in the form of $\sum_{j=0}^{d-1} |b_j\rangle\langle b_j|$. Without loss of generality, we can assume

$D = md$ where m is an integer. This assumption is always correct for qubit spaces. This means that the larger Hilbert space can be divided into m smaller subspaces each with dimension d. Let $\{|e_i\rangle\}_{i=0}^{d-1}$ be a subset of $\mathcal{H}^D$ which makes a complete set of bases for one of the d-dimensional subspaces. A projector projects $|\psi\rangle$ into one of the subspaces. As $|\psi\rangle$ has been picked at random and the subspaces are symmetric, the probability of falling into each subspace is the same and equal to $\frac{1}{m}$ which is $\frac{d}{D}$. Otherwise either the sum of all probabilities would not be 1 or the $|\psi\rangle$ has not been picked uniformly at random from $\mathcal{H}^D$. This shows that on average the probability of projecting a state ψ is $\frac{d}{D}$. This can also be seen by the fact that the sum of all projectors in a complete set of projectors is equal to one. In this case, we have

$$\sum_{i=0}^{D-1} \Pi_i = \mathbb{I} \tag{43}$$

By sandwiching $|\psi\rangle$ on both sides we have:

$$\sum_{i=0}^{D-1} \langle\psi| \Pi_i |\psi\rangle = 1. \tag{44}$$

Each $\langle\psi| \Pi_i |\psi\rangle$ is itself equal to $\sum_{j=0}^{d-1} |\langle\psi| d_{ij}\rangle|^2$ where $|d_{ij}\rangle$s are the bases associated to the subspace that the projector Π_i projects into. This corresponds to all the permutations of d number of the coefficient $|\alpha_i|^2$ which will be $\frac{1}{d}$ on average. Since we have $\sum_{i=0}^{D-1} \frac{Pr_{\Pi_i}}{d} = 1$, we can conclude that the average probability Pr_Π for all the projectors will be $\frac{d}{D}$ and the proof is complete. □

To establish our possibility result, we first present a preliminary theorem which demonstrates the security of the UqPUF considering an ideal test algorithm which asymptotically satisfies the notion of distance as defined in Definition 5.

Theorem 6 *For any unitary qPUF characterised by* $\text{UqPUF} = (\text{QGen}, \text{QEval}, \mathcal{T}_\delta^{ideal})$*, and any non-zero* δ*, the success probability of any QPT adversary* $\mathcal{A}$ *in the game* $\mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})$ *is bounded as follows:*

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})] \leq \frac{d+1}{D} \tag{45}$$

where D *is the dimension of the Hilbert space that the challenge quantum state is picked from, and* $0 \leq d \leq D-1$ *is the dimension of the largest subspace of* $\mathcal{H}^D$ *that the adversary can span in the learning phase of* $\mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})$*.*

Proof (Sketch): The complete proof can be found in Appendix D, here we only sketch the main idea. We are interested in the average success probability of the adversary running the game $\mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})$. Let the subspace spanned by the learnt queries be a d-dimensional subspace of $\mathcal{H}^D$ denoted by $\mathcal{H}^d$. We calculate the average fidelity of the adversary's estimated output state $|\omega\rangle$ and the correct output $|\psi^{out}\rangle$, over all choices of the qSel challenge state $|\psi\rangle$. We require this fidelity to be greater than a value δ imposed by the $\mathcal{T}_\delta^{ideal}$:

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})] = \underset{|\psi\rangle \in \mathcal{H}^D}{Pr}[F(|\omega\rangle, |\psi^{out}\rangle) \geq \delta]. \tag{46}$$

Note that because of the quantum nature of queries in the learning phase and the limited number of queries that the QPT adversary $\mathcal{A}$ can make, $\mathcal{A}$ might not have the classical description of the responses to his queries. So, we let $\mathcal{A}'$ be another QPT adversary who

has full knowledge of $\mathcal{H}^d$. It is obvious that the success probability of $\mathcal{A}'$ would be higher than the success probability of $\mathcal{A}$ due to the extra knowledge that A' has. So, we have

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A})] \leq Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A}')] \tag{47}$$

In rest of the proof, We calculate the success probability of $\mathcal{A}'$ which is the higher bound for the success probability of $\mathcal{A}$. We write this probability in terms of its partial probabilities for the states orthogonal to $\mathcal{H}^d$ and the rest of the space:

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A}')] = \underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] + \underset{|\psi\rangle \notin \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] Pr[|\psi\rangle \notin \mathcal{H}^{d^\perp}]. \tag{48}$$

The probability of projection into the orthogonal subspace and the conjugate subspace can be obtained by calling Lemma 1:

$$Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] = \frac{d^\perp}{D} \tag{49}$$

where $d^\perp = D - d$; And

$$Pr[|\psi\rangle \notin \mathcal{H}^{d^\perp}] = 1 - Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] = \frac{d}{D} \tag{50}$$

We also assume there exists a QPT algorithm that its average probability over all the states not in the orthogonal subspace to estimate their outputs with $F \geq \delta$ is 1, i.e. $\underset{|\psi\rangle \notin \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] = 1$.

Thus, the only remaining term to calculate is the probability that the average fidelity be greater than δ in the orthogonal subspace, i.e. $\underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta]$. We show in Appendix D that since the qSel challenge is chosen uniformly at random from $\mathcal{H}^D$, the best attack strategy to achieve the desired fidelity is choosing the output state uniformly at random from $\mathcal{H}^D$.

Then, we calculate the average fidelity according to Haar measure and show the average probability for non-zero fidelity is bounded by:

$$\underset{|\psi^{out}\rangle \in \mathcal{H}_{out}^{d^\perp}}{Pr}[F \neq 0] \leq \frac{1}{D - d} \tag{51}$$

So, for non-zero δ we also have,

$$\underset{|\psi^{out}\rangle \in \mathcal{H}_{out}^{d^\perp}}{Pr}[F \geq \delta] \leq \frac{1}{D - d} \tag{52}$$

As a result, the success probability of $\mathcal{A}$ is bounded by

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A})] \leq Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A}')] \leq \frac{d+1}{D} \tag{53}$$

And the theorem is proved. □

Theorem 7 ***(Any UqPUF provides quantum selective unforgeability)*** *Let the test algorithm $\mathcal{T}$ be defined according to Definition 4 and satisfy the condition $Err(\kappa_1, \kappa_2) = negl(\kappa_1, \kappa_2)$. Then, for any* $\mathrm{UqPUF} = (\mathrm{QGen}, \mathrm{QEval}, \mathcal{T})$ *and any QPT adversary, we have:*

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A})] = negl(\lambda). \tag{54}$$

Proof: Let $|\psi\rangle$ be quantum state chosen by the challenger in the selective challenge phase. Also, let $|\psi^{out}\rangle$ and $|\omega\rangle$ be the output of the UqPUF and the adversary $\mathcal{A}$ to $|\psi\rangle$, respectively. Note that the success probability of $\mathcal{A}$ in game $\mathcal{G}_{qSel}^{\text{UqPUF}}(\lambda, \mathcal{A})$ is equal to the probability of the test algorithm in outputting 1:

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})] = Pr[1 \leftarrow \mathcal{T}(|\psi^{out}\rangle^{\otimes\kappa_1}, |\omega\rangle^{\otimes\kappa_2})] \tag{55}$$

We denote $Pr[1 \leftarrow \mathcal{T}(|\omega\rangle^{\otimes\kappa_1}, |\psi^{out}\rangle^{\otimes\kappa_2})]$ with $Pr[1 \leftarrow \mathcal{T}]$ for simplicity. To calculate this probability, we consider two independent cases where leads the $\mathcal{T}$ outputs 1. We let δ be the threshold for $F(|\omega\rangle, |\psi^{out}\rangle)$ that helps us to write the $Pr[1 \leftarrow \mathcal{T}]$ as sum of two terms, i.e. the probability of $\mathcal{T}$ outputting 1 while $F \geq \delta$ and the probability of $\mathcal{T}$ outputting 1 while $F < \delta$:

$$Pr[1 \leftarrow \mathcal{T}] = Pr[1 \leftarrow \mathcal{T}, F \geq \delta] + Pr[1 \leftarrow \mathcal{T}, F < \delta] \tag{56}$$

Let $\delta = negl(\lambda)$ hence we have

$$\begin{aligned} Pr[1 \leftarrow \mathcal{T}] = {} & Pr[1 \leftarrow \mathcal{T} | F \geq negl(\lambda)] Pr[F \geq negl(\lambda)] \\ & + Pr[1 \leftarrow \mathcal{T} | F < negl(\lambda)] Pr[F < negl(\lambda)] \end{aligned} \tag{57}$$

and then from Theorem 6, it can be concluded that

$$Pr[F \geq negl(\lambda)] \leq \frac{d+1}{D} \tag{58}$$

where d is the dimension of the subspace spanned by the learnt queries and D is the dimension of the Hilbert space that the UqPUF is defined over it. Thus, $D = 2^n$ where n is the number of qubits in each input/output state. Since the adversary is a QPT adversary, the number of learnt queries and as a result the value of d should be polynomial in n, i.e. $d = poly(n)$.

Also, according to Definition 4, we have,

$$Pr[1 \leftarrow \mathcal{T} | F < negl(\lambda)] = Err(\kappa_1, \kappa_2) \tag{59}$$

And,

$$Pr[1 \leftarrow \mathcal{T} | F \geq negl(\lambda)] \leq F \tag{60}$$

Considering the equality cases and due to the fact that $Pr[F < negl(\lambda)] = 1 - Pr[F \geq negl(\lambda)]$,

$$Pr[1 \leftarrow \mathcal{T}] = Err(\kappa_1, \kappa_2)(1 - \frac{d+1}{D}) + negl(\lambda)\frac{d+1}{D} \tag{61}$$

Recall that $Err(\kappa_1, \kappa_2) = negl(\kappa_1, \kappa_2)$, $d = poly(n)$ and $D = 2^n$ and hence $\frac{d+1}{D} = negl(n)$ and the probability that the test algorithm outputs 1 is computed as

$$\begin{aligned} Pr[1 \leftarrow \mathcal{T}] &= negl(\kappa_1, \kappa_2)(1 - negl(n)) + negl(\lambda)negl(n) \\ &= negl(\kappa_1, \kappa_2) + negl(\lambda)negl(n) \end{aligned} \tag{62}$$

Let $\lambda = f(\kappa_1, \kappa_2, n)$, therefore we have

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\text{UqPUF}}(\lambda, \mathcal{A})] = Pr[1 \leftarrow \mathcal{T}] = negl(\lambda) \tag{63}$$

and the proof is complete. □

4 Discussion and Future works

In this section, we briefly discuss the relationship between our proposal and other types of PUFs, as well as the open questions and direction for future works.

Here, we briefly discuss how requirements and security properties defined for cPUFs and QR-PUFs [45, 46] in the literature differ from or relate to what we have defined as qPUF in this paper while leaving a concrete comparison between various PUF instances for future studies.

Most of the available PUF structures use digital encoding as their inputs and outputs so that they can easily be integrated with other functionalities in Integrated Circuits (ICs). This means their input-output pairs are bit-strings. As we can encode the bit strings in computational bases of the Hilbert space, the cPUFs can be considered as special types of Unitary qPUFs (UqPUFs) that can only operate on the computational bases, i.e. map the computational bases in their input domain to other computational bases in their output range. So, our result stating that no UqPUF provides quantum existential unforgeability also shows no cPUF, assuming that they can be queried by quantum states, can provide this security notion for $\mu \neq 1$.

According to [3], if a cPUF provides the min-entropy requirement (which imposes that the cPUF responses are linearly independent) then it can provide existential unforgeability [3] against classical adversaries with no quantum access to the cPUF. However, this requirement cannot be satisfied with most of the common cPUF structures as shown in [19, 44, 43, 28]. Instead of the min-entropy requirement that seems hard or impossible to be achieved, we only consider the basic assumption on PUFs that let the behaviour of PUF be unknown to anyone [42]; and instead of existential unforgeability property which seems impossible to be achieved for both cPUFs and qPUFs, we consider the selective unforgeability property which is a weaker, yet more relevant, notion than the existential one.

To the best of our knowledge, there is no study on quantum security of cPUFs in the literature. We emphasise given the speedy progress in quantum technology the investigation of the security of cPUFs against quantum adversaries is crucial. The security of silicon cPUFs and the other types of cPUFs that cannot be queried by quantum states can be explored in the *post-quantum (or standard) security model* where the quantum adversary has only classical interaction with the primitive while he has been equipped with a powerful quantum computer. However, for the other types of cPUF structures like optical PUFs that can naturally be queried with quantum states, the security of cPUFs need to be analysed in the quantum security model where the adversary in addition to having a quantum computer can have quantum access to the cPUF oracle. Note that quantum selective unforgeability of this type of cPUF structures can be investigated in the aforementioned model. We leave exploring these open questions for future studies.

Another main category of PUFs that can be represented via unitary transformations, is Quantum Read-out PUFs (QR-PUFs). The original definition of QR-PUFs considered cPUFs with quantumly-encoded challenge-response pairs. [45, 46]. The security of QR-PUF-based identification protocols has been investigated in specific security models, such as prepare-and-resend adversaries in [45, 46, 39, 22, 47, 38, 18] where either the full unitary transformation or equivalently the classical description of QR-PUF responses for any known challenge, is assumed to be public knowledge. The security of such PUF-based protocols relies on the bounds on the ability of an adversary to estimate an unknown quantum challenge sent by the verifier.

Although our current framework as it is, will not be directly applicable to all sorts

of protocols and scenarios in which QR-PUFs are defined and used due to specific sets of assumptions and adversarial models considered in these scenarios, we believe that an extended variant of QR-PUFs can be studied as a stand-alone primitive in our proposed framework. We call this extended class, Public-Database PUFs (or PDB-PUFs) which include any PUF that can be queried with quantum (or quantumly encoded) challenges, produce quantum states as responses and are modelled by a publicly known unitary transformation or a public database equivalently. Our framework provides security notions against general and quantum adversaries in the standard game-based model. Hence we can also investigate the security of PDB-PUFs, by relaxing the unknownness condition for this class.

It can easily be shown that in the case of PDB-PUFs the adversary has more knowledge compared to qPUFs, so, these PUFs cannot provide quantum existential unforgeability, either. But more interestingly, using our toolkit of the quantum emulation attack, one can also show that, provided that the classical description of the unitary or the responses to be known, PDB-PUFs do not even provide quantum selective unforgeability against QPT adversaries, even if the adversary is unable to efficiently estimate the challenge quantum state. To see why let us assume the challenger to be also an efficient quantum party. Hence a QPT adversary having knowledge over the database can efficiently span a subspace, including the challenge state, hence the approximate response can be produced with high fidelity using the universal quantum emulator as has been discussed in Section 2. We should mention that the feasibility of other quantum attacks with current technologies has been discussed in [45, 46, 39, 22, 47, 38, 18]. However, it remains an interesting open question when the quantum emulator attack presented in this paper can also be demonstrated on emerging quantum devices.

Another interesting direction for future work is whether the assumptions of QR-PUFs can be matched to the current framework to be able to study their provable security against stronger quantum adversaries. It seems that if one can assume the classical description of U_{QR} to be private and the challenge state can be chosen uniformly at random from the whole Hilbert space, the QR-PUFs like qPUFs can provide the quantum selective unforgeability. Although this remains an interesting open problem.

An important complementary question that we left open is the design of concrete qPUF construction based on the formal framework proposed in this work. Introducing a proper construction for quantum PUF would be much more complicated than their classical counterparts as one needs to deal with many complications of the quantum world such as decoherence. Although similar to the case of classical PUF, optical devices still remain good candidates for qPUFs and worth a formal study that would be able to show whether they satisfy all the requirements and properties of a secure qPUF. Moreover, some randomised circuit-based construction such as t-design can also be a suitable candidate for qPUF as we have recently explored [29]. Another challenge in the way of industrialising of the qPUFs is the need for quantum memory for some of the qPUF-based protocols. It is an interesting question that how much this resource can be reduced or even removed in different protocols. Finally, the current definition allows the study of unitary qPUFs while as also mentioned in the paper, by relaxing some of the requirements the framework could also allow for non-unitary qPUF which is another natural open question for the future studies.

References

[1] Andris Ambainis and Joseph Emerson. Quantum t-designs: t-wise independence in the quantum world. In *Proceedings of Twenty-Second Annual IEEE Conference on Computational Complexity (CCC'07)*, pages 129–140. IEEE, 2007. DOI: 10.1109/CCC.2007.26.

[2] Mohammad Hassan Ameri, Mahshid Delavar, and Javad Mohajeri. Provably secure and efficient PUF-based broadcast authentication schemes for smart grid applications. *International Journal of Communication Systems*, 32(8):e3935, 2019. DOI: 10.1002/dac.3935.

[3] Frederik Armknecht, Daisuke Moriyama, Ahmad-Reza Sadeghi, and Moti Yung. Towards a unified security model for physically unclonable functions. In *Proceedings of Cryptographers' Track at the RSA Conference*, pages 271–287. Springer, 2016. DOI: 10.1007/978-3-319-29485-8–16.

[4] Saikrishna Badrinarayanan, Dakshita Khurana, Rafail Ostrovsky, and Ivan Visconti. Unconditional uc-secure computation with (stronger-malicious) PUFs. In *Proceedings of Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 382–411. Springer, 2017. DOI: 10.1007/978-3-319-56620-7–14.

[5] Dan Boneh and Mark Zhandry. Secure signatures and chosen ciphertext security in a quantum computing world. In *Proceedings of Annual International Cryptology Conference*, pages 361–379. Springer, 2013. DOI: 10.1007/978-3-642-40084-1–21.

[6] Dagmar Bruss, Artur Ekert, and Chiara Macchiavello. Optimal universal quantum cloning and state estimation. *Physical Review Letters*, 81(12):2598, 1998. DOI: 10.1103/PhysRevLett.81.2598.

[7] Christina Brzuska, Marc Fischlin, Heike Schröder, and Stefan Katzenbeisser. Physically uncloneable functions in the universal composition framework. In *Proceedings of Annual International Cryptology Conference*, pages 51–70. Springer, 2011. DOI: 10.1007/978-3-642-22792-9–4.

[8] Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001. DOI: 10.1103/PhysRevLett.87.167902.

[9] Ran Canetti and Marc Fischlin. Universally composable commitments. In *Proceedings of Annual International Cryptology Conference*, pages 19–40. Springer, 2001. DOI: 10.1007/3-540-44647-8–2.

[10] Ulysse Chabaud, Eleni Diamanti, Damian Markham, Elham Kashefi, and Antoine Joux. Optimal quantum-programmable projective measurement with linear optics. *Physical Review A*, 98(6):062318, 2018. DOI: 10.1103/PhysRevA.98.062318.

[11] Chip-Hong Chang, Yue Zheng, and Le Zhang. A retrospective and a look forward: Fifteen years of physical unclonable function advancement. *IEEE Circuits and Systems Magazine*, 17(3):32–62, 2017. DOI: 10.1109/MCAS.2017.2713305.

[12] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Optimal cloning of unitary transformation. *Physical review letters*, 101(18):180504, 2008. DOI: 10.1103/PhysRevLett.101.180504.

[13] Christoph Dankert, Richard Cleve, Joseph Emerson, and Etera Livine. Exact and approximate unitary 2-designs and their application to fidelity estimation. *Physical Review A*, 80(1):012304, 2009. DOI: 10.1103/PhysRevA.80.012304.

[14] GM D'Ariano and P Lo Presti. Quantum tomography for measuring experimentally the matrix elements of an arbitrary quantum operation. *Physical Review Letters*, 86(19):4195, 2001. DOI: 10.1103/PhysRevLett.86.4195.

[15] Mahshid Delavar, Sattar Mirzakuchaki, Mohammad Hassan Ameri, and Javad Mohajeri. PUF-based solutions for secure communications in advanced metering infrastructure (ami). *International Journal of Communication Systems*, 30(9):e3195, 2017. DOI: 10.1002/dac.3195.

[16] Mahshid Delavar, Sattar Mirzakuchaki, and Javad Mohajeri. A ring oscillator-based PUF with enhanced challenge-response pairs. *Canadian Journal of Electrical and Computer Engineering*, 39(2):174–180, 2016. DOI: 10.1109/CJECE.2016.2521877.

[17] Mina Doosti, Farzad Kianvash, and Vahid Karimipour. Universal superposition of orthogonal states. *Physical Review A*, 96(5):052318, 2017. DOI: 10.1103/PhysRevA.96.052318.

[18] Lukas Fladung, Georgios M Nikolopoulos, Gernot Alber, and Marc Fischlin. Intercept-resend emulation attacks against a continuous-variable quantum authentication protocol with physical unclonable keys. *Cryptography*, 3(4):25, 2019. DOI: 10.3390/cryptography3040025.

[19] Fatemeh Ganji, Shahin Tajik, Fabian Fäßler, and Jean-Pierre Seifert. Strong machine learning attack against PUFs with no mathematical model. In *International Conference on Cryptographic Hardware and Embedded Systems*, pages 391–411. Springer, 2016. DOI: 10.1007/978-3-662-53140-2–19.

[20] Blaise Gassend, Dwaine Clarke, Marten Van Dijk, and Srinivas Devadas. Silicon physical random functions. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 148–160. ACM, 2002. DOI: 10.1145/586110.586132.

[21] Giulio Gianfelici, Hermann Kampermann, and Dagmar Bruß. Theoretical framework for physical unclonable functions, including quantum readout. *Physical Review A*, 101(4):042337, 2020. DOI: 10.1103/PhysRevA.101.042337.

[22] Sebastianus A Goorden, Marcel Horstmann, Allard P Mosk, Boris Škorić, and Pepijn WH Pinkse. Quantum-secure authentication of a physical unclonable key. *Optica*, 1(6):421–424, 2014. DOI: 10.1364/OPTICA.1.000421.

[23] Daniel Greenbaum and Zachary Dutton. Modeling coherent errors in quantum error correction. *Quantum Science and Technology*, 3(1):015007, 2017. DOI: 10.1088/2058-9565/aa9a06.

[24] Jorge Guajardo, Sandeep S Kumar, Geert-Jan Schrijen, and Pim Tuyls. Fpga intrinsic PUFs and their use for ip protection. In *Proceedings of International workshop on cryptographic hardware and embedded systems*, pages 63–80. Springer, 2007. DOI: 10.1007/978-3-540-74735-2–5.

[25] B. Halak. *Physically Unclonable Functions: From Basic Design Principles to Advanced Hardware Security Applications*. Springer International Publishing, 2019. DOI: 10.1007/978-3-319-76804-5.

[26] Charles Herder, Meng-Day Yu, Farinaz Koushanfar, and Srinivas Devadas. Physical unclonable functions and applications: A tutorial. *Proceedings of the IEEE*, 102(8):1126–1141, 2014. DOI: 10.1109/JPROC.2014.2320516.

[27] Jonathan Katz. Universally composable multi-party computation using tamper-proof hardware. In *Proceedings of Annual International Conference on the Theory*

and Applications of Cryptographic Techniques, pages 115–128. Springer, 2007. DOI: 10.1007/978-3-540-72540-4–7.

[28] Mahmoud Khalafalla and Catherine Gebotys. PUFs deep attacks: Enhanced modeling attacks using deep learning techniques to break the security of double arbiter PUFs. In *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 204–209. IEEE, 2019. DOI: 10.23919/DATE.2019.8714862.

[29] Niraj Kumar, Rawad Mezher, and Elham Kashefi. Efficient construction of quantum physical unclonable functions with unitary t-designs, 2021. arXiv:2101.05692.

[30] Weiqiang Liu, Lei Zhang, Zhengran Zhang, Chongyan Gu, Chenghua Wang, Maire O'neill, and Fabrizio Lombardi. Xor-based low-cost reconfigurable pufs for iot security. *ACM Transactions on Embedded Computing Systems (TECS)*, 18(3):1–21, 2019. DOI: 10.1145/32746665.

[31] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631, 2014. DOI: 10.1038/nphys3029.

[32] Roel Maes. *Physically Unclonable Functions: Constructions, Properties and Applications*. Springer-Verlag Berlin Heidelberg, 2016. DOI: 10.1007/978-3-642-41395-7–3.

[33] Cédric Marchand, Lilian Bossuet, Ugo Mureddu, Nathalie Bochard, Abdelkarim Cherkaoui, and Viktor Fischer. Implementation and characterization of a physical unclonable function for iot: a case study with the tero-PUF. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(1):97–109, 2017. DOI: 10.1109/TCAD.2017.2702607.

[34] Iman Marvian and Seth Lloyd. Universal quantum emulator. *arXiv preprint arXiv:1606.02734*, 2016. arXiv:1606.02734.

[35] Charis Mesaritakis, Marialena Akriotou, Alexandros Kapsalis, Evangelos Grivas, Charidimos Chaintoutis, Thomas Nikas, and Dimitris Syvridis. Physical unclonable function based on a multi-mode optical waveguide. *Scientific reports*, 8(1):9653, 2018. DOI: 10.1038/s41598-018-28008-6.

[36] Debdeep Mukhopadhyay. PUFs as promising tools for security in internet of things. *IEEE Design & Test*, 33(3):103–115, 2016. DOI: 10.1109/MDAT.2016.2544845.

[37] Michael A Nielsen and Isaac L Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 10th edition, 2010. DOI: 10.1017/CBO9780511976667.

[38] Georgios M Nikolopoulos. Continuous-variable quantum authentication of physical unclonable keys: Security against an emulation attack. *Physical Review A*, 97(1):012324, 2018. DOI: 10.1103/PhysRevA.97.012324.

[39] Georgios M Nikolopoulos and Eleni Diamanti. Continuous-variable quantum authentication of physical unclonable keys. *Scientific reports*, 7:46047, 2017. DOI: 10.1038/srep46047.

[40] Michał Oszmaniec, Andrzej Grudka, Michał Horodecki, and Antoni Wójcik. Creating a superposition of unknown quantum states. *Physical Review Letters*, 116(11), 2016. DOI: 10.1103/PhysRevLett.116.110403.

[41] Ravikanth Pappu, Ben Recht, Jason Taylor, and Neil Gershenfeld. Physical one-way functions. *Science*, 297(5589):2026–2030, 2002. DOI: 10.1126/science.1074376.

[42] Ulrich Rührmair and Daniel E Holcomb. PUFs at a glance. In *Proceedings of the conference on Design, Automation & Test in Europe*, page 347. European Design and Automation Association, 2014. DOI: 10.7873/DATE.2014.360.

[43] Ulrich Rührmair, Frank Sehnke, Jan Sölter, Gideon Dror, Srinivas Devadas, and Jürgen Schmidhuber. Modeling attacks on physical unclonable functions. In *Proceedings of the 17th ACM conference on Computer and communications security*, pages 237–249, 2010. DOI: 10.1145/1866307.1866335.

[44] Ulrich Ruhrmair and Jan Solter. PUF modeling attacks: An introduction and overview. In *2014 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, 2014. DOI: 10.7873/DATE2014.361.

[45] Boris Škorić. Quantum readout of physical unclonable functions. In *Proceedings of International Conference on Cryptology in Africa*, pages 369–386. Springer, 2010. DOI: 10.1007/978-3-642-12678-9–22.

[46] BORIS ŠKORIĆ. Quantum readout of physical unclonable functions. *International Journal of Quantum Information*, 10(01):1250001, 2012. DOI: 10.1142/S0219749912500013.

[47] Boris Škorić, Allard P Mosk, and Pepijn WH Pinkse. Security of quantum-readout PUFs against quadrature-based challenge-estimation attacks. *International journal of quantum information*, 11(04):1350041, 2013. DOI: 10.1142/S021974991350041X.

[48] Boris Škorić, Pepijn WH Pinkse, and Allard P Mosk. Authenticated communication from quantum readout of PUFs. *Quantum Information Processing*, 16(8):200, 2017. DOI: 10.1007/s11128-017-1649-0.

[49] G Edward Suh and Srinivas Devadas. Physical unclonable functions for device authentication and secret key generation. In *Proceedings of 44th ACM/IEEE Design Automation Conference*, pages 9–14. IEEE, 2007. 2007 44th ACM/IEEE Design Automation Conference.

[50] Lars Tebelmann, Michael Pehl, and Vincent Immler. Side-channel analysis of the tero PUF. In *International Workshop on Constructive Side-Channel Analysis and Secure Design*, pages 43–60. Springer, 2019. DOI: 10.1007/978-3-030-16350-1–4.

[51] Ravitej Uppu, Tom AW Wolterink, Sebastianus A Goorden, Bin Chen, Boris Škorić, Allard P Mosk, and Pepijn WH Pinkse. Asymmetric cryptography with physical unclonable keys. *Quantum Science and Technology*, 4(4):045011, 2019. DOI: 10.1088/2058-9565/ab479f.

[52] William K Wootters and Wojciech H Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802, 1982. DOI: 10.1038/299802a0.

[53] Yao Yao, Ming Gao, Mo Li, and Jian Zhang. Quantum cloning attacks against PUF-based quantum authentication systems. *Quantum Information Processing*, 15(8):3311–3325, 2016. DOI: 10.1007/s11128-016-1316-x.

[54] Robert Young, Utz Roedig, and Jonathan Roberts. Quantum physical unclonable function, 2019. Patent: US10148435B2.

[55] Karol Życzkowski and Hans-Jürgen Sommers. Average fidelity between random quantum states. *Physical Review A*, 71(3):032313, 2005. DOI: 10.1103/PhysRevA.71.032313.

A Background on Classical Physical Unclonable Functions

In this section, we briefly present the formal definition of Physical Unclonable Functions (PUFs) as found in the classical literature [3, 42, 7]. Let a $\mathcal{D}$-family be a set of physical devices generated through the same manufacturing process. Due to unavoidable variations during manufacturing, each device has some unique features that are not easily clonable. A Physical Unclonable Function (PUF) is an operation making these features observable and measurable by the holder of the device.

As in [3, 7], we formalize the manufacturing process of a PUF by defining the Gen algorithm that takes the security parameter λ as input and generates a PUF with an identifier **id**. Note that each time the Gen algorithm is run, a new PUF with new **id** is built. So, we have:

$$\mathrm{PUF}_{\mathbf{id}} \leftarrow \mathrm{Gen}(\lambda). \tag{64}$$

Also, we define the Eval algorithm that takes a challenge x and $\mathrm{PUF}_{\mathbf{id}}$ as inputs and generates the corresponding response $y_{\mathbf{id}}$ as output:

$$y_{\mathbf{id}} \leftarrow \mathrm{Eval}(\mathrm{PUF}_{\mathbf{id}}, x). \tag{65}$$

Due to variations in the environmental conditions, for any given $\mathrm{PUF}_{\mathbf{id}}$, the Eval algorithm may generate a different response to the same challenge x. It is required that this noise be bounded as follows; if $\mathrm{Eval}(\mathrm{PUF}_{\mathbf{id}}, x)$ is run several times, the maximum distance between the corresponding responses should at most be δ_r. This requirement is termed the *robustness requirement*.

Consider a family of PUF generated by the same Gen algorithm, and assume the algorithm Eval is run on all of them with a single challenge x. To be able to distinguish each $\mathrm{PUF}_{\mathbf{id}}$, it is required that the minimum distance between the corresponding responses be at least δ_u. This requirement is termed the *uniqueness requirement*.

The other requirement considered in [3] is *collision-resistance*. This imposes that whenever the Eval algorithm is run on $\mathrm{PUF}_{\mathbf{id}}$ with different challenges, the minimum distance between the different responses must be at least δ_c. The parameters δ_r, δ_u, δ_c are determined by the security parameter λ. Robustness, uniqueness and collision-resistance are crucial for correctness of cryptographic schemes built on top of PUFs. The conditions $\delta_r \leq \delta_u$ and $\delta_r \leq \delta_c$ must be satisfied to allow for distinguishing different challenges and PUFs [3].

According to the above, a $(\lambda, \delta_r, \delta_u, \delta_c)$-PUF is defined as a pair of algorithms: Gen and Eval that provides the robustness, uniqueness and collision-resistance requirements. We call a $(\lambda, \delta_r, \delta_u, \delta_c)$-PUF a Classical PUF (cPUF), if the Eval algorithm runs on classical information such as bit strings. Any classical function $f : \{0,1\}^n \to \{0,1\}^m$, including a cPUF's Eval, can be modelled as a unitary transformation as follows

$$\forall x \in \{0,1\}^n, \forall y \in \{0,1\}^m : U_f \left|x, y\right\rangle := \left|x, f(x) \oplus y\right\rangle \tag{66}$$

and thus a quantum adversary can query U_f on any desired quantum states such as the superposition of all the classical inputs.

B Proof of Theorem 2: Quantum Emulation Output

Here we give the full proof of Theorem 2 as follows.

Proof: We prove the theorem by induction. For the first block ($K = 1$), according to equation (9) and letting $|\chi_0\rangle = |\psi\rangle$ we have:

$$|\chi_1\rangle = \frac{1}{2}[(I - R(\phi_r))\,|\psi\rangle\,|0\rangle + R(\phi_i)(\mathbb{I} + R(\phi_r))\,|\psi\rangle\,|1\rangle] \tag{67}$$

where the term $I - R(\phi_r) = 2\,|\phi_r\rangle\,\langle\phi_r|$ projects the previous state to $|\phi_r\rangle$ with the coefficient $\langle\phi_r|\,\psi\rangle$ and the term $R(\phi_i)(I + R(\phi_r))$ is equal to:

$$R(\phi_i)(I + R(\phi_r)) = 2[I - |\phi_r\rangle\,\langle\phi_r| - 2\,|\phi_i\rangle\,\langle\phi_i| + 2\,\langle\phi_i|\,\phi_r\rangle\,|\phi_i\rangle\,\langle\phi_r|]. \tag{68}$$

Thus, the final relation between all the parameters in the first block is as follows.

$$|\chi_1\rangle = \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|0\rangle + |\psi\rangle\,|1\rangle - \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|1\rangle - 2\,\langle\phi_1|\,\psi\rangle\,|\phi_1\rangle\,|1\rangle + 2\,\langle\phi_r|\,\psi\rangle\,\langle\phi_r|\,\phi_1\rangle\,|\phi_1\rangle\,|1\rangle \tag{69}$$

As can be seen, it satisfies the form of equation (10) where the first sum is zero and in the second sum $g_{10} = -1, g_{11} = +1$, $l'_{10} = l'_{11} = 1$, $x'_{10} = z'_{10} = 0, y'_{10} = 1$, $x'_{11} = z'_{11} = 1$ and $y'_{11} = 0$.

Now we write $|\chi_K\rangle$ according to equation (9), assume $|\chi_{K-1}\rangle$ is written in form of equation (10) and show $|\chi_K\rangle$ also satisfies this equation.

$$\begin{aligned}|\chi_K\rangle = {} & \langle\phi_r|\,\chi_{K-1}\rangle\,|\phi_r\rangle\,|0\rangle + |\chi_{K-1}\rangle\,|1\rangle - \langle\phi_r|\,\chi_{K-1}\rangle\,|\phi_r\rangle\,|1\rangle - 2\,\langle\phi_K|\,\chi_{K-1}\rangle\,|\phi_K\rangle\,|1\rangle \\ & + 2\,\langle\phi_r|\,\chi_{K-1}\rangle\,\langle\phi_r|\,\phi_K\rangle\,|\phi_K\rangle\,|1\rangle\end{aligned} \tag{70}$$

By substituting $|\chi_{K-1}\rangle$ with its equivalent based on equation (10), we calculate each term in the above formula. Note that the coefficient in the third term is the same as the first one with a minus sign, and the ancillary state for the first term is $|0\rangle$ while for the third term is $|1\rangle$. Thus, we only show the details of the calculation for the first term:

$$\begin{aligned}&\langle\phi_r|\rangle\chi_{K-1}\,|\phi_r\rangle\,|0\rangle = \\ &\quad \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|0\rangle^{\otimes K} + \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|1\rangle^{\otimes K-1}\,|0\rangle - \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|1\rangle^{\otimes K-1}\,|0\rangle + \\ &\quad + \sum_{i=1}^{K-1}\sum_{j=0}^{i}[f_{ij}2^{l_{ij}}|\,\langle\phi_r|\,\psi\rangle|^{x_{ij}}|\,\langle\phi_i|\,\psi\rangle|^{y_{ij}}|\,\langle\phi_r|\,\phi_i\rangle|^{z_{ij}}]\,|\phi_r\rangle\,|q_{anc}(i,j)\rangle\,|0\rangle \\ &\quad + \sum_{i=1}^{K-1}\sum_{j=0}^{i}[g_{ij}2^{l'_{ij}}|\,\langle\phi_r|\,\psi\rangle|^{x'_{ij}}|\,\langle\phi_i|\,\psi\rangle|^{y'_{ij}}|\,\langle\phi_r|\,\phi_i\rangle|^{z'_{ij}+1}]\,|\phi_i\rangle\,|q'_{anc}(i,j)\rangle\,|0\rangle\,.\end{aligned} \tag{71}$$

The second term is calculated as follows:

$$\begin{aligned}|\chi_{K-1}\rangle\,|1\rangle = {} & \langle\phi_r|\,\psi\rangle\,|0\rangle^{\otimes K-1}\,|1\rangle + |\psi\rangle\,|1\rangle^{\otimes K} - \langle\phi_r|\,\psi\rangle\,|\phi_r\rangle\,|1\rangle^{\otimes K} + \\ & + \sum_{i=1}^{K-1}\sum_{j=0}^{i}[f_{ij}2^{l_{ij}}|\,\langle\phi_r|\,\psi\rangle|^{x_{ij}}|\,\langle\phi_i|\,\psi\rangle|^{y_{ij}}|\,\langle\phi_r|\,\phi_i\rangle|^{z_{ij}}]\,|\phi_r\rangle\,|q_{anc}(i,j)\rangle\,|1\rangle \\ & + \sum_{i=1}^{K-1}\sum_{j=0}^{i}[g_{ij}2^{l'_{ij}}|\,\langle\phi_r|\,\psi\rangle|^{x'_{ij}}|\,\langle\phi_i|\,\psi\rangle|^{y'_{ij}}|\,\langle\phi_r|\,\phi_i\rangle|^{z'_{ij}}]\,|\phi_i\rangle\,|q'_{anc}(i,j)\rangle\,|1\rangle\,.\end{aligned} \tag{72}$$

The forth term $-2\,\langle\phi_K|\,\chi_{K-1}\rangle\,|\phi_K\rangle\,|1\rangle$ has the coefficient $-2\,\langle\phi_K|\,\chi_{K-1}\rangle$, which produces the same sigma terms while only $l'_{i,j}, x'_{i,j}, y'_{i,j}$ and $z'_{i,j}$ are increased by one. The fifth term $2\,\langle\phi_r|\,\chi_{K-1}\rangle\,\langle\phi_r|\,\phi_K\rangle\,|\phi_K\rangle\,|1\rangle$ has the coefficient $2\,\langle\phi_r|\,\chi_{K-1}\rangle\,\langle\phi_r|\,\phi_K\rangle$ and similarly produces the same sigma terms where $l_{i,j}$, $x_{i,j}$, $y_{i,j}$ and $z_{i,j}$ are increased by one (Note that the $\langle\phi_r|\,\phi_K\rangle$ is itself one of the terms of the sigma). Finally by adding all these terms the equation (10) is obtained and the proof is complete. □

C Lemma for the Proof of Theorem 3

We establish the following lemma that we have used in the proof of theorem 3.

Lemma 2 *Let $\mathcal{E}$ be a CPT map of the for $\mathcal{E}(\rho) = (1-\epsilon)U\rho U^\dagger + \epsilon\tilde{\mathcal{E}}(\rho)$ where U is a unitary and $\tilde{\mathcal{E}}$ is a strictly contractive CPT map. Let ρ and σ be two arbitrary density matrices with trace distance $D = \mathcal{D}_{tr}(\rho,\sigma)$. Then the following inequality holds:*

$$\mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \leq \epsilon D \tag{73}$$

Proof: We note that the first part of the channel $\mathcal{E}$, which outputs density matrix $U\rho U^\dagger$ with probability $(1-\epsilon)^2$, is a unitary and preserves the distance. As a result, for a fixed value of ϵ and fixed arbitrary states ρ and σ, the difference between the trace distances of the output of $\mathcal{E}$ and the input states increases as $\tilde{\mathcal{E}}$ becomes more contractive. As the maximum contractivity of $\tilde{\mathcal{E}}$ occurs when $\tilde{\mathcal{E}} = \frac{I}{d}$, then the maximum difference between the output and input trace distances is satisfies for this instance of the channel. Let $\mathcal{E}'(\rho) = (1-\epsilon)U\rho U^\dagger + \epsilon\frac{I}{d}$. Then for a fixed ϵ we will have:

$$\mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \leq \mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}'(\rho),\mathcal{E}'(\sigma)) \tag{74}$$

Now we calculate $\mathcal{D}_{tr}(\mathcal{E}'(\rho),\mathcal{E}'(\sigma))$ using the definition of the trace distance which is $\mathcal{D}_{tr}(\rho,\sigma) = \frac{1}{2}tr(|\rho-\sigma|)$. And $|A| = \sqrt{A^\dagger A}$ for a positive semidefinite matrix A. We calculate the trace distance as:

$$\begin{aligned}
\mathcal{D}_{tr}(\mathcal{E}'(\rho),\mathcal{E}'(\sigma)) &= \frac{1}{2}tr[|\mathcal{E}'(\rho),\mathcal{E}'(\sigma)|] = \frac{1}{2}tr[|(1-\epsilon)U\rho U^\dagger + \epsilon\frac{I}{d} - (1-\epsilon)U\sigma U^\dagger - \epsilon\frac{I}{d}|] \\
&= (1-\epsilon)(\frac{1}{2}tr[|U\rho U^\dagger - U\sigma U^\dagger|]) = (1-\epsilon)\mathcal{D}_{tr}(U\rho U^\dagger, U\sigma U^\dagger) \\
&= (1-\epsilon)\mathcal{D}_{tr}(\rho,\sigma) \\
&= (1-\epsilon)D
\end{aligned} \tag{75}$$

Finally, we can relate the desired trace distance with the above value as:

$$\mathcal{D}_{tr}(\rho,\sigma) - \mathcal{D}_{tr}(\mathcal{E}(\rho),\mathcal{E}(\sigma)) \leq D - (1-\epsilon)D = \epsilon D \tag{76}$$

And the lemma has been proved. □

D Full Proof of Theorem 6

Proof: Let $\mathcal{A}$ be a QPT adversary playing the game $\mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda,\mathcal{A})$ where UqPUF is defined over $\mathcal{H}^D$. Let S_{in} and S_{out} be the input and output database of the adversary after the learning phase both with size k_1, respectively. Also, Let $\mathcal{H}^d$ be the d-dimensional Hilbert space spanned by elements of S_{in} where $d \leq k_1$ and $\mathcal{H}^d_{out}$ be the Hilbert space spanned by elements of S_{out} with the same dimension. $\mathcal{A}$ receives an unknown quantum state $|\psi\rangle$ as a challenge in the qSel challenge phase and tries to output a state $|\omega\rangle$ as close as possible to $|\psi^{out}\rangle$. We are interested in calculating the average probability that the fidelity of $\mathcal{A}$'s output state $|\omega\rangle$ and $|\psi^{out}\rangle$ be larger or equal to δ. We calculate this probability over all the possible states chosen uniformly at random from $\mathcal{H}^D$.

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda,\mathcal{A})] = \Pr_{|\psi\rangle\in\mathcal{H}^D}[F(|\omega\rangle, |\psi^{out}\rangle) \geq \delta] \tag{77}$$

We calculate this probability over all the possible states chosen uniformly at random from $\mathcal{H}^D$. We will show, for any $\delta \neq 0$, the success probability of $\mathcal{A}$ is negligible in λ.

According to the game definition, as the adversary selects states of the learning phase, the classical description of these states are known for him while the corresponding responses are unknown quantum states. Let $\mathcal{A}'$ be the adversary who also receives the classical description of the outputs, or the complete set of bases of $\mathcal{H}^d$ and $\mathcal{H}^d_{out}$. So, he will have a complete description of the map in the subspace; and as a result $\mathcal{A}'$ has a greater success probability than $\mathcal{A}$.

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda, \mathcal{A})] \leq Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda, \mathcal{A}')] \tag{78}$$

Therefore from now on throughout the proof, we calculate the success probability of $\mathcal{A}'$ who has full knowledge of the subspace.

Note that the adversary cannot enhance his knowledge of the subspace by entangling its local system to the challenges of the learning phase since the reduced density matrix of the challenge/response entangled state lies in the same subspace $\mathcal{H}^d$ and $\mathcal{H}^d_{out}$. Hereby upper-bounding the success probability of $\mathcal{A}$ with the success probability of $\mathcal{A}'$ who has the full knowledge of the subspace we have also included the entangled queries. Thus without loss of generality and to avoid complicated notations, we consider the adversary's estimated state as a pure state $|\omega\rangle$.

Now, we partition the set of all the challenges to two parts: the challenges that are completely orthogonal to $\mathcal{H}^d$ subspace, and the rest of the challenges that have non-zero overlap with $\mathcal{H}^d$. We denote the subspace of all the states orthogonal to $\mathcal{H}^d$ as $\mathcal{H}^{d^\perp}$. We calculate the success probability of $\mathcal{A}'$ in terms of the following partial probabilities:

$$\underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] \text{ and } \underset{|\psi\rangle \notin \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta]. \tag{79}$$

Because the probability of $|\psi\rangle$ being in any particular subset is independent of the adversary's learnt queries, the success probability of $\mathcal{A}'$ can be written as:

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda, \mathcal{A}')] = \underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] \times Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] + \underset{|\psi\rangle \notin \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] \times Pr[|\psi\rangle \notin \mathcal{H}^{d^\perp}] \tag{80}$$

where $Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] = 1 - Pr[|\psi\rangle \notin \mathcal{H}^{d^\perp}]$ denotes the probability of $|\psi\rangle$ that is picked uniformly at random from $\mathcal{H}^D$ being projected into the subspace of $\mathcal{H}^{d^\perp}$. From lemma 1, we know that this probability for any subspace, is equal to the ratio of the dimensions. As $\mathcal{H}^{d^\perp}$ is a $D-d$ dimensional subspace, $Pr[|\psi\rangle \in \mathcal{H}^{d^\perp}] = \frac{D-d}{D}$ and respectively $Pr[|\psi\rangle \notin \mathcal{H}^{d^\perp}] = \frac{d}{D}$. Also the probability is upper-bounded by the cases that the adversary can always win the game for $|\psi\rangle \notin \mathcal{H}^{d^\perp}$. So, we have,

$$Pr[1 \leftarrow \mathcal{G}^{\text{UqPUF}}_{\mathsf{qSel}}(\lambda, \mathcal{A}')] \leq \underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] \times (\frac{D-d}{D}) + \frac{d}{D} \tag{81}$$

Finally, the only term that should be calculated is $\underset{|\psi\rangle \in \mathcal{H}^{d^\perp}}{Pr}[F \geq \delta]$.

Note that any $|\psi\rangle \in \mathcal{H}^D$ can be written in any set of full bases of $\mathcal{H}^D$ as $|\psi\rangle = \sum_{i=1}^{D} c_i |e_i\rangle$. For any $|\psi\rangle \in \mathcal{H}^{d^\perp}$, the set of $\{|e_i\rangle\}_{i=1}^{D}$ can be the a union of the bases of $\mathcal{H}^d$, i.e. $\{|e_i^{in}\rangle\}_{i=1}^{d}$ and the bases of $\mathcal{H}^{d^\perp}$, i.e. $\{|e_i'\rangle\}_{i=d+1}^{D}$. Note that any state in $\mathcal{H}^{d^\perp}$ is orthogonal to all the $|e_i^{in}\rangle$s. Thus, we write as follows

$$|\psi\rangle = \sum_{i=1}^{d} c_i^{in} |e_i^{in}\rangle + \sum_{i=d+1}^{D} c_i' |e_i'\rangle \tag{82}$$

Recall that $|\psi\rangle \in \mathcal{H}^{d^\perp}$, so, $\langle\psi|\, e_i^{in}\rangle = 0$ and as a result $c_i^{in} = 0$. So,

$$|\psi\rangle = \sum_{i=d+1}^{D} c_i' \,|e_i'\rangle \tag{83}$$

Similarly for the output state $|\psi^{out}\rangle = \sum_{i=1}^{d} c_i^{out}\,|e_i^{out}\rangle + \sum_{i=d+1}^{D} \alpha_i\,|b_i\rangle$, as the unitary preserves the inner product, $c_i^{out} = \langle e_i^{out}|\,\psi^{out}\rangle = \langle e_i^{in}|\,U^\dagger U\,|\psi\rangle = \langle e_i^{in}|\,\psi\rangle = 0$, and the correct output state can be written as

$$|\psi^{out}\rangle = \sum_{i=d+1}^{D} \alpha_i\,|b_i\rangle \tag{84}$$

where $\{|b_i\rangle\}_{i=1}^{D-d}$ are a set of bases for $\mathcal{H}_{out}^{d^\perp}$. The output estimated by the adversary $\mathcal{A}'$ can be written as

$$|\omega\rangle = \sum_{i=1}^{d} \beta_i\,|e_i^{out}\rangle + \sum_{i=d+1}^{D} \gamma_i\,|q_i\rangle \tag{85}$$

where the first term represents part of the output state, that has been produced by $\mathcal{A}$ from the his learnt output subspace and the second term denotes the part lies in $\mathcal{H}_{out}^{d^\perp}$ with the set of bases $\{|q_i\rangle\}_{i=1}^{D-d}$. Based on the above argument, the fidelity of the first part is always zero as $\langle b_i|\, e_i^{out}\rangle = 0$.

Note that the normalization condition implies $\sum_{i=1}^{d} |\beta_i|^2 + \sum_{i=d+1}^{D} |\gamma_i|^2 = 1$. Thus for any state $|\omega\rangle$ that has a non-zero overlap with the learnt outputs, the fidelity with the correct state decreases. To make the $\mathcal{A}'$'s strategies optimal we assume $\sum_{i=1}^{D-d} \gamma_i\,|q_i\rangle \in \mathcal{H}_{out}^{d^\perp}$ where the normalization condition is $\sum_{i=1}^{D-d} |\gamma_i|^2 = 1$.

Since there are infinite choices for set of bases orthogonal to $\{|e_i^{out}\rangle\}_{i=1}^{d}$, there is no way to uniquely choose or obtain the rest of the bases to complete the set. Also, another input of the adversary is the state $|\psi\rangle$ which according to the game definition, is an unknown state from a uniform distribution. As a result, the choice of the $|q_i\rangle$ bases are also independent of $|e_i'\rangle$ or $|b_i\rangle$. Thus knowing a matching pair of $(|q_i\rangle, |b_i\rangle)$ increases the dimension of the known subspace by one that means the adversary has more information that it is assumed to have.

So, for each new challenge, $\mathcal{A}'$ produces a state $|\omega\rangle = \sum_{i=1}^{D-d} \gamma_i\,|q_i\rangle$ with a totally independent choice of bases. Without loss of generality we can fix the bases $|q_i\rangle$ for different $|\omega\rangle$. To calculate the success probability of $\mathcal{A}'$, we calculate the fidelity averaging over all the possible choices of ψ. As the unitary transformation preserves the distance, it maps a uniform distribution of states to a uniform distribution. This leads to a uniform distribution of all the possible $|\psi^{out}\rangle$. As a result, the average probability over all possible $|\psi\rangle$ is equal to the average probability over all possible $|\psi^{out}\rangle$.

$$\underset{|\psi\rangle\in\mathcal{H}^{d^\perp}}{Pr}[F \geq \delta] = \underset{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d^\perp}}{Pr}[F \geq \delta]. \tag{86}$$

Now, we show that the adversary $\mathcal{A}'$ also needs to output $|\omega\rangle$ according to the uniform distribution to win the game in the average case.

Let $\mathcal{A}'$ output the states according to a probability distribution $\mathfrak{D}$ which is not uniform. Then, by repeating the experiment asymptotically many times, the correct response $|\psi^{out}\rangle$ covers the whole $\mathcal{H}_{out}^{d^\perp}$ while $|\omega\rangle$ covers a subspace of $\mathcal{H}_{out}^{d^\perp}$. This decreases the average success probability of $\mathcal{A}'$. So, the best strategy for $\mathcal{A}'$ is to generate the states $|\omega\rangle$ such that they span the whole $\mathcal{H}_{out}^{d^\perp}$, i.e. generating them according to the uniform distribution.

Based on the above argument, and the fact that all the $|\omega\rangle$s are produced independently, we show that the average fidelity over all the $|\psi^{out}\rangle$ is equivalent to average fidelity over all the $|\omega\rangle$.

There are different methods for calculating the average fidelity [55], but most commonly the average fidelity can be written as:

$$\int\limits_{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d\perp}} |\langle\omega|\psi_x^{out}\rangle|^2 d\mu_x \tag{87}$$

where $d\mu$ is a measure based on which the reference state has been produced and parameterized. According to our uniformity assumption, the $d\mu$ here is the Haar measure. Note that $|\omega\rangle$ can be different for any new challenge. Now we rewrite the above average with the new parameters as:

$$\begin{aligned}
\int\limits_{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d\perp}} F(|\omega\rangle, |\psi_x^{out}\rangle) d\mu_x &= \int\limits_{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d\perp}} |\langle\omega|\psi_x^{out}\rangle|^2 d\mu_x \\
&= \int\limits_{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d\perp}} |\sum_{i=1}^{D-d} \overline{\gamma_i}\langle q_i|\psi_x^{out}\rangle|^2 d\mu_x \\
&= \int\limits_{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d\perp}} |\sum_{i=1}^{D-d} \overline{\gamma_{i_x}}\langle q_i|\psi^{out}\rangle|^2 d\mu_x \\
&= \int\limits_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}} |\langle\omega_x|\psi^{out}\rangle|^2 d\mu_x \\
&= \int\limits_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}} F(|\omega_x\rangle, |\psi^{out}\rangle) d\mu_x
\end{aligned} \tag{88}$$

The above equality holds since the fidelity is a symmetric function of two states and the measure of integral is the same for both cases. We use this equality for averaging all the possible outputs for one $|\psi^{out}\rangle$. Recall that we aim to calculate the probability of the average fidelity being greater than δ. To this end, we first calculate a more general probability that is the probability of the average fidelity to be non-zero. As we have

$$\Pr_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}}[F\neq 0] + \Pr_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}}[F=0] = 1, \tag{89}$$

we calculate the probability of the zero fidelity for simplicity. So,

$$\begin{aligned}
\Pr_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}}[F=0] &= \Pr_{|\omega\rangle\in\mathcal{H}_{out}^{d\perp}}[|\langle\omega|\psi^{out}\rangle|^2 = 0] \\
&= Pr[(\int |\sum_{i=1}^{D-d} \overline{\gamma_{i_x}}\langle q_i|\psi^{out}\rangle|^2 d\mu_x) = 0] \\
&= \Pr_x[(\sum_{i,j=1}^{D-d} \overline{\gamma_{i_x}}\alpha_j \langle q_{i_x}|b_j\rangle)^2 = 0]
\end{aligned} \tag{90}$$

Based on the Cauchy–Schwarz inequality we have the following inequality:

$$[\sum_{i,j=1}^{D-d} \overline{\gamma_{i_x}}\alpha_j \langle q_i|b_j\rangle]^2 \geq \sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|b_j\rangle|^2 \tag{91}$$

where,

$$\sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|\, b_j\rangle|^2 = \sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|\, b_j\rangle \langle b_j|\, q_i\rangle| = \sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|\, \Pi_j\, |q_i\rangle\,| \tag{92}$$

So, we have,

$$\underset{|\omega\rangle\in\mathcal{H}_{out}^{d^\perp}}{Pr}[F=0] \geq \underset{x}{Pr}[\sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|\, \Pi_j\, |q_i\rangle\,| = 0] \tag{93}$$

The smaller term is the probability of $|\omega\rangle$ being projected into the orthogonal subspace of a space that only includes $|\psi^{out}\rangle$ averaging over all the projectors. We call again Lemma 1. As the target subspace includes only one vector of the Hilbert space, the dimension of the orthogonal subspace is always one dimension less. Recall that $d^\perp = D - d$, the dimension of the intended orthogonal subspace is equal to $D - d - 1$. So,

$$\begin{aligned} &\underset{x}{Pr}[(\sum_{i,j=1}^{D-d} |\overline{\gamma_{i_x}}\alpha_j|^2 |\langle q_i|\, \Pi_j\, |q_i\rangle\,|) = 0] = \frac{D-d-1}{D-d} \Rightarrow \\ &\underset{|\omega\rangle\in\mathcal{H}_{out}^{d^\perp}}{Pr}[F=0] \geq \frac{D-d-1}{D-d} \end{aligned} \tag{94}$$

And as a result,

$$\underset{|\psi^{out}\rangle\in\mathcal{H}_{out}^{d^\perp}}{Pr}[|\langle\omega|\,\psi^{out}\rangle| \neq 0] \leq \frac{1}{D-d} \tag{95}$$

So, for any non-zero δ we have,

$$\underset{|\psi\rangle\in\mathcal{H}^{d^\perp}}{Pr}[|\langle\omega|\,\psi^{out}\rangle| \geq \delta] \leq \frac{1}{D-d} \tag{96}$$

Thus, the success probability of $\mathcal{A}'$ is

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A}')] = \frac{1}{D-d} \times (\frac{D-d}{D}) + \frac{d}{D} = \frac{d+1}{D} \tag{97}$$

And the success probability of $\mathcal{A}$ is bounded by $\frac{d+1}{D}$,

$$Pr[1 \leftarrow \mathcal{G}_{\mathsf{qSel}}^{\mathrm{UqPUF}}(\lambda, \mathcal{A})] \leq \frac{d+1}{D} \tag{98}$$

and the theorem has been proved. □