

# Building Quantum-One-Way Functions from Block Ciphers: Davies-Meyer and Merkle-Damgård Constructions

Akinori Hosoyamada and Kan Yasuda

NTT Secure Platform Laboratories,  
3-9-11, Midori-cho Musashino-shi, Tokyo 180-8585, Japan.  
{hosoyamada.akinori,yasuda.kan}@lab.ntt.co.jp

**Abstract.** We present hash functions that are almost optimally one-way in the quantum setting. Our hash functions are based on the Merkle-Damgård construction iterating a Davies-Meyer compression function, which is built from a block cipher. The quantum setting that we use is a natural extension of the classical ideal cipher model. Recent work has revealed that symmetric-key schemes using a block cipher or a public permutation, such as CBC-MAC or the Even-Mansour cipher, can get completely broken with quantum superposition attacks, in polynomial time of the block size. Since many of the popular schemes are built from a block cipher or a permutation, the recent findings motivate us to study such schemes that are provably secure in the quantum setting. Unfortunately, no such schemes are known, unless one relies on certain algebraic assumptions. In this paper we present hash constructions that are provably one-way in the quantum setting without algebraic assumptions, solely based on the assumption that the underlying block cipher is ideal. To do this, we reduce one-wayness to a problem of finding a fixed point and then bound its success probability with a distinguishing advantage. We develop a generic tool that helps us prove indistinguishability of two quantum oracle distributions.

**keywords** symmetric key cryptography, provable security, Merkle-Damgård, Davies-Meyer, one-wayness, non-invertibility, preimage-resistance, derangement, fixed point, indistinguishability, quantum ideal cipher model

## 1 Introduction

The epoch-making work by Shor [26] revealed that widely used cryptographic schemes such as RSA, DSA and ECDSA would become insecure when a practical quantum computer becomes available. Since then, researchers have become increasingly interested in so-called *post-quantum* cryptography. Today there exist several schemes that claim to provide post-quantum security. Some of them are based on computational problems that are seemingly hard to solve even with quantum computers, like the lattice-based cryptography based on the shortest vector problem or its variants. Others are based on the assumption that there exist post-quantum-secure symmetric-key primitives, e.g. digital signatures based on one-way hash functions.

**Two Levels of Post-Quantum Security.** There are two notions of security against adversaries with quantum computers: *standard* security and *quantum* security [34]. In this paper we focus on the quantum security, because it is stronger. In the standard-security setting we assume that adversaries have quantum computers but can make only classical queries to the oracles. On the other hand, in the quantum-security setting, adversaries are allowed to make quantum superposition queries. In other words, that a scheme provides quantum security means that it will remain secure even in the far future when all computations and communications are done in quantum superposition states.

**Post-Quantum Insecurity of Symmetric-Key Constructions.** On the negative side, it has turned out that a number of symmetric-key constructions as well as many public-key schemes can be broken in polynomial time (of the block size) if adversaries are allowed to make quantum superposition queries. For example, such adversaries can distinguish 3-round Feistel ciphers from random [18], recover keys of Even-Mansour ciphers [19], forge various message authentication codes like CBC-MAC [16], by making only polynomially many queries. These attacks tell us that in general there is no guarantee that the classical security of a symmetric-key scheme implies its quantum security.

**Quantum-Secure Schemes based on One-Way Functions.** On the positive side, previous work [34,7,27] has shown that, if we assume the existence of one-way functions that are hard to invert even with quantum computers, then we can come up with a wide range of quantum-secure schemes. These include pseudo-random functions, message authentication codes, universal one-way hash functions, one-time signatures, and EU-CMA signature schemes. Thus, the existence of quantum-secure one-way functions is fundamental, just as in the classical setting, and the cryptographic hash functions in use like SHA-3 [23] and SHA-2 [22] are considered to be possible candidates also for the instantiation of these quantum-secure one-way functions.

**Cryptographic Hash Functions Revisited.** Recall that cryptographic hash functions are normally constructed only with public, “keyless” primitives, either from a public permutation or a block cipher having no secret keys (i.e. key inputs are public). For example, SHA-3 is constructed from a public permutation, and SHA-2 is essentially based on a public block cipher. The generic security (indifferentiability) of the sponge construction used in SHA-3 is proven in the random permutation model, and the security (one-wayness and collision resistance) of Davies-Meyer construction adopted by the SHA-2 compression function is proven in the ideal cipher model.

However, as mentioned above, we should carefully note that the classical provable security of these hash functions may not carry over to the quantum setting. For example, recently Carstens et al. [9] gave an evidence that SHA-3 is not indifferentiable in the quantum setting, based on a conjecture. There-

fore, here we would like to pose a fundamental question: do we have a provably quantum-secure construction of one-way hash functions?

### 1.1 Our Contributions

Our answer is positive; in this paper we show that the Merkle-Damgård iteration with the Davies-Meyer compression function is a quantum-secure one-way hash function. This has been a popular design used in MD5, SHA-1 and SHA-2. Indeed, our construction is essentially identical to the modes of operation used in these traditional hash functions, except for minor differences in padding rules, initialization vectors, and input-size restrictions on the underlying block cipher.

Our contributions come in three steps. First, we fix a security model in which we prove our main result. Second, we develop a generic tool for bounding quantum oracle indistinguishability. Finally, we use the tool to prove our main result.

1. **Introducing the Quantum Ideal Cipher Model.** As the first step we introduce the *quantum ideal cipher* model, which, as the name suggests, naturally extends the ideal cipher model in the classical setting. Similarly to the classical case, we treat the underlying block cipher as an ideal cipher  $E$ , i.e.,  $E_k$  is a random permutation for each key  $k$ . We then allow quantum adversaries to make both forward and backward queries to the cipher. In our model, a table of all values for the ideal cipher  $E$  is determined at the beginning of each game, and the oracle that computes  $E_{(\cdot)}(\cdot)$  and  $E_{(\cdot)}^{-1}(\cdot)$  are given to the adversary. Following the style of previous work in the classical setting, we consider (quantum) information-theoretic adversaries that have no limitation on computational resources, such as time or the number of available qubits. We only bound the number  $q$  of queries that the adversary makes to its oracles.
2. **A Generic Tool for Quantum Indistinguishability.** The second step is to develop a proof tool to upper-bound quantum oracle distinguishing advantages. The tool can be applied to any pair  $(D_1, D_2)$  of distributions on an arbitrary (finite) set of functions (Proposition 3.1.) The tool enables us to obtain an upper bound by mere combinatorial enumeration and associated probability computations. There is a simplified version of the tool corresponding to the special case when  $D_1$  and  $D_2$  are distributions on a set of boolean functions (having some fixed domain size) with  $D_2$  being a degenerate distribution at the zero function (Proposition 3.2.) In fact this simplified version suffices to prove our main result. Our tool is developed by generalizing and integrating several existing techniques [5,28,2,15] corresponding to some limited cases of the simplified version. However, previous work treats only the case that  $D_1$  is some specific distributions, and no previous work seems suitable to our situation. We developed our tool so that it looks familiar to researchers on symmetric-key provable security (like coefficient-H technique).
3. **One-Wayness of Merkle-Damgård with Davies-Meyer.** The final but main contribution of this paper is to give almost optimal security bound for

quantum one-wayness of the Merkle-Damgård construction with a Davies-Meyer compression function. That is, any quantum query adversary needs to make about  $2^{n/2}$  queries to invert the function with  $n$ -bit output. This bound is almost optimal since the Grover search can find a preimage of random functions with  $O(2^{n/2})$  quantum queries, and it is proven that the Grover search is optimal strategy to find a preimage of random functions [15]. In our proof, the input length of functions can be exponentially long but must be fixed. We stress that this is the first proof for quantum security on symmetric key schemes based on public block ciphers.

**Technical Details.** In this paper we give exact security bounds without any asymptotic notation, because security parameters of symmetric-key schemes are usually fixed to some constant.

This paper considers two security notions: non-invertibility and one-wayness. When we say  $h : \{0, 1\}^s \rightarrow \{0, 1\}^n$  has one-wayness, we mean that any adversary cannot find a preimage of  $y = h(x)$ , where  $x$  is randomly chosen from  $\{0, 1\}^s$ .<sup>1</sup> On the other hand, when we say  $h$  has non-invertibility, we mean that any adversary cannot find a preimage of  $y$ , where  $y$  is randomly chosen from  $\{0, 1\}^n$ . These are similar but independent notions.

We firstly show non-invertibility of permutation with feedforward in the quantum ideal permutation model, secondly show both non-invertibility and one-wayness of Davies-Meyer constructions, and finally show both non-invertibility and one-wayness of Merkle-Damgård constructions. It might be unexpected that permutation with feedforward is non-invertible in the quantum setting although it uses only public permutation and XOR operation, which seems similar to the Even-Mansour ciphers that are broken by quantum superposition attacks.

Due to a technical reason, we need some restriction on usage of keys in Davies-Meyer construction. Similarly, we need a padding function for Merkle-Damgård construction. However, these do not mean restriction on available block ciphers. As a subsidiary result, we also show that any quantum query adversary needs to make about  $2^{n/2}$  queries to find a fixed point of a public random permutation (which allow adversaries to make both forward and backward quantum queries). This is the first result on quantum query lower bound for a property related to public random permutations.

Our proof strategy is to reduce the problem of breaking security notions to the problem of distinguishing oracle distributions on boolean functions. A similar strategy can be found in [15]. Then indistinguishability between quantum oracle distributions is shown using our new proof tool described above. To reduce problems on public random permutations to problems on boolean functions, we try to approximate the uniform distribution on random permutations by combining distributions on boolean functions with the uniform distribution on derangements (permutations without fixed points).

---

<sup>1</sup> This security notion is also called *preimage resistance* (see [25] for example).

## 1.2 Related Work

There already exist powerful tools that aim to give quantum security bounds for cryptographic schemes. These tools include “one-way to hiding” lemma and quantum random oracle programming by Unruh [30,29], the rank method and oracle indistinguishability frameworks by Zhandry [34,35,7]. These tools do not seem to consider the situation where adversaries can make both forward and backward queries to public permutations or block ciphers. There exists previous work [1] that proves quantum security of Even-Mansour ciphers in a model where adversaries make both forward and backward queries to the underlying permutation, but it should be noted that the proof [1] requires a quantum computational hardness assumption (the hidden shift problem.)

A quantum version of the random oracle model is proposed by Boneh et al, [6], and many schemes are proven to be secure in this model ([35,29], for example). Regarding symmetric key schemes, several papers on quantum security already exist. They include work on quantum security of Carter-Wegman MACs [7], quantum PRP-PRF switching lemma [36], quantum security of the CBC, OFB, CTR, and XTS modes of operation [3], quantum generic security of random hash functions [15], and quantum security of NMAC [28]. With a computational assumption that *hidden shift problem* is hard to solve even with quantum computers, it is shown that Even-Mansour ciphers and CBC-MAC, which are broken in polynomial time with quantum queries, can be modified to have quantum security [1]. For standard security, i.e., with the assumption that adversaries have quantum computers but can make only classical queries, XOR of PRPs are proven to be secure [20]. Unruh introduced a security notion named *collapsing*, which is a generalized notion of collision-resistant in the quantum setting [32]. Unruh showed that Merkle-Damgård constructions are collapsing if underlying constructions are collapsing [31]. Czajkowski et al. showed that sponge constructions are also collapsing [10] (Note that they assume building permutations are one-way permutations or functions, and do not treat the usual sponge functions that are constructed from public permutations). Recently Zhandry [33] showed indifferentiability of the Merkle-Damgård construction in the quantum random oracle model (compression functions are assumed to be random functions).

## 2 Preliminaries

In this section we describe notation and definitions. For readers who are not familiar with quantum terminology, we give a brief explanation on quantum computation in Appendix C.

**Notation.** Let  $[i, \dots, j]$  denote the set of integers  $\{i, i + 1, \dots, j\}$  for  $i < j$ , and  $[N]$  denote the set  $[1, \dots, N]$ . For sets  $X$  and  $Y$ , let  $\text{Func}(X, Y)$  be the set of functions from  $X$  to  $Y$ . For a set  $X$ , let  $\text{Perm}(X)$  be the set of permutations

on  $X$ . Let  $\text{Ciph}(m, n)$  denote the set

$$\{E \in \text{Func}(\{0, 1\}^m \times \{0, 1\}^n, \{0, 1\}^n) \mid E(k, \cdot) \in \text{Perm}(\{0, 1\}^n) \text{ for each } k\},$$

where “ $\cdot$ ” means arbitrary inputs.

We call an element of  $\text{Ciph}(m, n)$  an  $n$ -bit block cipher with an  $m$ -bit key. For each  $E \in \text{Ciph}(m, n)$  and  $k \in \{0, 1\}^m$ , let  $E_k$  denote the permutation  $E(k, \cdot)$ . For a distribution  $D$ , let  $\Pr_{x \sim D}[\text{event}]$  denote the probability that event occurs when  $x$  is sampled according to the distribution  $D$ . For two distributions  $D_1$  and  $D_2$ , let  $\Delta(D_1, D_2)$  denote the total variation distance  $D_1$  and  $D_2$ . Let  $\text{td}(\rho_1, \rho_2)$  denote the trace distance between density matrices  $\rho_1$  and  $\rho_2$ . For a random variable  $V$  that takes values in a set  $X$ , define a distribution  $D_V : X \rightarrow [0, 1]$  by  $D_V(x) = \Pr[V = x]$  for each  $x \in X$ . We call  $D_V$  the distribution of  $V$ . If we write  $x \stackrel{D}{\leftarrow} X$ , then it means to sample  $x$  according to the distribution  $D$  on  $X$ .

**Derangements.** A permutation  $P_0 \in \text{Perm}(X)$  is called a *derangement* if  $P_0$  has no fixed point, i.e. if there is no element  $x \in X$  such that  $P_0(x) = x$ . The set of derangements on a set  $X$  is denoted as  $\text{Der}(X)$ . The number of derangements on a set of size  $N$  is written as  $!N$ . The following formula is well-known [14]:

**Lemma 2.1.** *We have  $!N = N! \cdot \sum_{i=0}^N \frac{(-1)^i}{i!} = \lfloor \frac{N!}{e} + \frac{1}{2} \rfloor$ , where  $\lfloor \cdot \rfloor$  is the floor function.*

For completeness, we give a proof in Appendix B.

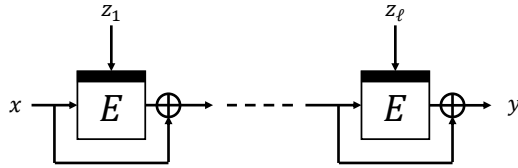
**Davies-Meyer and Merkle-Damgård Constructions.** For an  $n$ -bit block cipher  $E$  with an  $m$ -bit key, we define a function  $\text{DM}^E \in \text{Func}(\{0, 1\}^m \times \{0, 1\}^n, \{0, 1\}^n)$  by  $\text{DM}^E(z, x) = E_z(x) \oplus x$ . We call  $\text{DM}^E$  the *Davies-Meyer construction* made from  $E \in \text{Ciph}(m, n)$ . For a permutation  $P \in \text{Perm}(\{0, 1\}^n)$ , we define a function  $\text{FF}^P \in \text{Func}(\{0, 1\}^n)$  by  $\text{FF}^P(x) := P(x) \oplus x$ . We call the function  $\text{FF}^P$  as *permutation  $P$  with feedforward*. The function  $\text{FF}$  can be regarded as a “fixed-key” version of  $\text{DM}$ .

For a function  $h : \{0, 1\}^m \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  and an integer  $\ell > 0$ , the *Merkle-Damgård construction*  $\text{MD}_\ell^h : \{0, 1\}^m \times \{0, 1\}^{m\ell} \rightarrow \{0, 1\}^n$  is defined by

$$\text{MD}_\ell^h(x, z_1, \dots, z_\ell) := h(z_\ell, h(z_{\ell-1}, \dots, h(z_2, h(z_1, x)) \dots)), \quad (1)$$

where  $z_i \in \{0, 1\}^m$  for each  $i$ . We consider the special case when  $h$  is the Davies-Meyer compression function, i.e.,  $h(z, x) = \text{DM}^E(z, x)$  for an  $n$ -bit block cipher  $E \in \text{Ciph}(m, n)$ . Fig. 1 illustrates  $\text{MD}_\ell^{\text{DM}^E}$ , the combination of a Davies-Meyer compression function with the Merkle-Damgård iteration.

**Quantum oracles and quantum adversaries.** For a function  $f \in \text{Func}(\{0, 1\}^a, \{0, 1\}^b)$ , quantum oracle of  $f$  is defined as the unitary operator  $O_f$  such that  $O_f |x\rangle |y\rangle = |x\rangle |y \oplus f(x)\rangle$  for arbitrary  $x \in \{0, 1\}^a, y \in \{0, 1\}^b$ . By an abuse of



**Fig. 1.** The Merkle-Damgård construction with a Davies-Meyer compression function

notation, let  $O_f$  also denote the  $(a + b + c)$ -qubit unitary operator  $O_f \otimes I_c$  that maps  $|x\rangle|y\rangle|z\rangle$  to  $|x\rangle|y \oplus f(x)\rangle|z\rangle$  for any  $c$ .

This paper discusses on information theoretic quantum query adversary. That is, we fix a constant  $q$  and assume that a quantum adversary  $\mathcal{A}$  can make at most  $q$  quantum queries, but we assume no other limitation for  $\mathcal{A}$  about quantum computational resources such as time or the number of available qubits. Following the previous works that treat quantum oracle query adversary ([4,6,35,34,7,28], for example), we model  $\mathcal{A}$  as a sequence of unitary operators  $U_q O_f U_{q-1} \cdots O_f U_0$ . We write  $\mathcal{A}^O(x) = y$  for the event that a quantum adversary  $\mathcal{A}$  takes  $x$  as input, makes quantum queries to  $O$ , and finally outputs  $y$ .

If quantum oracle  $O$  is dependent on some distribution, then the state of a quantum query algorithm  $\mathcal{A}$  is described as a density operator. Suppose  $O = O_f$  for a function  $f$ , which is sampled according to a distribution  $D_1$  on  $\text{Func}(\{0, 1\}^a, \{0, 1\}^b)$ . Then, the state of  $\mathcal{A}$  with input  $x$  after the  $i$ -th query becomes  $|\phi_f^i\rangle := U_i O_f U_{i-1} O_f \cdots O_f U_0 |0, x, 0\rangle$  with probability  $p_1^f := \Pr_{F \sim D_1}[F = f]$ . This mixed state is described as

$$\rho_1^i = \sum_f p_1^f |\phi_f^i\rangle \langle \phi_f^i|. \quad (2)$$

**Quantum oracle distinguishing advantage.** Following previous works (see [34], for example), we define quantum oracle distinguishing advantage as follows. Let  $D_1, D_2$  be two distributions on a set of functions. Assume that a quantum algorithm  $\mathcal{A}$  is allowed to access the quantum oracle of a function that is chosen according to either  $D_1$  or  $D_2$ . Suppose  $\mathcal{A}$  can make at most  $q$  queries, and finally outputs the result 1 or 0. Then, we define the distinguishing advantage of  $\mathcal{A}$  by

$$\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) := \left| \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 1] - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 1] \right|.$$

In addition, we define

$$\text{Adv}_{D_1, D_2}^{\text{dist}}(q) := \max_{\mathcal{A}} \left\{ \text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \right\},$$

where the maximum is taken over all quantum-query algorithms, each making at most  $q$  quantum queries.

Distinguishing advantages can be bounded by the trace distance and total variational distance. Let  $\rho_1^i$  be the density operator defined by (2), and  $\rho_2^i$  be the density operator that is similarly defined according to the distribution  $D_2$ . Then we can show the following lemma:

**Lemma 2.2.** *For any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  queries,*

$$\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq \text{td}(\rho_1^q, \rho_2^q) \quad (3)$$

and

$$\text{td}(\rho_1^q, \rho_2^q) \leq \Delta(D_1, D_2) \quad (4)$$

hold.

The inequality (4) trivially follows from definitions and the proof of inequality (3) is also straightforward, but we give a proof in Appendix C.1 for readers who are not used to quantum computation.

## 2.1 Modeling Public Random Permutations and Block Ciphers in the Quantum Setting

To model public ideal permutations and block ciphers, here we introduce quantum ideal permutation model and quantum ideal cipher model, which are quantum versions of the classical ideal permutation model and ideal cipher model, respectively. There already exist works on quantum provable security [1] in the models that are essentially same to our quantum random permutation model. However, this is the first paper on provable security that treats ideal cipher model in the quantum setting. We begin with formalizing quantum oracles of public permutations and block ciphers, and then introduce quantum ideal permutation model and quantum ideal cipher model.

**Quantum oracles of public permutations and ciphers.** Here we describe how to formalize quantum oracles of public permutations and block ciphers. For an  $n$ -bit public permutation  $P$ , we define a function  $P^\pm : \{0, 1\} \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  by

$$P^\pm(b, x) = \begin{cases} P(x) & \text{if } b = 0, \\ P^{-1}(x) & \text{if } b = 1. \end{cases}$$

For a distribution  $D$  on  $\text{Perm}(\{0, 1\}^n)$ , let  $D^\pm$  be the associated distribution on  $\text{Func}(\{0, 1\} \times \{0, 1\}^n, \{0, 1\}^n)$  defined by  $D^\pm(f) = \Pr_{P \sim D}[P^\pm = f]$ . For any public permutation  $P$ , we assume that the quantum oracle  $O_{P^\pm}$  is available. This models the situation that both of forward and backward quantum queries to the public permutation  $P$  are allowed.

Similarly, if  $E$  is an  $n$ -bit block cipher with  $m$ -bit key, then we define a function  $E^\pm : \{0, 1\} \times \{0, 1\}^m \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  by

$$E^\pm(b, k, x) = \begin{cases} E_k(x) & \text{if } b = 0, \\ E_k^{-1}(x) & \text{if } b = 1. \end{cases}$$



For a distribution  $D$  on  $\text{Ciph}(m, n)$ , let  $D^\pm$  be the associated distribution on  $\text{Func}(\{0, 1\} \times \{0, 1\}^m \times \{0, 1\}^n, \{0, 1\}^n)$  defined by  $D^\pm(f) = \Pr_{E \sim D}[E^\pm = f]$ . For any public block cipher  $E$ , we assume that the quantum oracle  $O_{E^\pm}$  is available. This models the situation that both of forward and backward quantum queries to a block cipher  $E$  are allowed.

**Quantum ideal permutation model.** Assume that  $P$  is a public permutation which is chosen from  $\text{Perm}(\{0, 1\}^n)$  uniformly at random, and an adversary  $\mathcal{A}$  is allowed to make at most  $q$  quantum queries to  $P^\pm$ , for some fixed number  $q$ . We call this model as *quantum ideal permutation model*. We say that a scheme constructed from a public permutation is secure (with regard to some quantum security notion) up to  $q$  quantum queries if no such quantum information theoretic adversary can break the security notion. We say that  $P$  is an ideal permutation if we assume the situation that quantum adversaries can access quantum oracle of  $P$ , and  $P$  is chosen from  $\text{Perm}(\{0, 1\}^n)$  uniformly at random.

**Quantum ideal cipher model.** Assume that  $E$  is a public block cipher which is chosen from  $\text{Ciph}(m, n)$  uniformly at random, and an adversary  $\mathcal{A}$  is allowed to make at most  $q$  quantum queries to  $E^\pm$ , for some fixed number  $q$ . We call this model as *quantum ideal cipher model*. Security in this model is defined similarly as in the quantum ideal permutation model. Similarly, we say that  $E$  is an ideal cipher if we assume the situation that quantum adversaries can access quantum oracle of  $E$ , and  $E$  is chosen from  $\text{Ciph}(m, n)$  uniformly at random.

## 2.2 Two Security Notions: Non-Invertibility and One-Wayness.

This paper considers two security notions: non-invertibility and one-wayness. These are similar but independent notions (we give a separation proof in Appendix D for completeness). Let  $h^F : \{0, 1\}^s \rightarrow \{0, 1\}^n$  be a function that is constructed from a function (or permutation)  $F$ , and  $O$  be a quantum oracle that is defined depending on  $F$ . We assume  $F$  is chosen from a set of functions  $\mathcal{S}_F$  uniformly at random. The set  $\mathcal{S}_F$  and how the oracle  $O$  is related to  $F$  depend on security models.

If we consider the quantum ideal permutation model, then  $\mathcal{S}_F = \text{Perm}(\{0, 1\}^n)$ , and  $O$  is defined as the oracle of  $P^\pm$ . We will consider the case that  $h^F$  is a permutation with feedforward. Similarly, if we consider the quantum ideal cipher model, then  $\mathcal{S}_F = \text{Ciph}(m, n)$ , and  $O$  is defined as the oracle of  $E^\pm$ . We will consider the case that  $h^F$  is the Davies-Meyer constructions or Merkle-Damgård constructions.

*Non-invertibility.* For any quantum oracle query adversary  $\mathcal{A}$ , define the advantage of  $\mathcal{A}$  to invert the function  $h^F$  by

$$\text{Adv}_{h^F}^{\text{inv}}(\mathcal{A}) := \Pr_{F, y}[\mathcal{A}^O(y) = x \wedge h^F(x) = y], \quad (5)$$

where  $F \in \mathcal{S}_F$  and  $y \in \{0, 1\}^n$  are chosen uniformly at random. In addition, we define

$$\text{Adv}_{h^F}^{inv}(q) := \max_{\mathcal{A}} \{\text{Adv}_{h^F}^{inv}(\mathcal{A})\}, \quad (6)$$

where the maximum is taken over all quantum-query algorithms, each making at most  $q$  quantum queries.

*One-wayness.* Similarly, define the advantage of  $\mathcal{A}$  to break the one-wayness of the function  $h^F$  by

$$\text{Adv}_{h^F}^{ow}(\mathcal{A}) := \Pr_{F, x'} [\mathcal{A}^O(h^F(x')) = x \wedge h^F(x) = h^F(x')], \quad (7)$$

where  $F \in \mathcal{S}_F$  and  $x' \in \{0, 1\}^s$  are chosen uniformly at random. In addition, we define

$$\text{Adv}_{h^F}^{ow}(q) := \max_{\mathcal{A}} \{\text{Adv}_{h^F}^{ow}(\mathcal{A})\}, \quad (8)$$

where the maximum is taken over all quantum-query algorithms, each making at most  $q$  quantum queries.

*Trivial upper bounds.* We note here that there are trivial upper bounds of quantum query complexity for non-invertibility and one-wayness, if  $h^F$  is sufficiently random. The bound is given by simple application of the Grover search or its generalizations [13,8]. Given  $y$ , let consider to find  $x$  such that  $h^F(x) = y$ . Then, if  $2^s / |(h^F)^{-1}(y)| \approx 2^n$ , (which is the case when  $h^F$  is a truly random function and message space  $\{0, 1\}^s$  is much larger than range  $\{0, 1\}^n$ ) then we can find  $x$  such that  $h^F(x) = y$  with about  $\sqrt{2^n}$  quantum queries to  $h^F$ . We say  $h^F$  is almost optimally non-invertible or one-way if  $\text{Adv}_{h^F}^{inv}(q) = \tilde{O}(q/\sqrt{2^n})$  or  $\text{Adv}_{h^F}^{ow}(q) = \tilde{O}(q/\sqrt{2^n})$ , respectively, since these imply that there is no way which is significantly better than the generic attack (the Grover search) to break one-wayness of  $h^F$ .

### 3 A Tool for Quantum Oracle Indistinguishability

Here we give a tool to upper bound quantum oracle distinguishing advantages  $\text{Adv}_{D_1, D_2}^{dist}$  with only classical probability calculation and purely combinatorial enumeration (Proposition 3.1). Our tool can be applied to *any* distributions  $D_1, D_2$  on *any* (finite) set of functions  $\text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ . In later sections, to show non-invertibility and one-wayness of functions, we treat only the cases that  $c = 1$  and  $D_2$  is the degenerate distribution with support on the zero function. Our tool can be somewhat simplified in those cases, and thus we give a simplified version of our tool (Proposition 3.2) for later use. We believe that the generalized version (Proposition 3.1) itself is also useful to give some quantum security bound for other schemes or other security notions. To show that the generalized version is also useful, an application is given in Appendix F.

There already exist techniques to bound quantum oracle distinguishing advantages in the situations which are similar to our simplified version ( $c = 1$  and

$D_2$  is the degenerate distribution with support on the zero function), but existing works treat only the case that  $D_1$  is some specific distributions. (See proof of Lemma 37 in [2], proof of Lemma C.1 in [28], for example. Theorem 1 in [15] gives similar result as Lemma 37 in [2], but uses different analyzing technique by Zhandry [34].) On the other hand, our simplified tool (Proposition 3.2) enables us to treat *any* distribution  $D_1$  on a (finite) set of boolean functions.

This section is organized as follows. First, we explain our motivations to develop quantum proof tools. Second, we describe our main tool. Third, we briefly explain how to apply them to give quantum security bounds in later sections.

### 3.1 Motivations: the coefficient H technique

In the classical setting, there exist several proof tools to prove oracle indistinguishability of symmetric key schemes. The *coefficient-H technique* developed by Patarin [24] is one of the most powerful tools. Below we explain essence of the technique.

Suppose we want to upper bound  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A})$  for a (classical) information theoretic adversary  $\mathcal{A}$ , and distributions  $D_1, D_2$ . The technique allows  $\mathcal{A}$  to obtain *transcripts* including all input-output pairs defined by queries. Let  $\mathsf{T}_1, \mathsf{T}_2$  be the transcripts that correspond to the oracle distributions  $D_1$  and  $D_2$ , respectively. Then,  $\mathsf{T}_1, \mathsf{T}_2$  define distributions on a set of transcript  $\mathcal{T}$ . The coefficient-H technique divides  $\mathcal{T}$  into a *good set* **good** and *bad set* **bad**. Roughly speaking, the technique gives a bound  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq \epsilon + \Pr[\mathsf{T}_2 \in \text{bad}]$ . The parameter  $\epsilon$  is a small number that satisfies  $\Pr[\mathsf{T}_1 = \tau] / \Pr[\mathsf{T}_2 = \tau] \geq 1 - \epsilon$  for any good transcript  $\tau \in \text{good}$ . How good bound we can achieve depends on how well we define the set of transcripts  $\mathcal{T}$ , good sets **good**, and bad sets **bad**.

### 3.2 Our Main tool

Following the classical coefficient-H technique, we aim to develop a quantum proof tool so that: 1. It uses some good and bad sets, and 2. It gives an upper bound as a sum of an amount related to good events (like  $\epsilon$  in the coefficient-H technique), and a bad probability. In addition, we make our tool so that we can obtain an upper bound with only classical probability calculation and purely combinatorial enumeration. We first describe a generalized version that  $D_1$  and  $D_2$  can be any distributions, and then explain how it is simplified in the case  $c = 1$  and  $D_2$  is the degenerate distribution.

**Generalized version.** Let  $D_1, D_2$  be *any* distributions on *any* (finite) set of functions  $\text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ . In addition, let  $\bar{D}$  be an *arbitrary* distribution on the product space  $\text{Func}(\{0, 1\}^n, \{0, 1\}^c) \times \text{Func}(\{0, 1\}^n, \{0, 1\}^c)$  that satisfies

$$D_1(f) = \sum_g \bar{D}(f, g) \text{ for any } f \wedge D_2(g) = \sum_f \bar{D}(f, g) \text{ for any } g. \quad (9)$$

(In applications, even though  $D_1$  and  $D_2$  are given as independent distributions, we try to find a convenient distribution  $\bar{D}$ , just like we do so in the (classical) game-playing proof technique. See Appendix F for a concrete example.)

For each  $f, g \in \text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ , let  $p_1^f, p_2^g, p^{f \cdot g}$  denote  $\Pr_{F \sim D_1}[F = f]$ ,  $\Pr_{G \sim D_1}[G = g]$ , and  $\Pr_{(F, G) \sim \bar{D}}[(F, G) = (f, g)]$ , respectively. In addition, define a boolean function  $\delta(f, g) : \{0, 1\}^n \rightarrow \{0, 1\}$  by  $\delta(f, g)(x) = 1$  if and only if  $f(x) \neq g(x)$  for each pair  $(f, g)$ . Let  $\mathbf{0} \in \text{Func}(\{0, 1\}^n, \{0, 1\})$  be the zero function that maps  $x$  to 0 for any  $x$ . For each  $g \in \text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ , let  $\delta D|_g$  be the conditional distribution on  $\text{Func}(\{0, 1\}^n, \{0, 1\})$  defined by  $(\delta D|_g)(\gamma) = \Pr_{(F, G) \sim \bar{D}}[\delta(F, G) = \gamma | G = g]$  for any  $\gamma \in \text{Func}(\{0, 1\}^n, \{0, 1\})$ .

For each  $g \in \text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ , take a “bad” set  $\text{bad}^g \subset \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus \{\mathbf{0}\}$  arbitrarily (actually we select  $\text{bad}^g$  such that  $\Pr_{\Gamma \sim \delta D|_g}[\Gamma \in \text{bad}^g]$  is small), and define “good” set by  $\text{good}^g := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus (\{\mathbf{0}\} \cup \text{bad}^g)$ . Furthermore, decompose the good set  $\text{good}^g$  into smaller subsets  $\{\text{good}_\alpha^g\}_{\alpha \in A_g}$  (i.e.  $\text{good}^g = \bigcup_\alpha \text{good}_\alpha^g$  and  $\text{good}_\alpha^g \cap \text{good}_\beta^g = \emptyset$  for  $\alpha \neq \beta$ ) such that the conditional probability  $\Pr_{\Gamma \sim \delta D|_g}[\Gamma = \gamma | \Gamma \in \text{good}_\alpha^g]$  is independent of  $\gamma$  (in other words, for each  $\alpha \in A_g$ ,  $\Pr_{\Gamma \sim \delta D|_g}[\Gamma = \gamma] = \Pr_{\Gamma \sim \delta D|_g}[\Gamma = \gamma']$  holds for  $\gamma, \gamma' \in \text{good}_\alpha^g$ ). In addition, define  $\text{bad}_{all} \subset (\text{Func}(\{0, 1\}^n, \{0, 1\}^c))^2$  by  $\text{bad}_{all} := \{(f, g) | \delta(f, g) \in \text{bad}^g\}$ . For each  $g, \alpha \in A_g$  and  $\gamma \in \text{Func}(\{0, 1\}^n, \{0, 1\})$ , let  $p_{\delta D|_g}^{\text{good}_\alpha^g} := \Pr_{\Gamma \sim \delta D|_g}[\Gamma \in \text{good}_\alpha^g]$  and  $p_{\delta D|_g}^{\gamma | \text{good}_\alpha^g} := \Pr_{\Gamma \sim \delta D|_g}[\Gamma = \gamma | \Gamma \in \text{good}_\alpha^g]$  (by assumption,  $p_{\delta D|_g}^{\gamma | \text{good}_\alpha^g}$  is independent of  $\gamma$ ). Then the following proposition holds.

**Proposition 3.1 (Generalized version).** *Let  $D_1, D_2$  be any distributions on  $\text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ , and  $\bar{D}$  be any distribution that satisfies (9). Let  $\text{bad}_{all}$ ,  $\text{bad}^g$ ,  $\text{good}^g$ , and  $\{\text{good}_\alpha^g\}_{\alpha \in A_g}$  be the sets as stated above. Then, for any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  quantum queries,  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A})$  is upper bounded by*

$$2q \cdot \mathbf{E}_{G \sim D_2} \left[ \sum_{\alpha \in A_G} p_{\delta D|_G}^{\text{good}_\alpha^G} \sqrt{p_{\delta D|_G}^{\gamma | \text{good}_\alpha^G} \cdot \max_x |\{\gamma \in \text{good}_\alpha^G | \gamma(x) = 1\}|} \right] + 2q \cdot \Pr_{(F, G) \sim \bar{D}}[(F, G) \in \text{bad}_{all}]. \quad (10)$$

A proof of this proposition is given in Appendix E.

In later sections, we apply our tool only to the cases that  $c = 1$  and  $D_2$  is the degenerate distribution with support on the zero function  $\mathbf{0}$ . Description of our tool can be somewhat simplified in such cases, and below we give the simplified version for later use. To show that the generalized version itself is also useful, an application of Proposition 3.1 is given in Appendix F.

**Simplified version.** Now we describe a simplified version of our tool. Let  $D_1, D_2$  be distributions on a set of boolean functions  $\text{Func}(\{0, 1\}^n, \{0, 1\})$ , and  $D_2$  be the degenerate distribution with support on the zero function  $\mathbf{0}$ .  $D_1$  can be any distribution.

Take a “bad” set  $\text{bad} \subset \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus \{\mathbf{0}\}$  arbitrarily (actually we select  $\text{bad}$  such that  $\Pr_{F \sim D_1}[F \in \text{bad}]$  will be small), and define “good” set by  $\text{good} := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus (\{\mathbf{0}\} \cup \text{bad})$ . Furthermore, decompose the good set  $\text{good}$  into smaller subsets  $\{\text{good}_\alpha\}_\alpha$  (i.e.  $\text{good} = \bigcup_\alpha \text{good}_\alpha$  and  $\text{good}_\alpha \cap \text{good}_\beta = \emptyset$  for  $\alpha \neq \beta$ ) such that the conditional probability  $\Pr_{F \sim D_1}[F = f | F \in \text{good}_\alpha]$  is independent of  $f$  (in other words, for each  $\alpha$ ,  $\Pr_{F \sim D_1}[F = f] = \Pr_{F \sim D_1}[F = f']$  holds for  $f, f' \in \text{good}_\alpha$ ). Let  $p_1^{\text{good}_\alpha} := \Pr_{F \sim D_1}[F \in \text{good}_\alpha]$  and  $p_1^{f|\text{good}_\alpha} := \Pr_{F \sim D_1}[F = f | F \in \text{good}_\alpha]$  (by assumption,  $p_1^{f|\text{good}_\alpha}$  is independent of  $f$ ). Then, the following proposition holds, which enables us to bound advantages of quantum adversaries with only classical probability calculations and purely combinatorial enumeration, without any quantum arguments.

**Proposition 3.2 (Simplified version).** *Let  $D_1$  be any distribution on the set of boolean functions  $\text{Func}(\{0, 1\}^n, \{0, 1\})$ , and  $D_2$  be the degenerate distribution with support on the zero function. Let  $\text{bad}$ ,  $\text{good}$ , and  $\{\text{good}_\alpha\}_\alpha$  be the subsets of  $\text{Func}(\{0, 1\}^n, \{0, 1\})$  as stated above. Then, for any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  quantum queries,  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A})$  is upper bounded by*

$$2q \sum_{\alpha} p_1^{\text{good}_\alpha} \sqrt{p_1^{f|\text{good}_\alpha} \cdot \max_x |\{f \in \text{good}_\alpha \mid f(x) = 1\}|} + 2q \Pr_{F \sim D_1}[F \in \text{bad}]. \quad (11)$$

This proposition follows as an immediate corollary of the generalized version Proposition 3.1 as below.

*Proof (of Proposition 3.2).* Now,  $D_1$  and  $D_2$  are distributions on a set of boolean functions  $\text{Func}(\{0, 1\}^n, \{0, 1\})$ , and  $D_2$  is the degenerate distribution with support on the zero function  $\mathbf{0}$ . Let  $\text{bad}$ ,  $\text{good}$ , and  $\{\text{good}_\alpha\}_\alpha$  be the sets in Proposition 3.2.

We translate notations in Proposition 3.2 to those in Proposition 3.1. Let  $\bar{D}$  be the product distribution  $D_1 \times D_2$ . Let  $\text{bad}^g := \emptyset$ ,  $\text{good}_\alpha^g := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus \{\mathbf{0}\}$  for  $g \neq \mathbf{0}$ , and  $\text{bad}^{\mathbf{0}} := \text{bad}$ ,  $\text{good}_\alpha^{\mathbf{0}} := \text{good}_\alpha$ .

Then,  $\delta(f, \mathbf{0}) = f$  holds for any boolean function  $f$ ,  $\Pr_{G \sim D_2}[G = g] = 1$  holds if and only if  $g = \mathbf{0}$ , and  $\delta D|_{\mathbf{0}} = D_1$  holds. In addition, we have  $p_{\delta D|_{\mathbf{0}}}^{\text{good}_\alpha^{\mathbf{0}}} = p_1^{\text{good}_\alpha}$ , and  $p_{\delta D|_{\mathbf{0}}}^{f|\text{good}_\alpha^{\mathbf{0}}} = p_1^{f|\text{good}_\alpha}$  for any boolean function  $f$ . Moreover,  $\text{bad}_{\text{all}} = \{(f, \mathbf{0}) \mid f \in \text{bad}^{\mathbf{0}}\}$  holds, which implies that  $\Pr_{(F, G) \sim \bar{D}}[(F, G) \in \text{bad}_{\text{all}}] = \Pr_{F \sim D_1}[F \in \text{bad}]$ . Therefore Proposition 3.2 follows from Proposition 3.1.  $\square$

*Remark 3.1.* We do not claim that our tool is all-around. Actually the condition that the probability  $p_{\delta D|_g}^{\gamma|\text{good}_\alpha^g}$  is independent of  $\gamma$  (in the generalized version) and  $p_1^{f|\text{good}_\alpha}$  is independent of  $f$  (in the simplified version) implicitly means that  $D_1$  must have some “uniform” structure to obtain a good bound with our tool. See proofs of Lemma 4.3 and Lemma 5.1 for concrete examples.

### 3.3 How to give quantum security bound with our tool

Next, we describe how we apply Proposition 3.2 in later sections to give quantum security bounds, in a high-level fashion. Roughly speaking, we try to reduce a target problem to a problem of bounding distinguishing advantage between two distributions on a set of boolean functions, and then apply Proposition 3.2. This strategy itself is not new, but we believe our tool enables us to take the strategy for wider applications.

Let  $\mathcal{A}$  be a quantum query algorithm, and suppose that a problem to give a security proof is reduced to a problem to upper bound some distinguishing advantage  $\text{Adv}_{G_{\text{real}}, G_{\text{ideal}}}^{\text{dist}}(\mathcal{A})$ . We introduce intermediate distributions (i.e. intermediate games)  $G_1 = G_{\text{ideal}}, G_2, \dots, G_t = G_{\text{real}}$  such that  $\text{Adv}_{G_i, G_{i+1}}^{\text{dist}}(\mathcal{A})$  can be bounded using other techniques for  $1 \leq i \leq t-2$ . In addition, we assume  $\text{Adv}_{G_{t-1}, G_t}^{\text{dist}}(\mathcal{A})$  can be bounded by  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{B})$  for some distributions  $D_1, D_2$  on  $\text{Func}(\{0, 1\}^n, \{0, 1\})$ , and another quantum query algorithm  $\mathcal{B}$ . Then we have

$$\begin{aligned} \text{Adv}_{G_{\text{real}}, G_{\text{ideal}}}^{\text{dist}}(\mathcal{A}) &\leq \text{Adv}_{G_{t-1}, G_t}^{\text{dist}}(\mathcal{A}) + \sum_{i=1}^{t-2} \text{Adv}_{G_i, G_{i+1}}^{\text{dist}}(\mathcal{A}) \\ &\leq \text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{B}) + \sum_{i=1}^{t-2} \text{Adv}_{G_i, G_{i+1}}^{\text{dist}}(\mathcal{A}) \end{aligned} \quad (12)$$

Hence, if  $\text{Adv}_{G_i, G_{i+1}}^{\text{dist}}(\mathcal{A})$  can be upper bounded by other approaches for  $1 \leq i \leq t-2$ , then the remaining term can be bounded without any quantum argument, by using our tool. In later sections, we will upper bound  $\text{Adv}_{G_i, G_{i+1}}^{\text{dist}}(\mathcal{A})$  by total variation distance  $\Delta(G_i, G_{i+1})$ . (Remember that  $\text{Adv}_{D, D'}^{\text{dist}}(\mathcal{A}) \leq \Delta(D, D')$  holds for any distributions  $D$  and  $D'$  from Lemma 2.2.) Thus we upper bound the advantage  $\text{Adv}_{G_{\text{real}}, G_{\text{ideal}}}^{\text{dist}}(\mathcal{A})$  by purely combinatorial enumerating arguments.

## 4 Non-invertibility of Permutation with Feedforward in the Quantum Ideal Permutation Model

Now we apply the technique of Section 3 to show that permutation with feedforward is optimally non-invertible in the ideal permutation model. As one step in our proof, we also prove the difficulty to find a fixed point of random permutations (Proposition 4.1). We stress that this is the first results on quantum query lower bound for some property of random permutation  $P$  or some scheme constructed from  $P$ , in the model that both of forward and backward queries to permutation  $P$  are allowed. The goal of this section is to prove the following theorem.

**Theorem 4.1.** *Let  $n \geq 32$ . For any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  forward or backward queries to a public permutation  $P$ ,*

$$\text{Adv}_{\text{FFP}}^{\text{inv}}(\mathcal{A}) \leq \frac{4(e+1)(q+1)}{2^{n/2}} + \epsilon(n) \quad (13)$$

holds, where  $\epsilon(n) = \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!}$ . In particular,  $\mathcal{A}$  cannot invert  $\text{FF}^P$  with constant probability for  $q \ll 2^{n/2}$ .

*Remark 4.1.* We need the condition  $n \geq 32$  for technical reasons. This assumption is reasonable since block lengths of block ciphers usually satisfy it.

To show the above theorem, we begin with reducing the problem of finding a preimage of permutation with feedforward in the ideal permutation model to the problem of finding a fixed point of an ideal permutation. Let us define the advantage of a quantum algorithm  $\mathcal{A}$  to find a fixed point of an ideal permutation by

$$\text{Adv}_P^{\text{fixpt}}(\mathcal{A}) := \Pr[\mathcal{A}^{O_{P^\pm}}() = x \wedge P(x) = x],$$

here  $P$  is chosen uniformly at random, and

$$\text{Adv}_P^{\text{fixpt}}(q) := \max_{\mathcal{A}} \left\{ \text{Adv}_P^{\text{fixpt}}(\mathcal{A}) \right\},$$

where the maximum is taken over all quantum-query algorithms, each making at most  $q$  quantum queries.

**Lemma 4.1.** *For a quantum algorithm  $\mathcal{A}$  that makes at most  $q$  quantum queries to  $O_{P^\pm}$ , there exists a quantum algorithm  $\mathcal{B}$  that makes at most  $q$  quantum queries to  $O_{P^\pm}$  such that  $\text{Adv}_{\text{FF}^P}^{\text{inv}}(\mathcal{A}) = \text{Adv}_P^{\text{fixpt}}(\mathcal{B})$ .*

*Proof.* Given such algorithm  $\mathcal{A}$ , we construct  $\mathcal{B}$  with the desired properties. Firstly, before making queries,  $\mathcal{B}$  chooses  $y \in \{0, 1\}^n$  uniformly at random.  $\mathcal{B}$  is given the oracle  $O_{P^\pm}$  of the permutation  $P$ . Define another permutation  $P'$  by  $P'(x) = P(x) \oplus y$ . Then, the pair  $(P', y)$  follows the uniform distribution. If  $x$  satisfies  $\text{FF}_{P'}(x) = y$ , then  $P(x) = x$  holds. In addition,  $\mathcal{B}$  can simulate the quantum oracle  $O_{P'^\pm}$  using  $O_{P^\pm}$  with no simulation overhead.

Then  $\mathcal{B}$  runs  $\mathcal{A}$ , giving  $y$  as the target image. If  $\mathcal{A}$  makes queries, then  $\mathcal{B}$  answers using the oracle  $O_{P'^\pm}$ . Finally  $\mathcal{B}$  outputs the final output of  $\mathcal{A}$ . This algorithm  $\mathcal{B}$  obviously satisfies the desired property.  $\square$

From the above lemma, it suffices to upper bound  $\text{Adv}_P^{\text{fixpt}}$  to prove Theorem 4.1. Below we show the following proposition.

**Proposition 4.1.** *Let  $n \geq 32$ . For any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  forward or backward queries to a public permutation  $P$ ,*

$$\text{Adv}_P^{\text{fixpt}}(\mathcal{A}) \leq \frac{4(e+1)(q+1)}{2^{n/2}} + \epsilon(n) \quad (14)$$

holds, where  $\epsilon(n) = \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!}$ . In particular,  $\mathcal{A}$  cannot find a fixed point of  $P$  with constant probability for  $q \ll 2^{n/2}$ .

Next, we reduce the problem of finding a fixed point of permutations to the problem of distinguishing two oracle distributions: random permutations and random derangements (permutations without fixed point). Let  $U$  be the

uniform distribution on  $\text{Perm}(\{0,1\}^n)$ , and  $U_0$  be the uniform distribution on  $\text{Der}(\{0,1\}^n) \subseteq \text{Perm}(\{0,1\}^n)$ . Then

$$\text{Adv}_P^{\text{fixpt}}(q) \leq \text{Adv}_{U^\pm, U_0^\pm}^{\text{dist}}(q+1) \quad (15)$$

holds, since we can distinguish a permutation from derangements if we find its fixed point.

To upper bound  $\text{Adv}_{U^\pm, U_0^\pm}^{\text{dist}}(q+1)$ , we apply the technique introduced in Section 3. That is, we reduce the problem of distinguishing  $U^\pm$  and  $U_0^\pm$  to the problem of distinguishing two distributions  $\Lambda$  and  $\Lambda_0$  on  $\text{Func}(\{0,1\}^n, \{0,1\})$ , introducing intermediate distributions (or games).  $\Lambda$  is the distribution which is defined according to the distribution of fixed points of random permutations, and  $\Lambda_0$  is the degenerate distribution with support on the zero-function. To this end, in addition to  $\Lambda, \Lambda_0$ , below we define functions  $\Phi : \text{Der}(\{0,1\}^n) \times \text{Func}(\{0,1\}^n, \{0,1\}) \rightarrow \text{Perm}(\{0,1\}^n)$ ,  $\Phi' : \text{Der}(\{0,1\}^n) \times \text{Func}(\{0,1\}^n, \{0,1\}) \rightarrow \text{Func}(\{0,1\}^n, \{0,1\}^n)$ , and distributions  $D_{\text{num}}$  on  $[0, \dots, 2^n]$ ,  $U'_1$  on  $\text{Perm}(\{0,1\}^n)$ , and  $U'_2$  on  $\text{Func}(\{0,1\} \times \{0,1\}^n, \{0,1\}^n)$ . In the notation of Section 3,  $G_1 = G_{\text{ideal}} = U^\pm$ ,  $G_2 = U_1^{\pm'}$ ,  $G_3 = U'_2$ , and  $G_4 = G_{\text{real}} = U_0^\pm$ , and  $D_1 = \Lambda, D_2 = \Lambda_0$ .

Here we briefly explain motivations to introduce  $U'_1, U'_2$  and  $\Phi, \Phi'$ . Our goal is to reduce the problem of distinguishing  $U^\pm$  from  $U_0^\pm$  to the problem of distinguishing  $\Lambda$  from  $\Lambda_0$ . That is, we want a technique to simulate the oracle that follows the distribution  $U^\pm$  or  $U_0^\pm$  on  $\text{Func}(\{0,1\} \times \{0,1\}^n, \{0,1\}^n)$ , given the oracle that follows the distribution  $\Lambda$  or  $\Lambda_0$  on  $\text{Func}(\{0,1\}^n, \{0,1\})$ , respectively, without any knowledge that which of  $\Lambda$  and  $\Lambda_0$  is given. However, it is difficult to directly construct such a technique. Thus, we define an intermediate distribution  $U'_1$  that is close to  $U$ , and so that we can construct such a technique between  $U_1^{\pm'}$  and  $U_0^\pm$ . The technique is as follows. Firstly, we define a map  $\Phi : \text{Der}(\{0,1\}^n) \times \text{Func}(\{0,1\}^n, \{0,1\}) \rightarrow \text{Perm}(\{0,1\}^n)$  such that  $\Phi(P_0, f)$  follows  $U'_1$  if  $(P_0, f)$  follows  $(U_0, \Lambda)$ , and  $\Phi(P_0, f)$  follows  $U_0$  if  $(P_0, f)$  follows  $(U_0, \Lambda_0)$ , respectively (actually  $\Phi$  is firstly defined and then  $U'_1$  is defined using  $\Phi$ ). Secondly, given an oracle  $f$  that follows  $\Lambda$  or  $\Lambda_0$ , we choose  $P_0 \in \text{Der}(\{0,1\}^n)$  uniformly at random, and simulate the oracle of  $(\Phi(P_0, f))^\pm$ . Then, we can simulate the distributions  $U_1^{\pm'}$  or  $U_0^\pm$  according to which of  $\Lambda$  or  $\Lambda_0$  is given. However, there is a problem: simulation cost of  $U_1^{\pm'}$  might become very high. Thus we introduce another distribution  $U'_2$  and map  $\Phi'$ , to overcome the problem of simulation overhead. Details on simulation overhead will be explained later.

Now we give formal description of intermediate distributions and maps  $\Phi, \Phi'$ . In what follows, we identify a function  $F \in \text{Func}(\{0,1\}^n, \{0,1\}^n)$  with the associated graph  $G_F$  of which vertexes are  $n$ -bit strings. In the graph  $G_F$ , there is an edge from a vertex  $x$  to another vertex  $y$  if and only if  $F(x) = y$ . If  $F$  is a permutation  $P$ , then each connected component of  $G_P$  is a cycle, and isolated points correspond to fixed points of  $P$ .

**Distribution  $D_{\text{num}}$ .** Distribution  $D_{\text{num}}$  on  $[0, \dots, 2^n]$  is the distribution of the number of fixed points of random permutations.  $D_{\text{num}}$  is formally defined



by  $D_{num}(\lambda) := \Pr_{P \sim U}[\lambda = |\{x | P(x) = x\}|]$ . In other words,  $D_{num}$  is the distribution of the random variable that takes values in  $[0, \dots, 2^n]$  which is defined according to the following sampling.

1.  $P \xleftarrow{\$} \text{Perm}(\{0, 1\}^n)$
2.  $\lambda \leftarrow |\{x | P(x) = x\}|$
3. Return  $\lambda$ .

**Distribution  $\Lambda$ .** Distribution  $\Lambda$  on  $\text{Func}(\{0, 1\}^n, \{0, 1\})$  is defined according to the distribution of fixed points of random permutations. For  $P \in \text{Perm}(\{0, 1\}^n)$ , define  $f_P \in \text{Func}(\{0, 1\}^n, \{0, 1\})$  by  $f_P(x) = 1$  if and only if  $P(x) = x$ . Then,  $\Lambda$  is formally defined by  $\Lambda(f) := \Pr_{P \sim U}[f = f_P]$ . In other words,  $\Lambda$  is the distribution of the random variable that takes values in  $\text{Func}(\{0, 1\}^n, \{0, 1\})$ , which is defined according to the following sampling:

1.  $P \xleftarrow{\$} \text{Perm}(\{0, 1\}^n)$
2.  $f \leftarrow f_P$
3. Return  $f$ .

**Distribution  $\Lambda_0$ .** Distribution  $\Lambda_0$  on  $\text{Func}(\{0, 1\}^n, \{0, 1\})$  is the degenerate distribution with support on the zero-function  $\mathbf{0}$ , which maps  $x$  to 0 for any  $x$ . Formally,  $\Lambda_0$  is defined by  $\Lambda_0(g) := 1$  if and only if  $g = \mathbf{0}$ .

**Function  $\Phi$ .** Taking  $P_0 \in \text{Der}(\{0, 1\}^n)$  and  $f \in \text{Func}(\{0, 1\}^n, \{0, 1\})$  as inputs, we want to construct another permutation  $P = \Phi(P_0, f)$  which has, informally speaking, the following properties:

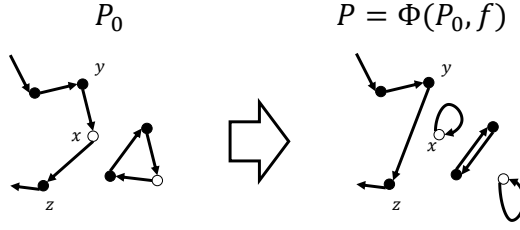
1.  $P(x) = x$  if and only if  $f(x) = 1$  holds with high probability when  $P_0$  and  $f$  are chosen uniformly at random.
2. If  $f(x) = 0$ , then  $P(x) = P_0(x)$  for almost all  $x$ .

This function  $\Phi$  is used later to approximate  $U$  by using  $U_0$  and  $\Lambda$ .

Formally, function  $\Phi : \text{Der}(\{0, 1\}^n) \times \text{Func}(\{0, 1\}^n, \{0, 1\}) \rightarrow \text{Perm}(\{0, 1\}^n)$  is defined by the following process.

1. Take  $P_0 \in \text{Perm}(\{0, 1\}^n)$ ,  $f \in \text{Func}(\{0, 1\}^n)$  as inputs.
2. For each  $x \in \{0, 1\}^n$ , define  $P(x)$  by:
3.   If  $f(x) = 1$
4.      $P(x) \leftarrow x$
5.   Else
6.     Calculate  $\min\{i | f(P_0^i(x)) = 0\}$ ,  $\text{cnt} \leftarrow \min\{i | f(P_0^i(x)) = 0\}$
7.      $P(x) \leftarrow P_0^{\text{cnt}}(x)$
8.   End If
9.  $\Phi(P_0, f) \leftarrow P$

Figure 2 illustrates how  $P = \Phi(P_0, f)$  is generated from  $P_0$  and  $f$ . Each element  $x$  such that  $f(x) = 1$  is converted to isolated points, and the edges  $y \rightarrow x, x \rightarrow z$  are converted to new edges  $y \rightarrow z, x \rightarrow x$ . By definition, images of  $\Phi$  are certainly in  $\text{Perm}(\{0, 1\}^n)$ . Note that  $\Phi(P_0, f)^{-1} = \Phi(P_0^{-1}, f)$  holds.



**Fig. 2.** How  $P = \Phi(P_0, f)$  is generated. White circle are the preimages of 1 by  $f$ .

**Function  $\Phi'$**   $\Phi'$  is a function which is defined to approximate  $U$  using  $U_0$  and  $\Lambda$  similarly as  $\Phi$ , but the approximation of  $\Phi'$  is more rough than that of  $\Phi$ . While outputs of  $\Phi$  are always permutations, outputs of  $\Phi'$  might not be permutations (although  $\Phi(P_0, f) = \Phi'(P_0, f)$  holds with high probability when  $P_0$  and  $f$  are sampled following  $U_0$  and  $\Lambda$ , as shown in Appendix H.)

Formally, function  $\Phi' : \text{Der}(\{0, 1\}^n) \times \text{Func}(\{0, 1\}^n, \{0, 1\}) \rightarrow \text{Func}(\{0, 1\}^n, \{0, 1\}^n)$  is defined by the following process.

1. Take  $P_0 \in \text{Perm}(\{0, 1\}^n)$ ,  $f \in \text{Func}(\{0, 1\}^n, \{0, 1\})$  as inputs.
2. For each  $x \in \{0, 1\}^n$ , define  $P(x)$  by:
  3. If  $f(x) = 1$
  4.  $P(x) \leftarrow x$
  5. Else If  $f(P_0(x)) = 1$
  6.  $P(x) \leftarrow P_0^2(x)$
  7. Else
  8.  $P(x) \leftarrow P_0(x)$
  9. End If
10.  $\Phi'(P_0, f) \leftarrow P$

We defined not only  $\Phi$  but also  $\Phi'$  to achieve low simulation overhead: Suppose we are given the oracle of  $f \in \text{Func}(\{0, 1\}^n, \{0, 1\})$ . Then, for any  $P_0 \in \text{Der}(\{0, 1\}^n)$  which we choose ourselves, we can operate one evaluation of the function  $\Phi'(P_0, f)$  with only two queries to  $f$ . On the other hand, we might need a lot of queries to  $f$  to evaluate  $\Phi(P_0, f)$  in Step 6 of the definition of  $\Phi$  (we need about  $2^n$  queries in the worst case). This is the reason why we introduced  $\Phi'$ .

For fixed  $P_0$  and  $f$ , we define  $P_2^{\pm} : \{0, 1\} \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  by

$$P_2^{\pm}(b, x) := \begin{cases} \Phi'(P_0, f)(x) & \text{if } b = 0, \\ \Phi'(P_0^{-1}, f)(x) & \text{if } b = 1. \end{cases}$$

$P_2^{\pm}$  can be regarded as an approximation of the function  $\Phi^{\pm}(P_0, f) \in \text{Func}(\{0, 1\} \times \{0, 1\}^n, \{0, 1\}^n)$ , which is defined by  $\Phi^{\pm}(P_0, f)(0, x) = \Phi(P_0, f)(x)$  and  $\Phi^{\pm}(P_0, f)(1, x) = \Phi(P_0^{-1}, f)(x)$ .

**Distribution  $U'_1$**  Distribution  $U'_1$  on  $\text{Perm}(\{0,1\}^n)$  is an approximation of the uniform distribution  $U$  that combines  $U_0$  with  $\Lambda$ . Formally,  $U'_1$  is defined by  $U'_1(P) = \Pr_{P_0 \sim U_0, f \sim \Lambda}[P = \Phi(P_0, f)]$ . In other words,  $U'_1$  is the distribution of the random variable that takes values in  $\text{Perm}(\{0,1\}^n)$  which is defined according to the following sampling:

1.  $P_0 \xleftarrow{U_0} \text{Perm}(\{0,1\}^n)$ ,  $f \xleftarrow{\Lambda} \text{Func}(\{0,1\}^n, \{0,1\})$
2.  $P \leftarrow \Phi(P_0, f)$

Note that if  $P$  is sampled following  $U'_1$ , we assume that a quantum adversary  $\mathcal{A}$  is given a quantum oracle of  $P^\pm : \{0,1\} \times \{0,1\}^n \rightarrow \{0,1\}^n$  (see Section 2.1).

**Distribution  $U'_2$**  Distribution  $U'_2$  on  $\text{Func}(\{0,1\} \times \{0,1\}^n, \{0,1\}^n)$  is another approximation of  $U$ , which is more “rough” than  $U'_1$ . Below, for  $F \in \text{Func}(\{0,1\} \times \{0,1\}^n, \{0,1\}^n)$ , the  $n$ -bit functions  $F(0, \cdot), F(1, \cdot)$  are denoted by  $F^+$  and  $F^-$ . Then, formally,  $U'_2$  is defined by  $U'_2(F) = \Pr_{P_0 \sim U_0, f \sim \Lambda}[F^+ = \Phi'(P_0, f) \wedge F^- = \Phi'(P_0^{-1}, f)]$ . In other words,  $U'_2$  is the distribution of the random variable that takes values in  $\text{Func}(\{0,1\} \times \{0,1\}^n, \{0,1\}^n)$  which is defined according to the following sampling:

1.  $P_0 \xleftarrow{U_0} \text{Perm}(\{0,1\}^n)$ ,  $f \xleftarrow{\Lambda} \text{Func}(\{0,1\}^n, \{0,1\})$
2.  $F^+ \leftarrow \Phi'(P_0, f)$ ,  $F^- \leftarrow \Phi'(P_0^{-1}, f)$

Now the preparation to use the technique in Section 3 is completed. We reduce the problem of distinguishing  $U$  from  $U_0$  to the problem of distinguishing  $\Lambda$  and  $\Lambda_0$ . Now we have the following inequalities.

$$\begin{aligned} \text{Adv}_{U^\pm, U_0^\pm}^{\text{dist}}(\mathcal{A}) &\leq \text{Adv}_{U^\pm, U_1^{\pm}}^{\text{dist}}(\mathcal{A}) + \text{Adv}_{U_1^{\pm}, U_2'}^{\text{dist}}(\mathcal{A}) + \text{Adv}_{U_2', U_0^\pm}^{\text{dist}}(\mathcal{A}) \\ &\leq \Delta(U^\pm, U_1^{\pm}) + \Delta(U_1^{\pm}, U_2') + \text{Adv}_{U_2', U_0^\pm}^{\text{dist}}(\mathcal{A}). \end{aligned} \quad (16)$$

Next, we show the following lemma.

**Lemma 4.2.** *For a quantum algorithm  $\mathcal{A}$  to distinguish  $U_2'$  from  $U_0^\pm$  that makes at most  $q$  quantum queries, we can construct a quantum algorithm  $\mathcal{B}$  to distinguish  $\Lambda$  from  $\Lambda_0$  that makes at most  $2q$  queries and satisfies*

$$\text{Adv}_{U_2', U_0^\pm}^{\text{dist}}(\mathcal{A}) = \text{Adv}_{\Lambda, \Lambda_0}^{\text{dist}}(\mathcal{B}).$$

*Proof.* We give a quantum algorithm  $\mathcal{B}$  that satisfies the desired properties.  $\mathcal{B}$  is given a quantum oracle  $O_f$ , where  $f$  is sampled according to  $\Lambda$  or  $\Lambda_0$ . Before making queries,  $\mathcal{B}$  chooses a derangement  $P_0$  uniformly at random. Then,  $\mathcal{B}$  runs  $\mathcal{A}$ .  $\mathcal{B}$  answers to queries of  $\mathcal{A}$  by calculating  $\Phi'(P_0, f)$  and  $\Phi'(P_0^{-1}, f)$ . By definition of  $\Phi'$ ,  $\mathcal{B}$  can calculate one evaluation of  $\Phi'(P_0, f)$  (and  $\Phi'(P_0^{-1}, f)$ ) with two queries to  $O_f$ . Finally,  $\mathcal{B}$  outputs what  $\mathcal{A}$  outputs.

Since  $\mathcal{A}$  makes at most  $q$  queries,  $\mathcal{B}$  makes at most  $2q$  queries.  $\mathcal{B}$  perfectly simulates the distributions  $U_2'$  and  $U_0^\pm$  according to which of  $\Lambda$  and  $\Lambda_0$  is given. Thus  $\text{Adv}_{U_2', U_0^\pm}^{\text{dist}}(\mathcal{A}) = \text{Adv}_{\Lambda, \Lambda_0}^{\text{dist}}(\mathcal{B})$  holds.  $\square$

From the above lemma and the inequalities (16), we have

$$\text{Adv}_{U^\pm, U_0^\pm}^{\text{dist}}(q) \leq \Delta(U^\pm, U_1^{\pm}) + \Delta(U_1^{\pm}, U_2') + \text{Adv}_{\Lambda, \Lambda_0}^{\text{dist}}(2q). \quad (17)$$

The three terms in the right hand side are upper bounded as in the following lemmas.

**Lemma 4.3.**  $\text{Adv}_{\Lambda, \Lambda_0}^{\text{dist}}(q) \leq \frac{2(e+1)q}{2^{n/2}}$

**Lemma 4.4.**  $\Delta(U^\pm, U_1^{\pm}) \leq \frac{8n^3}{2^n - 2n + 1} + \frac{16n^3}{2^n} + \frac{e+1}{n!}$  for  $n \geq 32$ .

**Lemma 4.5.**  $\Delta(U_1^{\pm}, U_2') \leq \frac{32n^3}{2^n} + \frac{2(e+1)}{n!}$  for  $n \geq 32$ .

Thus we have

$$\text{Adv}_{U^\pm, U_0^\pm}^{\text{dist}}(q) \leq \frac{4(e+1)q}{2^{n/2}} + \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!}.$$

Combining this inequality and inequality (15), we obtain the desired bound (14) in Theorem 4.1.

To complete the proof, we give a proof of Lemma 4.3. Proofs of Lemma 4.4 and 4.5 will be given in Appendix G and Appendix H.

**Proof of Lemma 4.3.** To prove the Lemma 4.3, we use Proposition 3.2. Let us define a set of functions  $\text{good}, \text{bad} \subset \text{Func}(\{0, 1\}^n, \{0, 1\})$  by  $\text{good} := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus \{0\}$ , and  $\text{bad} := \emptyset$ . In addition, for each integer  $\lambda > 0$ , define  $\text{good}_\lambda \subset \text{good}$  by  $f \in \text{good}_\lambda$  if and only if  $|f^{-1}(1)| = \lambda$ . Then,  $\bigcup_\lambda \text{good}_\lambda = \text{good}$  and  $\text{good}_{\lambda_1} \cap \text{good}_{\lambda_2} = \emptyset$  for  $\lambda_1 \neq \lambda_2$ . Moreover, the conditional probability  $\Pr_{F \sim \Lambda}[F = f | F \in \text{good}_\lambda]$  is independent on  $f$  due to the symmetry of the distribution  $\Lambda$ . Therefore we can apply Proposition 3.2.

Let  $p_1^{\text{good}_\lambda} := \Pr_{F \sim \Lambda}[F \in \text{good}_\lambda]$  and  $p_1^{f|\text{good}_\lambda} := \Pr_{F \sim \Lambda}[F = f | F \in \text{good}_\lambda]$ . For each fixed  $x$ , the number of boolean function  $f$  such that  $f(x) = 1 \wedge |f^{-1}(1)| = \lambda$  is exactly  $\binom{2^n - 1}{\lambda - 1}$ . Hence we have

$$\max_x |\{f \in \text{good}_\lambda \mid f(x) = 1\}| = \binom{2^n - 1}{\lambda - 1}. \quad (18)$$

In addition,

$$p_1^{f|\text{good}_\lambda} = \frac{1}{\binom{2^n}{\lambda}}. \quad (19)$$

hold.

Next, we upper bound  $p_1^{\text{good}_\lambda} = \Pr_{f \sim \Lambda}[f \in \text{good}_\lambda] = \Pr_{a \sim D_{\text{num}}}[a = \lambda]$ . For any fixed  $\lambda$ , we have

$$\begin{aligned} \Pr_{a \sim D_{\text{num}}}[a = \lambda] &= \frac{\binom{2^n}{\lambda}!(2^n - \lambda)}{2^n!} = \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{1}{\lambda!} \leq \frac{(2^n - \lambda)!/e + 1}{(2^n - \lambda)!} \cdot \frac{1}{\lambda!} \\ &= \left(1 + \frac{e}{(2^n - \lambda)!}\right) \cdot \frac{1}{e} \cdot \frac{1}{\lambda!} \leq \frac{1+e}{e} \cdot \frac{1}{\lambda!} \end{aligned} \quad (20)$$

(Remember that  $!N$  denotes the number of derangements on a set of size  $N$  and  $!N = \lfloor \frac{N!}{e} + \frac{1}{2} \rfloor$  holds. see Section 2.) Thus we have

$$p_1^{\text{good}_\lambda} = \Pr_{f \sim \Lambda}[f \in \text{good}_\lambda] \leq \frac{1+e}{e} \cdot \frac{1}{\lambda!}. \quad (21)$$

From Proposition 3.2, equality (18), (19), and inequality (21), since  $\Pr_{f \sim \Lambda}[f \in \text{bad}] = 0$  we have

$$\begin{aligned} \text{Adv}_{\Lambda, \Lambda_0}^{\text{dist}}(q) &\leq 2q \cdot \sum_{0 < \lambda} p_1^{\text{good}_\lambda} \sqrt{p_1^{f|\text{good}_\lambda} \cdot \max_x \{|\{f \mid f(x) = 1 \wedge f \in \text{good}_\lambda\}|\}} \\ &\leq 2q \cdot \sum_{0 < \lambda} \frac{1+e}{e} \cdot \frac{1}{\lambda!} \sqrt{\frac{\binom{2^n-1}{\lambda-1}}{\binom{2^n}{\lambda}}} \leq \frac{2q(1+e)}{e} \cdot \sum_{0 < \lambda} \frac{1}{\lambda!} \sqrt{\frac{\lambda}{2^n}} \\ &= \frac{2q(1+e)}{e} \cdot \sum_{0 < \lambda} \frac{1}{\sqrt{\lambda(\lambda-1)!}} \sqrt{\frac{1}{2^n}} \\ &\leq \frac{2q(1+e)}{e} \cdot \sum_{0 \leq \lambda} \frac{1}{\lambda!} \sqrt{\frac{1}{2^n}} = \frac{2(e+1)q}{\sqrt{2^n}}, \end{aligned} \quad (22)$$

which is the desired bound. Hence Lemma 4.3 follows.  $\square$

*Remark 4.2.* In this section we showed the non-invertibility of  $\text{FF}^P$  but did not show the one-wayness, because it seems difficult to reduce the one-wayness to the non-invertibility for the case of a permutation with feedforward. For Davies-Meyer construction, on the other hand, we can reduce its one-wayness to the non-invertibility by upper-bounding the total variation distance between the distribution of the game to break the one-wayness and that of the game to break the non-invertibility. Unfortunately, for permutations with feedforward, this strategy cannot be applied since the total variation distance between the two corresponding distributions would become very large.

## 5 Security of Davies-Meyer Constructions in the Quantum Ideal Cipher Model

This section gives proofs for security of Davies-Meyer constructions in the quantum ideal cipher model. We begin with showing non-invertibility, and then prove one-wayness. Our result in this section is the first proof for quantum security of functions based on public block ciphers.

### 5.1 Non-Invertibility of Davies-Meyer

Non-invertibility in the ideal cipher model is shown in the similar way as in the proof for non-invertibility of permutation with feedforward in Section 4. We show the following theorem.

**Theorem 5.1 (Non-invertibility of Davies-Meyer).** *Let  $n \geq 32$ . For any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  queries to a block cipher  $E$ ,*

$$\text{Adv}_{\text{DM}^E}^{\text{inv}}(\mathcal{A}) \leq 4(q+1) \left( \frac{n^{1/2}}{2^{n/2}} + \frac{2^m(e+1)}{n!} \right) + 2^m \epsilon(n) \quad (23)$$

*holds, where  $\epsilon(n) = \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!}$ . In particular,  $\mathcal{A}$  cannot invert  $\text{DM}^E$  with constant probability if  $\frac{2^m}{2^n} \ll 1$  and  $q \ll 2^{n/2}/n^{1/2}$ .*

*Remark 5.1.* In the above theorem, security bound is valid only for the case that key length  $m$  is less than block length  $n$ . (We do not know if there exist any attacks that exploit long key lengths. The condition that key length should be shorter than the block length comes from limitation of our proof technique.) However, even if  $m \geq n$ , then we can achieve the same bound if we restrict key space. That is, if we are given  $n$ -bit block ciphers with  $m$ -bit key and  $m \geq n$ , we use only the keys of which all bits are 0 except for the first  $n/2$ -bits, for example. Then we can construct non-invertible functions with  $3n/2$ -bit input and  $n$ -bit output.

We cannot get rid of this restriction on usage of key space since there are terms of order  $O(n^3 \cdot 2^{m-n})$  in our bound (23), which come from Lemma 4.4 and 4.5. The bound of Lemma 4.4 cannot be essentially improved, since  $\Delta(U, U'_1) \geq \frac{1}{4e \cdot 2^n}$  holds (see Appendix J for more details). Thus, if we want to get rid of the restriction, then we have to use other proof strategies.

Let  $U_E$  be the uniform distribution on  $\text{Ciph}(m, n)$ , and  $U_{E,0}$  be the distribution on  $\text{Ciph}(m, n)$  defined by  $U_{E_0}(E) = \prod_k U_0(E_k)$  (i.e., when  $E$  is sampled according to  $U_{E,0}$ , then  $E_k$  is sampled according to  $U_0$  for each key  $k$ .) We say that a pair  $(z, x)$  is a fixed point of a block cipher  $E$  if  $E_z(x) = x$ . Let us define the advantage of a quantum algorithm  $\mathcal{A}$  to find a fixed point of an ideal block cipher  $E$  by

$$\text{Adv}_E^{\text{fixpt}}(\mathcal{A}) := \Pr_{E \sim U_E} [\mathcal{A}^{O_{E^\pm}}() = (z, x) \wedge E_z(x) = x],$$

and

$$\text{Adv}_E^{\text{fixpt}}(q) := \max_{\mathcal{A}} \left\{ \text{Adv}_E^{\text{fixpt}}(\mathcal{A}) \right\},$$

where the maximum is taken over all quantum-query algorithms, each making at most  $q$  quantum queries.

Then, similarly as in the proof for permutation with feedforward, we have

$$\text{Adv}_{\text{DM}^E}^{\text{inv}}(q) \leq \text{Adv}_E^{\text{fixpt}}(q) \leq \text{Adv}_{U_{E^\pm}, U_{E,0}^\pm}^{\text{dist}}(q+1). \quad (24)$$

To upper bound  $\text{Adv}_{U_{E^\pm}, U_{E,0}^\pm}^{\text{dist}}$ , we introduce distributions  $D_{E, \text{num}}, \Lambda_E, \Lambda_{E,0}, U'_{E,1}, U'_{E,2}$ , which are essentially product distributions of  $D_{\text{num}}, \Lambda, \Lambda_0, U'_1, U'_2$ , respectively.

**Distribution  $D_{E,num}$ .** Distribution  $D_{E,num}$  on  $([0, \dots, 2^n])^{\times 2^m}$  is the product distribution  $D_{num} \times \dots \times D_{num}$ , i.e.  $D_{E,num}$  is defined by  $D_{E,num}(\lambda_0, \dots, \lambda_{2^m-1}) := D_{num}(\lambda_0) \times \dots \times D_{num}(\lambda_{2^m-1})$ .  $D_{E,num}$  can be regarded as the distribution of the number of fixed points of ideal ciphers.

**Distribution  $\Lambda_E$ .** Distribution  $\Lambda_E$  on the set  $\text{Func}(\{0, 1\}^m \times \{0, 1\}^n, \{0, 1\}) = (\text{Func}(\{0, 1\}^n, \{0, 1\}))^{2^m}$  is defined as the product distribution  $\Lambda \times \dots \times \Lambda$ , i.e.  $\Lambda_E$  is defined by  $\Lambda_E(F) := \Lambda_E(F(0, \cdot)) \times \Lambda_E(F(1, \cdot)) \times \dots \times \Lambda_E(F(2^m - 1, \cdot))$ .  $\Lambda_E$  can be regarded as the distribution of fixed points of ideal ciphers.

**Distribution  $\Lambda_{E,0}$ .** Distribution  $\Lambda_{E,0}$  on  $\text{Func}(\{0, 1\}^m \times \{0, 1\}^n, \{0, 1\})$  is the degenerate distribution with support on the zero-function  $\mathbf{0}$ .

**Distribution  $U'_{E,1}$**  Distribution  $U'_{E,1}$  on  $\text{Ciph}(m, n)$  is defined by  $U'_{E,1}(E) := \prod_{k \in \{0, 1\}^m} U'_1(E_k)$ . That is, when  $E$  is sampled according to  $U'_{E,1}$ , then  $E_k$  is chosen according to  $U'_1$  independently for each key  $k$ . Similarly as  $U'_1$  is an approximation of  $U$ ,  $U'_{E,1}$  can be regarded as an approximation of  $U_E$ .

**Distribution  $U'_{E,2}$**  Distribution  $U'_{E,2}$  on  $\text{Func}(\{0, 1\} \times \{0, 1\}^m \times \{0, 1\}^n, \{0, 1\}^n)$  is defined by  $U'_{E,2}(F) = \prod_{k \in \{0, 1\}^m} U'_2(F(\cdot, k, \cdot))$ . That is,  $U'_{E,2}$  is the distribution of the random variable that is defined by the following sampling.

1. For each  $z \in \{0, 1\}^m$ , do:
2.  $G_z \leftarrow^{U'_2} \text{Func}(\{0, 1\} \times \{0, 1\}^n, \{0, 1\}^n)$
3.  $F(b, z, x) \leftarrow G_z(b, x)$  for each  $b \in \{0, 1\}$ ,  $z \in \{0, 1\}^m$ ,  $x \in \{0, 1\}^n$ .
4. Return  $F$

Similarly as  $U'_2$  is a rough approximation of  $U^\pm$ ,  $U'_{E,2}$  can be regarded as a rough approximation of  $U_E^\pm$ .

Now we apply the technique introduced in Section 3. Similarly as inequality (17), we can show that

$$\text{Adv}_{U_E^\pm, U_{E,0}^\pm}^{dist}(q) \leq \Delta(U_E^\pm, U_{E,1}^\pm) + \Delta(U_{E,1}^\pm, U_{E,2}^\pm) + \text{Adv}_{\Lambda_E, \Lambda_{E,0}}^{dist}(2q),$$

holds. In addition, since  $U, U'_{E,1}, U'_{E,2}$  are essentially the product distributions of  $U, U'_1, U'_2$ , from Lemma 4.4 and Lemma 4.5 we have

$$\begin{aligned} \text{Adv}_{U_E^\pm, U_{E,0}^\pm}^{dist}(q) &\leq 2^m \Delta(U^\pm, U_1^\pm) + 2^m \Delta(U_1^\pm, U_2^\pm) + \text{Adv}_{\Lambda_E, \Lambda_{E,0}}^{dist}(2q) \\ &\leq 2^m \left( \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!} \right) + \text{Adv}_{\Lambda_E, \Lambda_{E,0}}^{dist}(2q). \end{aligned} \tag{25}$$

Thus, to prove Theorem 5.1, it suffices to show the following lemma.

**Lemma 5.1.**

$$\text{Adv}_{\Lambda_E, \Lambda_{E,0}}^{\text{dist}}(q) \leq 2q \left( \frac{n^{1/2}}{2^{n/2}} + \frac{2^m(e+1)}{n!} \right)$$

*Proof.* To prove the Lemma 5.1, again we use our tool in Section 3. Let us define a set of functions  $\text{good} \subset \text{Func}(\{0,1\}^m \times \{0,1\}^n, \{0,1\})$  by  $f \in \text{good}$  if and only if  $f \neq \mathbf{0}$  and  $\lambda_z = |f_z^{-1}(1)| < n$  for all  $z \in \{0,1\}^m$ , where  $f_z(\cdot) = f(z, \cdot)$ . Let  $\text{bad} := \text{Func}(\{0,1\}^m \times \{0,1\}^n, \{0,1\}) \setminus (\text{good} \cup \{\mathbf{0}\})$ . In addition, for each sequence of integers  $\lambda_S = (\lambda_0, \lambda_1, \dots, \lambda_{2^m-1})$ , define  $\text{good}_{\lambda_S} \subset \text{good}$  by  $f \in \text{good}_{\lambda_S}$  if and only if  $f_z^{-1}(1) = \lambda_z$  for all  $0 \leq z \leq 2^m - 1$ . For simplicity, we write  $\lambda_S < n$  if and only if  $\lambda_z < n$  for all  $0 \leq z \leq 2^m - 1$ . Similarly, we write  $0 < \lambda_S$  if and only if  $\lambda_z > 0$  for all  $0 \leq z \leq 2^m - 1$ . Then,  $\bigcup_{0 < \lambda_S < n} \text{good}_{\lambda_S} = \text{good}$  and  $\text{good}_{\lambda_S} \cap \text{good}_{\lambda_{S'}} = \emptyset$  for  $\lambda_S \neq \lambda_{S'}$ . The conditional probability  $\Pr_{F \sim \Lambda_E}[F = f | f \in \text{good}_{\lambda_S}]$  is independent on  $f$  due to the symmetry of the distribution  $\Lambda_E$ . Therefore we can apply Proposition 3.2 with  $D_1 = \Lambda_E$  and  $D_2 = \Lambda_{E,0}$ .

Define

$$p_1^{\text{good}_{\lambda_S}} := \Pr_{f \sim \Lambda_E} [f \in \text{good}_{\lambda_S}] \quad (26)$$

and

$$p_1^{f|\text{good}_{\lambda_S}} := \Pr_{F \sim \Lambda_E} [F = f | F \in \text{good}_{\lambda_S}]. \quad (27)$$

Now we upper bound  $\Pr_{f \sim \Lambda_E}[f \in \text{bad}]$ . Note that  $\Pr_{f \sim \Lambda_E}[f \in \text{bad}] \leq 2^m \Pr_{f \sim \Lambda} [|f^{-1}(1)| \geq n]$  holds since  $\Lambda_E$  is product distribution of  $\Lambda$ . In addition, from inequality (21) we have

$$\Pr_{f \sim \Lambda} [|f^{-1}(1)| \geq \lambda_0] \leq \frac{e+1}{e} \sum_{\lambda \geq \lambda_0} \frac{1}{\lambda!} \leq \frac{e+1}{e} \frac{e}{\lambda_0!} = \frac{e+1}{\lambda_0!}, \quad (28)$$

where we used the fact  $\sum_{\lambda \geq \lambda_0} \frac{1}{\lambda!} \leq \frac{e}{\lambda_0!}$  (see Lemma A.1). Thus we have

$$\Pr_{f \sim \Lambda_E} [f \in \text{bad}] \leq \frac{2^m(1+e)}{n!}. \quad (29)$$

Next, we upper bound  $p_1^{f|\text{good}_{\lambda_S}} \cdot \max_{(z,x)} |\{f \in \text{good}_{\lambda_S} | f(z,x) = f_z(x) = 1\}|$ . For each fixed  $w \in \{0,1\}^m$ ,  $x \in \{0,1\}^n$  and  $\lambda_S = (\lambda_0, \dots, \lambda_{2^m-1})$ , the number of boolean function  $f \in \text{good}_{\lambda_S}$  such that  $f_w(x) = 1$  is equal to

$$\binom{2^n-1}{\lambda_w-1} \cdot \prod_{z \neq w \in \{0,1\}^m} \binom{2^n}{\lambda_z} = \frac{\lambda_w}{2^n} \cdot \prod_{z \in \{0,1\}^m} \binom{2^n}{\lambda_z}. \quad (30)$$

Thus for each sequence  $\lambda_S < n$  we have

$$\begin{aligned} \max_{(z,x)} |\{f \in \text{good}_{\lambda_S} | f(z,x) = f_z(x) = 1\}| &= \max_{(z,x)} \left\{ \frac{\lambda_z}{2^n} \cdot \prod_{z \in \{0,1\}^m} \binom{2^n}{\lambda_z} \right\} \\ &\leq \frac{n}{2^n} \cdot \prod_{z \in \{0,1\}^m} \binom{2^n}{\lambda_z} \end{aligned} \quad (31)$$



Hence, for each sequence  $\lambda_S < n$  we have

$$\begin{aligned} & p_{\Delta}^{f|\text{good}_{\lambda_S}} \cdot \max_{(z,x)} |\{f \in \text{good}_{\lambda_S} \mid f(z,x) = f_z(x) = 1\}| \\ & \leq \frac{1}{\prod_{z \in \{0,1\}^m} \binom{2^n}{\lambda_z}} \cdot \frac{n}{2^n} \cdot \prod_{z \in \{0,1\}^m} \binom{2^n}{\lambda_z} = \frac{n}{2^n}. \end{aligned} \quad (32)$$

From Proposition 3.2, and inequalities (29) and (32),  $\text{Adv}_{A, A_0}^{\text{dist}}(q)$  is upper bounded by

$$\begin{aligned} & 2q \cdot \sum_{\lambda_S < n} p_1^{\text{good}_{\lambda_S}} \sqrt{p_1^{f|\text{good}_{\lambda_S}} \cdot \max_{(z,x)} |\{f \in \text{good}_{\lambda_S} \mid f_z(x) = 1\}|} \\ & \quad + 2q \cdot \Pr_{f \sim \Lambda_E} [f \in \text{bad}] \\ & \leq 2q \cdot \sum_{\lambda_S < n} p_1^{\text{good}_{\lambda_S}} \sqrt{\frac{n}{2^n}} + 2q \cdot \frac{2^m(e+1)}{n!} \leq 2q \left( \sqrt{\frac{n}{2^n}} + \frac{2^m(e+1)}{n!} \right), \end{aligned} \quad (33)$$

which completes the proof.  $\square$

## 5.2 One-Wayness of Davies-Meyer

Next, we show that Davies-Meyer constructions are also quantum one-way in the quantum ideal cipher model.

**Theorem 5.2 (One-wayness of Davies-Meyer).** *Let  $n \geq 32$  and  $m \leq n^2$ . For any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  queries to a block cipher  $E$ ,*

$$\text{Adv}_{\text{DM}^E}^{\text{ow}}(\mathcal{A}) \leq 4(q+1) \left( \frac{n^{1/2}}{2^{n/2}} + \frac{2^m(e+1)}{n!} \right) + 2^m \epsilon(n) + \frac{2n+1}{2^{m/3+1}} + \frac{n^2}{2^{m-2}} \quad (34)$$

*holds, where  $\epsilon(n) = \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!}$ . In particular,  $\mathcal{A}$  cannot find a preimage of  $\text{DM}^E$  with constant probability if  $\frac{2^m}{2^n} \ll 1$  and  $q \ll 2^{n/2}/n^{1/2}$ .*

*Remark 5.2.* Here we need an additional condition  $m \leq n^2$  for technical reasons. This assumption is reasonable since usual block ciphers satisfy it.

*Proof.* Let  $U_n$  be the uniform distribution on  $\{0,1\}^n$  and  $V$  be the distribution on  $\text{Ciph}(m, n) \times \{0,1\}^n$  which is defined by  $V(E, y) = \Pr_{e \sim U_E, (z,x) \sim U_{m+n}} [e = E \wedge \text{DM}^E(z, x) = y]$ . That is,  $V$  is the distribution of the random variable which is defined by the following sampling:

1.  $E \xleftarrow{U_E} \text{Ciph}(m, n)$ ,  $z \xleftarrow{\$} \{0,1\}^m$ ,  $x \xleftarrow{\$} \{0,1\}^n$
2.  $y \leftarrow \text{DM}^E(z, x)$
3. Return  $(E, y)$

Then  $\text{Adv}_{\text{DM}^E}^{\text{ow}}(\mathcal{A}) = \Pr_{(E,y) \sim V}[\mathcal{A}^{O_{E^\pm}}(y) = (z', x') \wedge \text{DM}^E(z', x') = y]$  is upper bounded by

$$\begin{aligned} & \Pr_{E \sim U_E, y \sim U_n}[\mathcal{A}^{O_{E^\pm}}(y) = (z', x') \wedge \text{DM}^E(z', x') = y] \\ & + \left| \Pr_{(E,y) \sim V}[\mathcal{A}^{O_{E^\pm}}(y) = (z', x') \wedge \text{DM}^E(z', x') = y] \right. \\ & \quad \left. - \Pr_{E \sim U_E, y \sim U_n}[\mathcal{A}^{O_{E^\pm}}(y) = (z', x') \wedge \text{DM}^E(z', x') = y] \right| \\ & \leq \text{Adv}_{\text{DM}^E}^{\text{inv}}(\mathcal{A}) + \Delta(V, (U_E, U_n)). \end{aligned} \tag{35}$$

Hence Theorem 5.2 follows from Theorem 5.1 and the following lemma.

**Lemma 5.2.**  $\Delta(V, (U_E, U_n)) \leq \frac{2n+1}{2^{m/3+1}} + \frac{n^2}{2^{m-2}}$  for  $n \geq 32$  and  $m \leq n^2$ .

A proof of this lemma is given in Appendix I.  $\square$

## 6 Security of Merkle-Damgård with Davies-Meyer Constructions

This section shows that the combination of Davies-Meyer constructions with the Merkle-Damgård constructions are optimally non-invertible and one-way in the quantum ideal cipher model.

Merkle-Damgård construction is the most basic construction to convert compression functions, which have fixed input length, to a function with (variable) long input lengths. In particular, lots of popular hash functions like SHA-2 [22] are based on the Merkle-Damgård constructions, and use Davies-Meyer constructions as compression functions. Merkle-Damgård construction with MD-compliant padding is proven to be collision resistant hash function when underlying compression function is collision-resistant [12]. However, there is no guarantee that Merkle-Damgård constructions (with MD-compliant padding) become one-way (preimage resistant) or second preimage resistant hash functions even if underlying compression functions are one-way (preimage resistant) or second preimage resistant. Actually there is an attack that finds a second preimage with complexity less than  $2^n$  [17].

Since usual Merkle-Damgård constructions do not guarantee one-wayness even in classical settings, in this paper we fix input length. Input length can be very long (actually we will construct functions of which input bit length are exponential of  $n$ ), but must be fixed.

This section assumes that we are given an ideal block cipher  $E \in \text{Ciph}(m, n)$  with  $m \leq n^2$ . For a positive number  $r$  ( $r$  means ‘‘rate’’) with  $1 < r < n$  and  $\ell \geq 1$ , define a padding function  $\text{pad}_{r,\ell} : \{0, 1\}^n \times \{0, 1\}^{\frac{n}{r} \cdot \ell} \rightarrow \{0, 1\}^n \times \{0, 1\}^{m\ell}$  by

$$\text{pad}_{r,\ell} : x \| z_1 \| \cdots \| z_\ell \mapsto x \| z_1 \| 0 \| \cdots \| z_i \| (i-1) \| \cdots \| z_\ell \| (\ell-1),$$

where  $z_i \in \{0, 1\}^{\frac{n}{r}}$  and we assume that each integer  $i$  is expressed as an  $(m-n/r)$ -bit string. Let us define a function  $H_{r,\ell}^E : \{0, 1\}^{n+\frac{n}{r} \cdot \ell} \rightarrow \{0, 1\}^n$  by

$$H_{r,\ell}^E(M) := \text{MD}_\ell^{\text{DM}^E}(\text{pad}_{r,\ell}(M)).$$

The following theorem claims that  $H_{r,\ell}^E$  has both non-invertibility and one-wayness.

**Theorem 6.1 (Security of Merkle-Damgård with Davies-Meyer).** *Let  $n \geq 32$  and  $m \leq n^2$ . Assume  $E \in \text{Ciph}(m, n)$  is an ideal cipher. For any quantum adversary  $\mathcal{A}$  that makes at most  $q$  queries to  $E$ ,*

$$\text{Adv}_{H_{r,\ell}^E}^{\text{inv}}(\mathcal{A}) \leq 4(q+1) \left( \frac{n^{1/2}}{2^{n/2}} + \frac{2^{n/r}(e+1)}{n!} \right) + \epsilon(r, n) \quad (36)$$

and

$$\text{Adv}_{H_{r,\ell}^E}^{\text{ow}}(\mathcal{A}) \leq 4(q+1) \left( \frac{n^{1/2}}{2^{n/2}} + \frac{2^{n/r}(e+1)}{n!} \right) + \epsilon(r, n) + \delta(r, \ell, n) \quad (37)$$

holds, where  $\epsilon(r, n) = 2^{n/r} \left( \frac{8n^3}{2^n - 2n + 1} + \frac{48n^3}{2^n} + \frac{3(e+1)}{n!} \right)$  and  $\delta(r, \ell, n) = \ell \cdot \left( \frac{2n+1}{2^{n/3r+1}} + \frac{n^2}{2^{n/r-2}} \right)$ . In particular, if  $\ell \ll 2^{\frac{2n}{3r}}$ , then  $\mathcal{A}$  cannot find a preimage of  $H_{r,\ell}^E$  with constant probability for  $q \ll 2^{n/2}/n^{1/2}$ .

*Remark 6.1.* We need padding function  $\text{pad}_{r,\ell}$  to restrict key space for each message block (see Remark 5.1). Our padding function pads different numbers for different message blocks so that the  $i$ -th compression function and the  $j$ -th compression function become essentially independent for  $i \neq j$ .

*Proof.* Firstly we show non-invertibility, i.e. inequality (36). Non-invertibility of  $H_{r,\ell}^E$  is reduced to non-invertibility of the Davies-Meyer construction of the last block. By using an adversary  $\mathcal{A}$  to invert  $H_{r,\ell}^E$ , we construct an adversary  $\mathcal{B}$  to invert a Davies-Meyer construction  $\text{DM}^{E'}$ , where  $E' \in \text{Ciph}(n/r, n)$ .

At the beginning of a game,  $\mathcal{B}$  receives randomly chosen  $y \in \{0, 1\}^n$  as an input. In addition,  $\mathcal{B}$  has oracle access to an ideal cipher  $E' \in \text{Ciph}(n/r, n)$ .  $\mathcal{B}$  simulates an oracle of ideal cipher  $E \in \text{Ciph}(m, n)$  as follows.  $\mathcal{B}$  chooses  $\tilde{E} \in \text{Ciph}(m, n)$  uniformly at random, and define  $E \in \text{Ciph}(m, n)$  by

$$E(k, x) = \begin{cases} E'(z, x) & \text{if } k = z\|\ell \text{ for some } z \in \text{Ciph}(n/r, n), \\ \tilde{E}(k, x) & \text{otherwise.} \end{cases} \quad (38)$$

The distribution of  $E$  equals to the uniform distribution.  $\mathcal{B}$  runs  $\mathcal{A}$ , giving  $y$  as the target image.  $\mathcal{B}$  answers queries of  $\mathcal{A}$  by using  $E$ . After  $\mathcal{A}$  outputs a message  $M = x\|z_1\|\cdots\|z_\ell \in \{0, 1\}^{n+\frac{n}{r}\cdot\ell}$ ,  $\mathcal{B}$  calculates  $x_{\ell-1} := H_{r,\ell-1}^E(x\|z_1\|\cdots\|z_{\ell-1})$  and outputs  $(z_\ell, x_{\ell-1})$ . Note that calculation of  $x_{\ell-1}$  does not need any query to  $E'$ . Since  $\text{DM}^{E'}(z_\ell\|\ell, H_{r,\ell-1}^E(x\|z_1\|\cdots\|z_{\ell-1})) = H_{r,\ell}^E(M) = y$  holds, we have  $\text{Adv}_{H_{r,\ell}^E}^{\text{inv}}(\mathcal{A}) = \text{Adv}_{\text{DM}^{E'}}^{\text{inv}}(\mathcal{B})$ , and we obtain the desired bound (36) from Theorem 5.1.

Next we show one-wayness, i.e. inequality (37). Similarly as in Section 5, we reduce one-wayness to non-invertibility. Again, let  $U_n$  be the uniform distribution on  $\{0, 1\}^n$ . Let  $V_1$  be the distribution of the random variable which takes values in  $\text{Ciph}(m, n) \times \{0, 1\}^n$  and is defined by the following sampling:

1.  $E \xleftarrow{U_E} \text{Ciph}(m, n), M \xleftarrow{\$} \{0, 1\}^{n + \frac{n}{r} \cdot \ell}$
2.  $y \leftarrow H_{r, \ell}^E(M)$
3. return  $(E, y)$

Then we have

$$\text{Adv}_{H_{r, \ell}^E}^{\text{ow}}(\mathcal{A}) \leq \text{Adv}_{H_{r, \ell}^E}^{\text{inv}}(\mathcal{A}) + \Delta(V_1, (U_E, U_n)). \quad (39)$$

Below we upper bound  $\Delta(V_1, (U_E, U_n))$  by using intermediate distributions  $V_2, \dots, V_\ell$ . For  $2 \leq i \leq \ell$ , let  $V_i$  be the distribution of the random variable which takes values in  $\{0, 1\}^n$  and is defined by the following sampling:

1.  $x \| z_i \| \dots \| z_\ell \xleftarrow{\$} \{0, 1\}^{n + \frac{n}{r}(\ell - i + 1)}$
2.  $h_{i-1} \leftarrow x$
3. For  $j = i, \dots, \ell$ , do:
4.  $h_j \leftarrow \text{DM}^E((z_i \| i), h_{j-1})$
5.  $y \leftarrow h_\ell$

Note that the above definition is valid even for  $i = 1$ , and the resulting distribution is equal to  $V_1$ . By definition of our padding function `pad`, function distributions of the compression functions which process the  $i$ -th block and  $j$ -th block are essentially independent for  $i \neq j$ . Thus, by Lemma 5.2 we have

$$\Delta(V_i, V_{i+1}), \Delta(V_\ell, (U_E, U_n)) \leq \frac{2n+1}{2^{n/3r+1}} + \frac{n^2}{2^{n/r-2}} \quad (40)$$

for  $1 \leq i \leq \ell - 1$ . Hence  $\Delta(V_1, (U_E, U_n))$  is upper bounded by

$$\sum_{i=1}^{\ell-1} \Delta(V_i, V_{i+1}) + \Delta(V_\ell, (U_E, U_n)) \leq \ell \cdot \left( \frac{2n+1}{2^{n/3r+1}} + \frac{n}{2^{n/r-2}} \right). \quad (41)$$

Thus inequality (37) follows from inequality (39) and (41).  $\square$

## References

1. Alagic, G., Russell, A.: Quantum-secure symmetric-key cryptography based on hidden shifts. In: *Advances in Cryptology - EUROCRYPT 2017 - 36th Annual International Conference on the Theory and Applications of Cryptographic Techniques*, Paris, France, April 30 - May 4, 2017, Proceedings, Part III. pp. 65–93 (2017)
2. Ambainis, A., Rosmanis, A., Unruh, D.: Quantum attacks on classical proof systems: The hardness of quantum rewinding. In: *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014*, Philadelphia, PA, USA, October 18–21, 2014. pp. 474–483 (2014)
3. Anand, M.V., Targhi, E.E., Tabia, G.N., Unruh, D.: Post-quantum security of the CBC, CFB, OFB, CTR, and XTS modes of operation. In: *Post-Quantum Cryptography - 7th International Workshop, PQCrypto 2016*, Fukuoka, Japan, February 24–26, 2016, Proceedings. pp. 44–63 (2016)

4. Beals, R., Buhrman, H., Cleve, R., Mosca, M., de Wolf, R.: Quantum lower bounds by polynomials. In: 39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11, 1998, Palo Alto, California, USA. pp. 352–361 (1998)
5. Bennett, C.H., Bernstein, E., Brassard, G., Vazirani, U.V.: Strengths and weaknesses of quantum computing. *SIAM J. Comput.* **26**(5), 1510–1523 (1997)
6. Boneh, D., Dagdelen, Ö., Fischlin, M., Lehmann, A., Schaffner, C., Zhandry, M.: Random oracles in a quantum world. In: Advances in Cryptology - ASIACRYPT 2011 - 17th International Conference on the Theory and Application of Cryptology and Information Security, Seoul, South Korea, December 4-8, 2011. Proceedings. pp. 41–69 (2011)
7. Boneh, D., Zhandry, M.: Quantum-secure message authentication codes. In: Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings. pp. 592–608 (2013)
8. Boyer, M., Brassard, G., Høyer, P., Tapp, A.: Tight bounds on quantum searching. *Fortsch. Phys.* **46**(4-5), 493–505 (June 1998)
9. Carstens, T.V., Ebrahimi, E., Tabia, G.N., Unruh, D.: On quantum indistinguishability. IACR Cryptology ePrint Archive, Report 2018/257 (2018)
10. Czakajowski, J., Bruinderink, L.G., Hülsing, A., Schaffner, C., Unruh, D.: Post-quantum security of the sponge construction. In: Post-Quantum Cryptography - 9th International Conference, PQCrypto 2018, Fort Lauderdale, FL, USA, April 9-11, 2018, Proceedings. pp. 185–204 (2018)
11. Deutsch, D., Jozsa, R.: Rapid solution of problems by quantum computation. *Proc. R. Soc. Lond. A* **439**(1907), 553–558 (1992)
12. Goldwasser, S., Bellare, M.: Lecture notes on cryptography. Summer course Cryptography and computer security at MIT (1996-2008)
13. Grover, L.K.: A fast quantum mechanical algorithm for database search. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996. pp. 212–219 (1996)
14. Hassani, M.: Derangements and applications. *Journal of Integer Sequences* **6**(2) (2003)
15. Hülsing, A., Rijneveld, J., Song, F.: Mitigating multi-target attacks in hash-based signatures. In: Public-Key Cryptography - PKC 2016 - 19th IACR International Conference on Practice and Theory in Public-Key Cryptography, Taipei, Taiwan, March 6-9, 2016, Proceedings, Part I. pp. 387–416 (2016)
16. Kaplan, M., Leurent, G., Leverrier, A., Naya-Plasencia, M.: Breaking symmetric cryptosystems using quantum period finding. In: Robshaw, M., Katz, J. (eds.) CRYPTO 2016, Part II. LNCS, vol. 9815, pp. 207–237. Springer (2016)
17. Kelsey, J., Schneier, B.: Second preimages on n-bit hash functions for much less than  $2^n$  work. In: Advances in Cryptology - EUROCRYPT 2005, 24th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, Proceedings. pp. 474–490 (2005)
18. Kuwakado, H., Morii, M.: Quantum distinguisher between the 3-round Feistel cipher and the random permutation. In: IEEE International Symposium on Information Theory, ISIT 2010, June 13-18, 2010, Austin, Texas, USA, Proceedings. pp. 2682–2685 (2010)
19. Kuwakado, H., Morii, M.: Security on the quantum-type even-mansour cipher. In: Proceedings of the International Symposium on Information Theory and its Applications, ISITA 2012, Honolulu, HI, USA, October 28-31, 2012. pp. 312–316 (2012)

20. Mennink, B., Szepieniec, A.: XOR of PRPs in a quantum world. In: Lange, T., Takagi, T. (eds.) PQCrypto 2017. LNCS, vol. 10346, pp. 367–383. Springer (2017)
21. Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press (2010)
22. NIST: Fips pub 180-2. National Institute of Standards and Technology (2002)
23. NIST: Fips pub 202. National Institute of Standards and Technology (2014)
24. Patarin, J.: The "coefficients H" technique. In: Selected Areas in Cryptography, 15th International Workshop, SAC 2008, Sackville, New Brunswick, Canada, August 14-15, Revised Selected Papers. pp. 328–345 (2008)
25. Rogaway, P., Shrimpton, T.: Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance. In: Fast Software Encryption, 11th International Workshop, FSE 2004, Delhi, India, February 5-7, 2004, Revised Papers. pp. 371–388 (2004)
26. Shor, P.W.: Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.* **26**(5), 1484–1509 (1997)
27. Song, F.: A note on quantum security for post-quantum cryptography. In: Post-Quantum Cryptography - 6th International Workshop, PQCrypto 2014, Waterloo, ON, Canada, October 1-3, 2014. Proceedings. pp. 246–265 (2014)
28. Song, F., Yun, A.: Quantum security of NMAC and related constructions - PRF domain extension against quantum attacks. In: Advances in Cryptology - CRYPTO 2017 - 37th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 20-24, 2017, Proceedings, Part II. pp. 283–309 (2017)
29. Unruh, D.: Quantum position verification in the random oracle model. In: Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II. pp. 1–18 (2014)
30. Unruh, D.: Revocable quantum timed-release encryption. In: Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings. pp. 129–146 (2014)
31. Unruh, D.: Collapse-binding quantum commitments without random oracles. In: Advances in Cryptology - ASIACRYPT 2016 - 22nd International Conference on the Theory and Application of Cryptology and Information Security, Hanoi, Vietnam, December 4-8, 2016, Proceedings, Part II. pp. 166–195 (2016)
32. Unruh, D.: Computationally binding quantum commitments. In: Advances in Cryptology - EUROCRYPT 2016 - 35th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Vienna, Austria, May 8-12, 2016, Proceedings, Part II. pp. 497–527 (2016)
33. Zhandry, M.: How to record quantum queries, and applications to quantum indistinguishability. IACR Cryptology ePrint Archive, Report 2018/276 (2018)
34. Zhandry, M.: How to construct quantum random functions. In: 53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012. pp. 679–687 (2012)
35. Zhandry, M.: Secure identity-based encryption in the quantum random oracle model. In: Advances in Cryptology - CRYPTO 2012 - 32nd Annual Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2012. Proceedings. pp. 758–775 (2012)
36. Zhandry, M.: A note on the quantum collision and set equality problems. *Quantum Info. Comput.* **15**(7-8), 557–567 (May 2015)

## A Approximation of Napier's Constant $e$ and $1/e$ .

For convenience, we give approximation bounds for Napier's constant  $e$  and its inverse  $1/e$  as the following lemma.

**Lemma A.1.** *For any non-negative constant  $\lambda_0$ ,*

$$\sum_{i=\lambda_0}^{\infty} \frac{1}{i!} = \left| e - \sum_{i=0}^{\lambda_0-1} \frac{1}{i!} \right| \leq \frac{e}{\lambda_0!} \quad (42)$$

and

$$\left| \sum_{i=\lambda_0}^{\infty} \frac{(-1)^i}{i!} \right| = \left| \frac{1}{e} - \sum_{i=0}^{\lambda_0-1} \frac{(-1)^i}{i!} \right| \leq \frac{1}{\lambda_0!} \quad (43)$$

hold.

*Proof.* For the Napier's constant  $e$ , we have

$$\sum_{i=\lambda_0}^{\infty} \frac{1}{i!} \leq \frac{1}{\lambda_0!} \sum_{i=0}^{\infty} \frac{1}{i!} \leq \frac{1}{\lambda_0!} \cdot e.$$

For the inverse of Napier's constant  $1/e$ , we have

$$\left| \sum_{i=\lambda_0}^{\infty} \frac{(-1)^i}{i!} \right| = \frac{1}{\lambda_0!} \left| 1 - \frac{1}{\lambda_0 + 1} + \frac{1}{(\lambda_0 + 2)(\lambda_0 + 1)} - \dots \right| \leq \frac{1}{\lambda_0!},$$

which completes the proof.  $\square$

## B Proof of Lemma 2.1

*Proof (of Lemma 2.1).* First, we show that

$$!N = (N-1)!(N-1)!(N-2) \quad (44)$$

holds. This equation holds since

$$\begin{aligned} \text{Der}([N]) &= \prod_{i=2}^N \left( \{P_0 \in \text{Der}([N]) \mid P_0(1) = i \wedge P_0(i) = 1\} \right. \\ &\quad \left. \prod \{P_0 \in \text{Der}([N]) \mid P_0(1) = i \wedge P_0(i) \neq 1\} \right) \\ &\cong \prod_{i=2}^N \left( \text{Der}([N-2]) \prod \text{Der}([N-1]) \right) \end{aligned} \quad (45)$$

holds, where  $A \cong B$  means that  $A$  and  $B$  are isomorphic as sets.

Next, we show

$$!N = N! \cdot \sum_{i=0}^N \frac{(-1)^i}{i!} \quad (46)$$

by induction. If  $N = 1$  or  $N = 2$ , then (46) holds since  $|\text{Der}([1])| = 0$  and  $|\text{Der}([2])| = 1$ . Suppose (46) holds for  $1 \leq N \leq k-1$ . Then, from equation (44) we have

$$\begin{aligned} !k &= (k-1)!(k-1)!(k-2) \\ &= (k-1) \cdot \left( (k-1)! \cdot \sum_{i=0}^{k-1} \frac{(-1)^i}{i!} + (k-2)! \cdot \sum_{i=0}^{k-2} \frac{(-1)^i}{i!} \right) \\ &= k! \cdot \frac{k-1}{k} \left( \sum_{i=0}^k \frac{(-1)^i}{i!} - \frac{(-1)^k}{k!} \right) \\ &\quad + k! \cdot \frac{1}{k} \left( \sum_{i=0}^k \frac{(-1)^i}{i!} - \frac{(-1)^{k-1}}{(k-1)!} - \frac{(-1)^k}{k!} \right) \\ &= k! \cdot \sum_{i=0}^k \frac{(-1)^i}{i!} - \frac{(k-1)(-1)^k}{k} - (-1)^{k-1} - \frac{(-1)^k}{k} \\ &= k! \cdot \sum_{i=0}^k \frac{(-1)^i}{i!}. \end{aligned} \quad (47)$$

Thus equation (46) holds.

Finally, we show

$$N! \cdot \sum_{i=0}^N \frac{(-1)^i}{i!} = \left\lfloor \frac{N!}{e} + \frac{1}{2} \right\rfloor \quad (48)$$

holds. Equation (48) is equivalent to the following inequalities:

$$N! \cdot \sum_{i=0}^N \frac{(-1)^i}{i!} - \frac{1}{2} \leq \frac{N!}{e} < N! \cdot \sum_{i=0}^N \frac{(-1)^i}{i!} + \frac{1}{2}. \quad (49)$$

Furthermore, (49) holds if the inequality

$$\left| \frac{1}{e} - \sum_{i=0}^N \frac{(-1)^i}{i!} \right| < \frac{1}{2N!} \quad (50)$$

holds. In fact inequality (50) follows from Lemma A.1 for  $N \geq 1$ .  $\square$

## C Notes on Quantum Computation

Below we briefly explain basics on quantum computation. See textbooks ([21], for example) on quantum computation for more details. A quantum system



is a complex Hilbert space  $\mathcal{H}$  of dimension  $2^d$  for some integer  $d > 0$ . The inner product and norm of the Hilbert spaces are denoted as  $\langle \cdot | \cdot \rangle$  and  $\| \cdot \|$ , respectively. Each quantum state of the system is described as a unit vector  $|\phi\rangle$  (i.e.,  $\| |\phi\rangle \| = 1$ ), which we call *d-qubit quantum state*. An orthonormal basis of  $\mathcal{H}$  is fixed, and each vector of the basis is labeled as  $|x\rangle$ ,  $x \in \{0, 1\}^d$ . This basis  $\{|x\rangle\}_{x \in \{0, 1\}^d}$  is called as *computational basis* of  $\mathcal{H}$ . Each  $|x\rangle$  corresponds to a classical  $d$ -bit string  $x$ . The hybrid quantum system that are composed of two quantum subsystems  $\mathcal{H}_A$  and  $\mathcal{H}_B$  is described as the tensor product  $\mathcal{H}_A \otimes \mathcal{H}_B$ . For quantum states  $|\phi_A\rangle \in \mathcal{H}_A$  and  $|\phi_B\rangle \in \mathcal{H}_B$ , the joined state corresponds to the state  $|\phi_A\rangle |\phi_B\rangle := |\phi_A\rangle \otimes |\phi_B\rangle$ .

The map  $|x\rangle \mapsto |x_1\rangle |x_2\rangle \cdots |x_d\rangle$  determines an isomorphism of Hilbert spaces between  $\mathcal{H}$  and  $(\mathbb{C}^2)^{\otimes d} = \mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$ , where  $x = x_1 x_2 \cdots x_d$  and  $x_i \in \{0, 1\}$  for each  $i$ . Here we assume the inner product of  $\mathbb{C}^2$  is the usual Hermitian inner product defined by  $\langle \binom{a}{b} | \binom{c}{d} \rangle = \bar{a}c + \bar{b}d$ , where  $\bar{a}$  is the conjugate of  $a$ . Computational basis  $\{|0\rangle, |1\rangle\}$  of  $\mathbb{C}^2$  is set as  $|0\rangle := \binom{1}{0}$  and  $|1\rangle := \binom{0}{1}$ . We identify  $\mathcal{H}$  with  $\mathbb{C}^2 \otimes \cdots \otimes \mathbb{C}^2$  through this isomorphism. We regard  $|x\rangle$  as a column vector, and  $\langle x|$  as the conjugate row vector that corresponds to  $|x\rangle$ .

For each orthonormal basis  $S = \{|s_0\rangle, \dots, |s_{2^d-1}\rangle\}$ , we can perform *measurement* of a quantum state  $|\phi\rangle$  according to the basis  $S$ , and obtain the result  $|s_i\rangle$  with probability  $p_i = |\langle s_i | \phi \rangle|^2$ . (Actually measurements are generalized to positive-operator valued measurement (POVM). However, here we treat only the measurement using some fixed basis, for simplicity. See [21] for more details.) After the measurement, the state collapses into  $|s_i\rangle$ .

Quantum algorithms are described as unitary operators that act on some quantum systems. We consider that a quantum algorithm  $\mathcal{A}$  acts on a product space  $\mathcal{H}_{work} \otimes \mathcal{H}_{in} \otimes \mathcal{H}_{out}$ . Here,  $\mathcal{H}_{work}$ ,  $\mathcal{H}_{in}$ ,  $\mathcal{H}_{out}$  correspond to working register, input register, output register, respectively. Given input  $|\psi\rangle \in \mathcal{H}_{in}$ , the final state of  $\mathcal{A}$  before the measurement is  $|\phi_{last}\rangle = \mathcal{A} |0\rangle |\psi\rangle |0\rangle$ . If classical data  $y$  is given as an input, then it is converted to qubits  $|y\rangle \in \mathcal{H}_{in}$  and then given to  $\mathcal{A}$ . When we measure  $|\phi_{last}\rangle$ , we obtain  $(w, y, x)$  with probability  $|\langle w, y, x | \phi_{last} \rangle|^2$ , and the output of  $\mathcal{A}$  is  $x$ . In particular, output becomes  $x$  with probability  $\sum_{w, y} |\langle w, y, x | \phi_{last} \rangle|^2$ . If  $\mathcal{A}$  can access to a quantum oracle of a function  $O_f$ , then we consider that  $O_f$  acts on (a subspace of)  $\mathcal{H}_{work}$ . Quantum query algorithm and its states are modeled and described as in Section 2.

Suppose that the quantum system  $\mathcal{H}$  becomes the state  $|\phi_i\rangle$  with probability  $p_i$  ( $\sum_i p_i = 1$ ). This state is called *mixed state*, whereas a state which occurs with probability 1 is called *pure state*. The mixed state of  $\mathcal{H}$  is described as an Hermitian operator  $\rho$  called *density operator* defined as

$$\rho := \sum_i p_i |\phi_i\rangle \langle \phi_i|. \quad (51)$$

For example, the state of  $\mathcal{H}$  after measuring  $|\phi\rangle$  according to the basis  $S$  becomes the state  $|s_i\rangle$  with probability  $|\langle s_i | \phi \rangle|^2$ , which is described as the density operator  $\sum_i |\langle s_i | \phi \rangle|^2 |s_i\rangle \langle s_i|$ . If the system is initially in the state  $|\phi_i\rangle$  with probability  $p_i$ , then the state after an evolution described by an operator  $U$  becomes  $U |\phi_i\rangle$  with

probability  $p_i$ . In particular, the evolution of the mixed state by  $U$  is described as  $\rho \mapsto U\rho U^\dagger$ .

Generally, the probability that we obtain  $|s_j\rangle$  when we measure the mixed state  $\rho = \sum_i p_i |\phi_i\rangle \langle \phi_i|$  with the basis  $S$  is equal to  $\langle s_j | \rho | s_j \rangle$ . In fact, if we measure (with the basis  $S$ ) the mixed state that becomes the state  $|\phi_i\rangle$  with probability  $p_i$ , then we obtain the result  $|s_j\rangle$  with probability

$$\begin{aligned} \sum_i p_i |\langle s_j | \phi_i \rangle|^2 &= \sum_i p_i \langle s_j | \phi_i \rangle \langle \phi_i | s_j \rangle \\ &= \langle s_j | \left( \sum_i p_i |\phi_i\rangle \langle \phi_i| \right) | s_j \rangle = \langle s_j | \rho | s_j \rangle. \end{aligned} \quad (52)$$

For two mixed states  $\rho_1$  and  $\rho_2$  of a quantum system  $\mathcal{H}$ , *trace distance* between  $\rho_1$  and  $\rho_2$  is defined by

$$\mathbf{td}(\rho_1, \rho_2) := \frac{1}{2} \|\rho_1 - \rho_2\|_{tr} = \frac{1}{2} \text{Tr} \left[ \sqrt{(\rho_1 - \rho_2)^\dagger (\rho_1 - \rho_2)} \right] = \frac{1}{2} \sum_i |\lambda_i|, \quad (53)$$

where  $\|\cdot\|_{tr}$  is the trace norm, and  $\lambda_i$  are the eigenvalues of the Hermitian operator  $(\rho_1 - \rho_2)$ . Trace distance  $\mathbf{td}$  satisfies axioms of distance function. For pure state  $|\psi\rangle$ , we write  $\mathbf{td}(|\psi\rangle, \rho_2)$  instead of  $\mathbf{td}(|\psi\rangle \langle \psi|, \rho_2)$ , for simplicity. This paper uses two properties of trace distance. The first one is unitary invariance. That is,  $\mathbf{td}(U\rho_1 U^\dagger, U\rho_2 U^\dagger) = \mathbf{td}(\rho_1, \rho_2)$  holds for any unitary operation  $U$ . The second one is joint convexity. For sets of density matrices  $\{\rho_i\}_{i \in I}$ ,  $\{\sigma_i\}_{i \in I}$  and non-negative integers  $\{\alpha_i\}_{i \in I}$  such that  $\sum_i \alpha_i = 1$ ,

$$\mathbf{td}\left(\sum_i \alpha_i \rho_i, \sum_i \alpha_i \sigma_i\right) \leq \sum_i \alpha_i \mathbf{td}(\rho_i, \sigma_i) \quad (54)$$

holds (see Theorem 9.3 and equality (9.50) in [21]). In addition, we use the following lemma by Ambainis et al [2].

**Lemma C.1 ([2], Lemma 36).** *For pure states  $|\phi\rangle$  and  $|\psi\rangle$ ,  $\mathbf{td}(|\phi\rangle, |\psi\rangle) \leq \|\phi\rangle - |\psi\rangle\|$  holds.*

Trace distance  $\mathbf{td}$  is a generalization of classical total variation distance  $\Delta$ . For example, let  $D_1, D_2$  be distributions on the basis  $S = \{|s_0\rangle, \dots, |s_{2^d-1}\rangle\}$ , and let  $p_1^i = \Pr_{v \sim D_1}[v = |s_i\rangle]$ ,  $p_2^i = \Pr_{v \sim D_2}[v = |s_i\rangle]$ . Let  $\rho_1$  be the density operator which corresponds to the state that becomes  $|s_i\rangle$  with probability  $p_1^i$ . Similarly, let  $\rho_2$  be the density operator which corresponds to  $\{|s_i\rangle, p_2^i\}$ . Then  $\rho_1 - \rho_2$  is equal to  $\sum_i (p_1^i - p_2^i) |s_i\rangle \langle s_i|$ , of which eigenvalues are  $(p_1^i - p_2^i)$ . Thus we have  $\mathbf{td}(\rho_1, \rho_2) = \frac{1}{2} \sum_i |p_1^i - p_2^i| = \Delta(D_1, D_2)$ , which implies that the trace distance  $\mathbf{td}$  is a generalization of the classical statistical distance  $\Delta$ .

### C.1 Proof of Lemma 2.2

This section shows Lemma 2.2. Before give a proof of Lemma 2.2, we show the following lemma.

**Lemma C.2.** For any Hermitian matrix  $A$  and vector  $|\phi\rangle$ ,

$$|\langle\phi|A|\phi\rangle| \leq \langle\phi|\sqrt{A^\dagger A}|\phi\rangle$$

holds.

*Proof (of Lemma C.2).* Let  $\{\lambda_i\}$  be eigenvalues of  $A$  and  $\{|v_i\rangle\}$  be the corresponding orthonormal eigenvector basis such that  $A|v_i\rangle = \lambda_i|v_i\rangle$ . Then we have  $\sqrt{A^\dagger A}|v_i\rangle = |\lambda_i||v_i\rangle$ . Let  $\{\alpha_i\}$  be the set of complex numbers that satisfies  $|\phi\rangle = \sum_i \alpha_i |v_i\rangle$ . Then we have  $|\langle\phi|A|\phi\rangle| = \left| \sum_{i,j} \alpha_j^\dagger \alpha_i \langle v_j|A|v_i\rangle \right| = \left| \sum_i |\alpha_i|^2 \lambda_i \right| \leq \sum_i |\alpha_i|^2 |\lambda_i| = \langle\phi|\sqrt{A^\dagger A}|\phi\rangle$ , which completes the proof.  $\square$

Next, we show Lemma 2.2. That is, we show

$$\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq \text{td}(\rho_1^q, \rho_2^q) \leq \Delta(D_1, D_2) \quad (55)$$

holds for any quantum algorithm  $\mathcal{A}$  that makes at most  $q$  quantum queries, where  $\rho_1^q$  is the density operator defined by (2), and  $\rho_2^q$  is the density operator that is similarly defined according to the distribution  $D_2$ .

Remember that a quantum algorithm  $\mathcal{A}$  acts on a product space  $\mathcal{H}_{work} \otimes \mathcal{H}_{in} \otimes \mathcal{H}_{out}$ , and quantum oracle  $O_f$  acts on (a subspace of)  $\mathcal{H}_{work}$ . Since now we consider the adversaries that output 0 or 1, and take no input, below  $\mathcal{H}_{out}$  corresponds to a 1-qubit quantum system, and we omit  $\mathcal{H}_{in}$  for simplicity. (That is, we consider that  $\mathcal{A}$  acts on a space  $\mathcal{H}_{work} \otimes \mathbb{C}^2$ .) The probability that  $\mathcal{A}$  outputs a result  $b$  when  $f$  is chosen according to the distribution  $D_i$  is equal to  $\sum_w \langle w, b | \rho_i^q | w, b \rangle$ .

*Proof (of Lemma 2.2).* First, we show  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq \text{td}(\rho_1^q, \rho_2^q)$ . We have

$$\begin{aligned}
\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) &= \left| \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 1] - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 1] \right| \\
&= \frac{1}{2} \left| \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 1] - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 1] \right| \\
&\quad + \frac{1}{2} \left| \left( 1 - \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 1] \right) - \left( 1 - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 1] \right) \right| \\
&= \frac{1}{2} \left| \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 1] - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 1] \right| \\
&\quad + \frac{1}{2} \left| \Pr_{f \sim D_1} [\mathcal{A}^{O_f}() = 0] - \Pr_{g \sim D_2} [\mathcal{A}^{O_g}() = 0] \right| \\
&= \frac{1}{2} \left| \sum_w \langle w, 1 | \rho_1^q | w, 1 \rangle - \sum_w \langle w, 1 | \rho_2^q | w, 1 \rangle \right| \\
&\quad + \frac{1}{2} \left| \sum_w \langle w, 0 | \rho_1^q | w, 0 \rangle - \sum_w \langle w, 0 | \rho_2^q | w, 0 \rangle \right| \\
&= \frac{1}{2} \sum_{b=0,1} \left| \sum_w \langle w, b | \rho_1^q | w, b \rangle - \sum_w \langle w, b | \rho_2^q | w, b \rangle \right| \\
&\leq \frac{1}{2} \sum_{w,b} |\langle w, b | (\rho_1^q - \rho_2^q) | w, b \rangle| \tag{56}
\end{aligned}$$

If we let  $A_{x,y}$  to denote the  $(x,y)$ -component of a matrix  $A$ , then  $A_{x,y} = \langle x | A | y \rangle$  holds, and thus  $\text{Tr}[A] = \sum_x A_{x,x} = \sum_x \langle x | A | x \rangle$  holds. Therefore, from Lemma C.2 we have

$$\begin{aligned}
\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) &\leq \frac{1}{2} \sum_{w,b} |\langle w, b | (\rho_1^q - \rho_2^q) | w, b \rangle| \\
&\leq \frac{1}{2} \sum_{w,b} \langle w, b | \sqrt{(\rho_1^q - \rho_2^q)^\dagger (\rho_1^q - \rho_2^q)} | w, b \rangle \\
&= \frac{1}{2} \text{Tr} \left[ \sqrt{(\rho_1 - \rho_2)^\dagger (\rho_1 - \rho_2)} \right] = \text{td}(\rho_1, \rho_2). \tag{57}
\end{aligned}$$

Hence  $\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq \text{td}(\rho_1^q, \rho_2^q)$  holds.

Next, we show  $\text{td}(\rho_1^q, \rho_2^q) \leq \Delta(D_1, D_2)$ . By definition of  $\rho_1^q, \rho_2^q$ , and  $\text{td}$ , we have

$$\begin{aligned}
\text{td}(\rho_1^q, \rho_2^q) &= \frac{1}{2} \left\| \sum_f p_1^f |\phi_f^q\rangle \langle \phi_f^q| - \sum_g p_2^g |\phi_g^q\rangle \langle \phi_g^q| \right\|_{\text{tr}} \\
&= \frac{1}{2} \left\| \sum_f (p_1^f - p_2^f) |\phi_f^q\rangle \langle \phi_f^q| \right\|_{\text{tr}}, \tag{58}
\end{aligned}$$

where  $\|\cdot\|_{tr}$  denotes the trace norm. From subadditivity of trace norm we have

$$\begin{aligned} \frac{1}{2} \left\| \sum_f (p_1^f - p_2^f) |\phi_f^q\rangle \langle \phi_f^q| \right\|_{tr} &\leq \frac{1}{2} \sum_f (p_1^f - p_2^f) \left\| |\phi_f^q\rangle \langle \phi_f^q| \right\|_{tr} \\ &= \frac{1}{2} \sum_f (p_1^f - p_2^f) = \Delta(D_1, D_2). \end{aligned} \quad (59)$$

Hence  $\text{td}(\rho_1^q, \rho_2^q) \leq \Delta(D_1, D_2)$  holds, and Lemma 2.2 follows.  $\square$

## D Separation of Non-Invertibility and One-Wayness

This section roughly explains that non-invertibility and one-wayness are separate notions.

**Non-invertible  $\not\Rightarrow$  One-way.** Assume that there exists a function  $h : \{0, 1\}^n \rightarrow \{0, 1\}^n$  that is non-invertible. Define another function  $h' : \{0, 1\}^n \times \{0, 1\} \rightarrow \{0, 1\}^n$  by  $h'(x, 0) = 0^n$  and  $h'(x, 1) = h(x)$  for  $x \in \{0, 1\}^n$ . Then  $h'$  is obviously non-invertible. However,  $h'$  does not have one-wayness since  $\Pr_{x,b}[h'(x,b) = 0^n] \geq 1/2$  holds, and we can easily find a preimage of  $h'$  with high probability.

**One-way  $\not\Rightarrow$  Non-Invertible.** Assume that there exists a function  $h : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n$  that is one-way. Define another function  $h' : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^n \times \{0, 1\}$  by  $h'(a, 0^n) = (a, 0)$  and  $h'(a, b) = (h(a, b), 1)$  for  $a \in \{0, 1\}^n$  and  $b \in \{0, 1\}^n \setminus \{0^n\}$ . Then  $h'$  is obviously one-way. However,  $h'$  does not have non-invertibility since  $\Pr_{y,\beta}[\beta = 0] = 1/2$  holds, and we can easily invert  $h'$  with high probability.

## E Proof of Proposition 3.1

This section gives a proof of Proposition 3.1. In our proof,  $D_1$  and  $D_2$  can be any distributions, which generalize existing analyses (see proof of Lemma 37 in [2], proof of Lemma C.1 in [28], for example) that usually treat only some specific distributions. Theorem 1 in [15] gives similar result as Lemma 37 in [2], but uses different analyzing technique (Theorem 7.2 in [34] by Zhandry).

Remember that  $p_1^f, p_2^g, p^{f,g}$  denote  $\Pr_{F \sim D_1}[F = g]$ ,  $\Pr_{G \sim D_1}[G = f]$ , and  $\Pr_{(F,G) \sim \mathcal{D}}[(F,G) = (f,g)]$ , respectively. Boolean function  $\delta(f,g) : \{0, 1\}^n \rightarrow \{0, 1\}$  is defined by  $\delta(f,g)(x) = 1$  if and only if  $f(x) \neq g(x)$  for each pair  $(f,g)$ . If the oracle  $O_f$  is chosen according to the distribution  $D_1$ , the state of  $\mathcal{A}$  after the  $i$ -th query is denoted as  $\rho_1^i = \sum_f p_1^f |\phi_f^i\rangle \langle \phi_f^i|$ , where  $|\phi_f^i\rangle = U_i O_f U_{i-1} O_f \cdots O_f U_0 |0\rangle$ . Similarly,  $\rho_2^i$  denotes the state of  $\mathcal{A}$  that corresponds to the distribution  $D_2$ .

Let  $\Pi_\gamma$  denote the projection  $\sum_{z:\gamma(z)=1} |z\rangle\langle z| \otimes I$  for a function  $\gamma \in \text{Func}(\{0,1\}^n, \{0,1\})$ . If  $\gamma$  is the zero-function  $\mathbf{0}$  such that  $\gamma(x) = 0$  for any  $x$ , then we define  $\Pi_\gamma$  as the zero operator. Before showing Proposition 3.1, we prove the following proposition.

**Proposition E.1 (General bound).** *For any distributions  $D_1$  and  $D_2$ ,*

$$\text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) \leq 2 \sum_{f,g, 0 \leq i \leq q-1} p^{f,g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\|$$

holds.

*Proof (of Proposition E.1).* For  $f, g \in \text{Func}(\{0,1\}^n, \{0,1\}^c)$ , define

$$|\psi_{1,f,g}^i\rangle := |\phi_f^i\rangle, |\psi_{2,f,g}^i\rangle := |\phi_g^i\rangle. \quad (60)$$

Due to joint convexity of trace distance (see inequality (54)), we have

$$\begin{aligned} & \text{td}(\rho_1^q, \rho_2^q) \\ &= \text{td}\left(\sum_f p_1^f |\phi_f^q\rangle\langle\phi_f^q|, \sum_g p_2^g |\phi_g^q\rangle\langle\phi_g^q|\right) \\ &= \text{td}\left(\sum_f \left(\sum_g p^{f,g}\right) |\phi_f^q\rangle\langle\phi_f^q|, \sum_g \left(\sum_f p^{f,g}\right) |\phi_g^q\rangle\langle\phi_g^q|\right) \\ &= \text{td}\left(\sum_{f,g} p^{f,g} |\psi_{1,f,g}^q\rangle\langle\psi_{1,f,g}^q|, \sum_{f,g} p^{f,g} |\psi_{2,f,g}^q\rangle\langle\psi_{2,f,g}^q|\right) \\ &\leq \sum_{f,g} p^{f,g} \text{td}\left(|\psi_{1,f,g}^q\rangle\langle\psi_{1,f,g}^q|, |\psi_{2,f,g}^q\rangle\langle\psi_{2,f,g}^q|\right) \\ &= \sum_{f,g} p^{f,g} \text{td}\left(|\phi_f^q\rangle\langle\phi_f^q|, |\phi_g^q\rangle\langle\phi_g^q|\right). \end{aligned} \quad (61)$$

For  $f, g \in \text{Func}(\{0,1\}^n, \{0,1\}^c)$ , define  $f \oplus g \in \text{Func}(\{0,1\}^n, \{0,1\}^c)$  by  $(f \oplus g)(x) = f(x) \oplus g(x)$ . Then  $O_g O_f = O_{f \oplus g}$  holds since

$$O_g O_f |x\rangle|y\rangle = O_g |x\rangle|y \oplus f(x)\rangle = |x\rangle|y \oplus g(x) \oplus f(x)\rangle = O_{f \oplus g} |x\rangle|y\rangle \quad (62)$$

holds for any  $x \in \{0,1\}^n, y \in \{0,1\}^c$ . Thus, due to invariance of trace distance under unitary operations, we have

$$\begin{aligned} \text{td}(|\phi_f^{i+1}\rangle, |\phi_g^{i+1}\rangle) &= \text{td}(U_{i+1} O_f |\phi_f^i\rangle, U_{i+1} O_g |\phi_g^i\rangle) \\ &= \text{td}(O_f |\phi_f^i\rangle, O_g |\phi_g^i\rangle) \\ &\leq \text{td}(O_f |\phi_f^i\rangle, O_f |\phi_g^i\rangle) + \text{td}(O_f |\phi_g^i\rangle, O_g |\phi_g^i\rangle) \\ &= \text{td}(|\phi_f^i\rangle, |\phi_g^i\rangle) + \text{td}(O_g O_f |\phi_g^i\rangle, O_g O_g |\phi_g^i\rangle) \\ &= \text{td}(|\phi_f^i\rangle, |\phi_g^i\rangle) + \text{td}(O_{f \oplus g} |\phi_g^i\rangle, |\phi_g^i\rangle) \end{aligned} \quad (63)$$

for any  $i$ . Therefore, by induction on  $i$  we have

$$\mathbf{td}(|\phi_f^q\rangle, |\phi_g^q\rangle) \leq \sum_{i=0}^{q-1} \mathbf{td}(O_{f \oplus g} |\phi_g^i\rangle, |\phi_g^i\rangle). \quad (64)$$

Here, from Lemma C.1 we have

$$\begin{aligned} \mathbf{td}(O_{f \oplus g} |\phi_g^i\rangle, |\phi_g^i\rangle) &= \mathbf{td}(O_{f \oplus g} \Pi_{\delta(f,g)} |\phi_g^i\rangle + O_{f \oplus g} (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle, \\ &\quad \Pi_{\delta(f,g)} |\phi_g^i\rangle + (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle) \\ &\leq \|(O_{f \oplus g} \Pi_{\delta(f,g)} |\phi_g^i\rangle + O_{f \oplus g} (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle) \\ &\quad - (\Pi_{\delta(f,g)} |\phi_g^i\rangle + (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle)\|. \end{aligned} \quad (65)$$

In addition,

$$O_{f \oplus g} (I - \Pi_{\delta(f,g)}) = I - \Pi_{\delta(f,g)} \quad (66)$$

holds since, if  $f(x) \neq g(x)$ , then  $(I - \Pi_{\delta(f,g)}) |x\rangle |y\rangle = |x\rangle |y\rangle - |x\rangle |y\rangle = 0$  for any  $y \in \{0, 1\}^c$ , and if  $f(x) = g(x)$  then  $O_{f \oplus g} (I - \Pi_{\delta(f,g)}) |x\rangle |y\rangle = O_{f \oplus g} (|x\rangle |y\rangle - 0) = |x\rangle |y \oplus f(x) \oplus g(x)\rangle = |x\rangle |y\rangle = (I - \Pi_{\delta(f,g)}) |x\rangle |y\rangle$  for any  $y \in \{0, 1\}^c$ . Therefore we have

$$\begin{aligned} &\|(O_{f \oplus g} \Pi_{\delta(f,g)} |\phi_g^i\rangle + O_{f \oplus g} (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle) - (\Pi_{\delta(f,g)} |\phi_g^i\rangle + (I - \Pi_{\delta(f,g)}) |\phi_g^i\rangle)\| \\ &= \|O_{f \oplus g} \Pi_{\delta(f,g)} |\phi_g^i\rangle - \Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &\leq \|O_{f \oplus g} \Pi_{\delta(f,g)} |\phi_g^i\rangle\| + \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &= 2 \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\|. \end{aligned} \quad (67)$$

From inequality (64), (65), and (67) we have

$$\mathbf{td}(|\phi_f^q\rangle, |\phi_g^q\rangle) \leq 2 \sum_{0 \leq i \leq q-1} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\|. \quad (68)$$

Eventually,

$$\mathbf{td}(\rho_1^q, \rho_2^q) \leq 2 \sum_{f,g, 0 \leq i \leq q-1} p^{f,g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \quad (69)$$

follows from inequality (61) and (68). Since  $\text{Adv}_{D_1, D_2}^{\text{dist}}(q) \leq \mathbf{td}(\rho_1^q, \rho_2^q)$  holds from Lemma 2.2, Proposition E.1 follows.  $\square$

Next, we give a proof of Proposition 3.1. Remember that  $p_{\delta D|g}^{\gamma|\text{good}_\alpha^g}$  is independent of  $\gamma$ , by assumption.

*Proof (of Proposition 3.1).*

Let  $p^{f|g} := \Pr_{(F,G) \sim D}[F = f | G = g]$ . From Proposition E.1 we have

$$\begin{aligned} \text{Adv}_{D_1, D_2}^{\text{dist}}(\mathcal{A}) &\leq 2 \sum_{0 \leq i \leq q-1, f, g} p^{f,g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &\leq 2 \sum_{0 \leq i \leq q-1} \sum_g p_2^g \sum_f p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\|. \end{aligned} \quad (70)$$

Now we upper bound  $\sum_f p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\|$  for fixed  $g$ . Note that

$$\begin{aligned} \sum_{f:\delta(f,g)=\gamma} p^{f|g} &= \sum_{f:\delta(f,g)=\gamma} \Pr_{(F,G)\sim\bar{D}} [F = f|G = g] \\ &= \Pr_{(F,G)\sim\bar{D}} [\delta(F, G) = \gamma|G = g] = \Pr_{\Gamma\sim\delta D|g} [\Gamma = \gamma] \end{aligned} \quad (71)$$

hold for any  $\gamma \in \text{Func}(\{0, 1\}^n, \{0, 1\})$ . Thus we have

$$\begin{aligned} &\sum_f p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &= \sum_{f:\delta(f,g)\in\text{good}^g} p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| + \sum_{f:\delta(f,g)\in\text{bad}^g} p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &\quad + \sum_{f:\delta(f,g)=\mathbf{0}} p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| \\ &\leq \sum_{\alpha\in A_g} \sum_{f:\delta(f,g)\in\text{good}_\alpha^g} p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| + \sum_{f:\delta(f,g)\in\text{bad}^g} p^{f|g} \cdot 1 \\ &\quad + \sum_{f:\delta(f,g)=\mathbf{0}} p^{f|g} \cdot 0 \\ &= \sum_{\alpha\in A_g} \sum_{\gamma\in\text{good}_\alpha^g} \left( \sum_{f:\delta(f,g)=\gamma} p^{f|g} \right) \|\Pi_\gamma |\phi_g^i\rangle\| + \sum_{\gamma\in\text{bad}^g} \left( \sum_{f:\delta(f,g)=\gamma} p^{f|g} \right) \\ &= \sum_{\alpha\in A_g} \sum_{\gamma\in\text{good}_\alpha^g} \Pr_{\Gamma\sim\delta D|g} [\Gamma = \gamma] \|\Pi_\gamma |\phi_g^i\rangle\| \\ &\quad + \sum_{\gamma\in\text{bad}^g} \Pr_{(F,G)\sim\bar{D}} [\delta(F, G) = \gamma|G = g] \\ &= \sum_{\alpha\in A_g} \sum_{\gamma\in\text{good}_\alpha^g} \Pr_{\Gamma\sim\delta D|g} [\Gamma \in \text{good}_\alpha^g] \cdot \Pr_{\Gamma\sim\delta D|g} [\Gamma = \gamma|\Gamma \in \text{good}_\alpha^g] \|\Pi_\gamma |\phi_g^i\rangle\| \\ &\quad + \Pr_{(F,G)\sim\bar{D}} [\delta(F, G) \in \text{bad}^g|G = g] \\ &= \sum_{\alpha\in A_g} p_{\delta D|g}^{\text{good}_\alpha^g} \sum_{\gamma\in\text{good}_\alpha^g} p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\| \\ &\quad + \Pr_{(F,G)\sim\bar{D}} [\delta(F, G) \in \text{bad}^g|G = g]. \end{aligned} \quad (72)$$

Moreover, since  $p_{\delta D|g}^{\gamma|\text{good}_\alpha^g}$  does not depend on  $\gamma$ , from Jensen's inequality we have

$$\begin{aligned} \sum_{\gamma\in\text{good}_\alpha^g} p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\| &\leq \sqrt{\sum_{\gamma\in\text{good}_\alpha^g} p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\|^2} \\ &= \sqrt{p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \sum_{\gamma\in\text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\|^2}. \end{aligned} \quad (73)$$



For each  $x \in \{0, 1\}^n$ , define the projector  $\Pi_x$  by  $\Pi_x := |x\rangle\langle x| \otimes I$ . Then we have

$$\begin{aligned} \|\Pi_\gamma |\phi_g^i\rangle\|^2 &= \langle \phi_g^i | \Pi_\gamma^\dagger \Pi_\gamma | \phi_g^i \rangle = \langle \phi_g^i | \Pi_\gamma | \phi_g^i \rangle = \langle \phi_g^i | \left( \sum_{x:\gamma(x)=1} |x\rangle\langle x| \otimes I \right) | \phi_g^i \rangle \\ &= \left( \sum_{x:\gamma(x)=1} \langle \phi_g^i | |x\rangle\langle x| \otimes I | \phi_g^i \rangle \right) = \sum_{x:\gamma(x)=1} \|\Pi_x |\phi_g^i\rangle\|^2, \end{aligned} \quad (74)$$

and

$$\sum_x \|\Pi_x |\phi_g^i\rangle\|^2 = \sum_x \langle \phi_g^i | \Pi_x | \phi_g^i \rangle = \langle \phi_g^i | \sum_x \Pi_x | \phi_g^i \rangle = \langle \phi_g^i | \phi_g^i \rangle = 1. \quad (75)$$

Hence

$$\begin{aligned} \sum_{\gamma \in \text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\|^2 &= \sum_{\gamma \in \text{good}_\alpha^g} \sum_{x:\gamma(x)=1} \|\Pi_x |\phi_g^i\rangle\|^2 = \sum_x \sum_{\gamma(x)=1 \wedge \gamma \in \text{good}_\alpha^g} \|\Pi_x |\phi_g^i\rangle\|^2 \\ &= \sum_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}| \|\Pi_x |\phi_g^i\rangle\|^2 \\ &\leq \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}| \sum_x \|\Pi_x |\phi_g^i\rangle\|^2 \\ &= \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}| \end{aligned} \quad (76)$$

holds. Thus, from inequality (73), (76) we have

$$\sum_{\gamma \in \text{good}_\alpha^g} p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \|\Pi_\gamma |\phi_g^i\rangle\| \leq \sqrt{p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \cdot \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}|} \quad (77)$$

Therefore, from inequalities (77) and (72), inequality

$$\begin{aligned} \sum_f p^{f|g} \|\Pi_{\delta(f,g)} |\phi_g^i\rangle\| &\leq \sum_{\alpha \in A_g} p_{\delta D|g}^{\text{good}_\alpha^g} \sqrt{p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \cdot \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}|} \\ &\quad + \Pr_{(F,G) \sim \bar{D}} [\delta(F,G) \in \text{bad}^g \mid G = g] \end{aligned} \quad (78)$$

follows for each  $g \in \text{Func}(\{0, 1\}^n, \{0, 1\}^c)$ .

From inequality (70) and (78), we have

$$\begin{aligned}
& \text{Adv}_{D_1, D_2}^{dist}(\mathcal{A}) \\
& \leq 2 \sum_{0 \leq i \leq q-1} \sum_g p_2^g \sum_{\alpha \in A_g} p_{\delta D|g}^{\text{good}_\alpha^g} \sqrt{p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \cdot \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}|} \\
& \quad + 2 \sum_{0 \leq i \leq q-1} \sum_g p_2^g \Pr_{(F,G) \sim \bar{D}} [\delta(F,G) \in \text{bad}^g \mid G = g] \\
& \leq 2q \sum_g p_2^g \sum_{\alpha \in A_g} p_{\delta D|g}^{\text{good}_\alpha^g} \sqrt{p_{\delta D|g}^{\gamma|\text{good}_\alpha^g} \cdot \max_x |\{\gamma \in \text{good}_\alpha^g \mid \gamma(x) = 1\}|} \\
& \quad + 2q \sum_g \Pr_{(F,G) \sim \bar{D}} [\delta(F,G) \in \text{bad}^G \wedge G = g] \\
& \leq 2q \cdot \mathbf{E}_{G \sim D_2} \left[ \sum_{\alpha \in A_G} p_{\delta D|G}^{\text{good}_\alpha^G} \sqrt{p_{\delta D|G}^{\gamma|\text{good}_\alpha^G} \cdot \max_x |\{\gamma \in \text{good}_\alpha^G \mid \gamma(x) = 1\}|} \right] \\
& \quad + 2q \cdot \Pr_{(F,G) \sim \bar{D}} [(F,G) \in \text{bad}_{all}], \tag{79}
\end{aligned}$$

which completes the proof of Proposition 3.1.  $\square$

## F An Application of Our Generalized Tool

Here we give an application to demonstrate usefulness of our generalized tool (Proposition 3.1). For an integer  $i$ , define the distribution  $D_i^c$  as the uniform distribution on  $\{f \in \text{Func}(\{0,1\}^n, \{0,1\}) \mid |f^{-1}(1)| = i\}$ . Fix parameters  $s, t$  such that  $0 \leq s < 2^n$  and  $0 < t \leq 2^n - s$ . We consider how difficult it is to solve the following problem:

*Problem F.1.* Suppose that  $f$  is chosen according to  $D_{s+t}^c$  or  $D_s^c$ , and we are given the oracle of  $f$ . Then, judge according to which distribution  $f$  is chosen.

Note that, if  $s = 0$ , then this problem matches the famous database search problem (of decision version). In this case, the tight bound of the number of query to solve Problem F.1 is known as  $\Theta(\sqrt{2^n/t})$  [8] (upper bound is given by Grover's algorithm). Moreover, if  $s = t = 2^{n-1}$ , then Problem F.1 can be solved with only *one* query by the Deutsch-Jozsa algorithm [11].

Below, firstly we show that our simplified tool (Proposition 3.2) gives an almost tight bound for the number of queries to solve Problem F.1 in the case  $s = 0$ . Secondly we explain the limitation of the simplified tool with the case  $s > 0$  as an example. Finally we show usefulness of our generalized tool, by showing that the generalized case gives a meaningful upper bound of oracle distinguishing advantage for Problem F.1, even if  $s$  is large.

**Application of the simplified tool to the case  $s = 0$ .** If  $s = 0$ , then  $D_s^c$  becomes the degenerate distribution with support on the zero function, and thus

we can apply our simplified tool to upper bound oracle distinguishing advantages between  $D_t^c$  and  $D_0^c$ . We show the following claim.

*Claim.* If  $s = 0$ , then oracle distinguishing advantage for Problem F.1 is upper bounded as

$$\text{Adv}_{D_t^c, D_0^c}^{\text{dist}}(q) \leq 2q\sqrt{\frac{t}{2^n}}. \quad (80)$$

In particular, an adversary has to make  $\Omega(\sqrt{2^n/t})$  queries to distinguish  $D_t^c$  from  $D_0^c$  with constant probability.

*Proof.* Define  $\text{good} = \text{good}_t := \{f \mid |f^{-1}(1)| = t\}$ ,  $\text{good}_\alpha := \emptyset$  for  $\alpha \neq t$ , and  $\text{bad} := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus (\text{good} \cup \{\mathbf{0}\})$ . Then  $\Pr_{F \sim D_1}[F = f \mid F \in \text{good}_f] = \binom{2^n}{t}^{-1}$  is independent on  $f$ , and we can apply Proposition 3.2.

Now we have  $\Pr_{F \sim D_1}[F \in \text{bad}] = 0$ ,  $p_1^{\text{good}_t} = \Pr_{F \sim D_1}[F \in \text{good}_t] = 1$ ,  $p_1^{\text{good}_\alpha} = \Pr_{F \sim D_1}[F \in \text{good}_\alpha] = 0$  for  $\alpha \neq t$ , and  $p_1^{f|\text{good}_t} = \Pr_{F \sim D_1}[F = f \mid F \in \text{good}_t] = \binom{2^n}{t}^{-1}$  for  $f \in \text{good}_t$ . In addition,  $|\{f \in \text{good}_t \mid f(x) = 1\}| = \binom{2^n-1}{t-1}$  holds for any  $x \in \{0, 1\}^n$ . Thus, from Proposition 3.2, we have

$$\begin{aligned} & \text{Adv}_{D_t^c, D_0^c}^{\text{dist}}(\mathcal{A}) \\ & \leq 2q \cdot \sum_{\alpha} p_1^{\text{good}_\alpha} \sqrt{p_1^{f|\text{good}_\alpha} \cdot \max_x |\{f \in \text{good}_\alpha \mid f(x) = 1\}|} \\ & \quad + 2q \cdot \Pr_{f \sim D_1}[f \in \text{bad}] \\ & \leq 2q \cdot 1 \cdot \sqrt{p_1^{f|\text{good}_t} \cdot \max_x |\{f \in \text{good}_t \mid f(x) = 1\}|} + 2q \cdot 0 \\ & \leq 2q \cdot \sqrt{\binom{2^n}{t}^{-1} \cdot \binom{2^n-1}{t-1}} = 2q \cdot \sqrt{\frac{t}{2^n}} \end{aligned} \quad (81)$$

for any quantum adversary  $\mathcal{A}$  that makes at most  $q$  quantum queries.  $\square$

**Limitation of the simplified tool.** Next, we explain the limitation of the simplified tool. If we want to upper bound  $\text{Adv}_{D_{s+t}^c, D_s^c}^{\text{dist}}$  for the case  $s > 0$  only with the simplified tool, then we have to additionally introduce the degenerate distribution  $D_0^c$  with support on the zero function, since the simplified tool can only be applied to the case that one of the two distributions is  $D_0^c$ . Thus, from inequality (80), the upper bound which we can obtain only with the simplified tool becomes

$$\begin{aligned} \text{Adv}_{D_{s+t}^c, D_s^c}^{\text{dist}}(q) & \leq \text{Adv}_{D_{s+t}^c, D_0^c}^{\text{dist}}(q) + \text{Adv}_{D_0^c, D_s^c}^{\text{dist}}(q) \\ & \leq 2q\sqrt{\frac{s+t}{2^n}} + 2q\sqrt{\frac{s}{2^n}} = O\left(q\sqrt{\frac{s+t}{2^n}}\right). \end{aligned} \quad (82)$$

However, this bound (82) does not reflect the intuition that Problem F.1 will be hard to solve if  $t$  is small, even if  $s$  is large. For example, if  $s = 2^{n-1}$  and  $t = 1$ , then the right hand side of inequality (82) becomes  $O(q)$ , which is a meaningless bound.

**Usefulness of the generalized tool.** Next, we show that the generalized tool gives a better upper bound than (82). We show the following claim.

*Claim.* Oracle distinguishing advantage for Problem F.1 is upper bounded as

$$\text{Adv}_{D_{s+t}^c, D_s^c}^{dist}(q) \leq 2q \sqrt{\frac{t}{2^n - s}}. \quad (83)$$

In particular, an adversary has to make  $\Omega(\sqrt{(2^n - s)/t})$  queries to distinguish  $D_{s+t}^c$  from  $D_s^c$  with constant probability.

*Proof (of Claim).* Let  $\bar{D}$  be the distribution on  $\text{Func}(\{0, 1\}^n, \{0, 1\}) \times \text{Func}(\{0, 1\}^n, \{0, 1\})$  which is defined by the following sampling:

1.  $g \xleftarrow{D_s^c} \text{Func}(\{0, 1\}^n, \{0, 1\})$ . Let  $x_1^g, \dots, x_s^g$  be the preimages of 1 by  $g$ .
2. Choose  $y_1, \dots, y_t$ , which are different from each other, from  $\{0, 1\}^n \setminus \{x_1^g, \dots, x_s^g\}$  uniformly at random.
3. Define  $f$  by  $f(x) = 1$  if and only if  $x = x_i^g$  or  $x = y_j$  for some  $i, j$ .
4. Return  $(f, g)$ .

Then obviously  $D_{s+t}^c(f) = \sum_g \bar{D}(f, g)$  for any  $f$  and  $D_s^c(g) = \sum_f \bar{D}(f, g)$  for any  $g$  hold.

For each fixed  $g$ , define  $A_g := \{t\}$  and

$$\text{good}_t^g := \{\gamma \in \text{Func}(\{0, 1\}^n, \{0, 1\}) \mid |\gamma^{-1}(1)| = t \wedge \gamma^{-1}(1) \cap g^{-1}(1) = \emptyset\},$$

$$\text{good}^g := \text{good}_t^g,$$

$$\text{bad}^g := \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus (\text{good}^g \cup \{\mathbf{0}\}).$$

Then we have

$$\max_x |\{\gamma \in \text{good}_t^g \mid \gamma(x) = 1\}| = \binom{2^n - s - 1}{t - 1}. \quad (84)$$

In addition, by straightforward calculations it can be shown that

$$p_{\delta D|g}^{\gamma \text{good}_t^g} = \binom{2^n - s}{t}^{-1}, p_{\delta D|g}^{\text{good}_t^g} = 1, \Pr_{(F,G) \sim \bar{D}}[(F,G) \in \text{bad}_{all}] = 0. \quad (85)$$

Thus, from Proposition 3.1 we have

$$\begin{aligned} \text{Adv}_{D_{s+t}^c, D_s^c}^{dist}(q) &\leq 2q \mathbf{E}_{g \sim D_2} \left[ 1 \cdot \sqrt{\binom{2^n - s - 1}{t - 1} \cdot \binom{2^n - s}{t}^{-1}} \right] \\ &= 2q \sqrt{\frac{t}{2^n - s}}, \end{aligned} \quad (86)$$

which is the desired bound.  $\square$

Now, the new bound (83) reflects the intuition that problem will be hard to solve if  $t$  is small. For example, if  $s = 2^{n-1}$ , then the right hand side of inequality (83) becomes  $O(\frac{q\sqrt{t}}{2^{\frac{(n-1)}{2}}})$ . In the case  $t = 2^{n-1}$ , then  $O(\frac{q\sqrt{t}}{2^{\frac{(n-1)}{2}}}) = O(q)$  matches the bound by the Deutsch-Jozsa algorithm, and  $O(\frac{q\sqrt{t}}{2^{\frac{(n-1)}{2}}})$  decreases as  $t$  decreases, which reflects our intuition. Thus the bound (83) can be regarded as a meaningful bound.

*Remark F.1.* Our bound (83) is asymmetric: The situation  $(s, t)$  and  $(2^n - s - t, t)$  are the same up to complementing the functions, but our bound for the situation  $(2^n - s - t, t)$  differs from the bound (83) for  $(s, t)$ . This means that our bound is not tight.

## G Proof of Lemma 4.4

*Proof (of Lemma 4.4).* Since  $\Delta(U^\pm, U_1^\pm) = \Delta(U, U_1)$  holds, below we upper bound  $\Delta(U, U_1)$ . In this proof, let  $\text{good} \subset \text{Perm}(\{0, 1\}^n)$  denote the subset defined by  $P \in \text{good}$  if and only if the number of fixed points of  $P$  is less than  $2n$ . In addition, define a set  $\text{bad}$  by  $\text{bad} := \text{Perm}(\{0, 1\}^n) \setminus \text{good}$ .

Then we have

$$\begin{aligned} \Delta(U, U_1) &= \frac{1}{2} \sum_{P \in \text{good}} |\Pr_{\sigma \sim U}[\sigma = P] - \Pr_{\sigma \sim U_1}[\sigma = P]| \\ &\quad + \frac{1}{2} \sum_{P \in \text{bad}} |\Pr_{\sigma \sim U}[\sigma = P] - \Pr_{\sigma \sim U_1}[\sigma = P]| \\ &\leq \frac{1}{2} \left| \sum_{P \in \text{good}} \Pr_{\sigma \sim U_1}[\sigma = P] - \frac{1}{2^n} \right| + \frac{1}{2} \Pr_{P \sim U}[P \in \text{bad}] + \frac{1}{2} \Pr_{P \sim U_1}[P \in \text{bad}]. \end{aligned} \tag{87}$$

Next, we upper bound  $\frac{1}{2} \Pr_{P \sim U}[P \in \text{bad}] + \frac{1}{2} \Pr_{P \sim U_1}[P \in \text{bad}]$ . By definition of  $U_1$ ,

$$\Pr_{P \sim U_1}[P \in \text{bad}] = \Pr_{P_0 \sim U_0, f \sim \Lambda}[\Phi(P_0, f) \in \text{bad}]$$

holds. In addition, by definition of  $\Phi$ , the number of fixed points of  $\Phi(P_0, f)$  is less than or equal to  $2 \cdot |f^{-1}(1)|$ . Therefore  $\Pr_{P_0 \sim U_0, f \sim \Lambda}[\Phi(P_0, f) \in \text{bad}] \leq \Pr_{f \sim \Lambda}[|f^{-1}(1)| \geq n] = \Pr_{\lambda \sim D_{\text{num}}}[\lambda \geq n]$ . Thus, from inequality (20), we have

$$\begin{aligned} \frac{1}{2} \Pr_{P \sim U}[P \in \text{bad}] + \frac{1}{2} \Pr_{P \sim U_1}[P \in \text{bad}] &\leq \frac{1}{2} \Pr_{\lambda \sim D_{\text{num}}}[\lambda \geq 2n] + \frac{1}{2} \Pr_{\lambda \sim D_{\text{num}}}[\lambda \geq n] \\ &\leq \Pr_{\lambda \sim D_{\text{num}}}[\lambda \geq n] \leq \sum_{\lambda \geq n} \frac{e+1}{e} \frac{1}{\lambda!} \\ &\leq \frac{1+e}{n!}, \end{aligned} \tag{88}$$

where we used the fact  $\sum_{\lambda \geq n} 1/\lambda! \leq e/n!$  (see Lemma A.1).

Next, we upper bound  $\frac{1}{2} \left| \sum_{P \in \text{good}} \Pr_{\sigma \sim U'_1} [\sigma = P] - \frac{1}{2^{n!}} \right|$ . Define subsets  $\text{irr}_1, \text{irr}_2, \text{irr} \in \text{Der}(\{0, 1\}^n) \times \text{Func}(\{0, 1\}^n, \{0, 1\})$  as follows:

$$\text{irr}_1 := \{(P_0, f) \mid P_0(x) = y \text{ for some } x, y \in f^{-1}(1)\}, \quad (89)$$

$$\text{irr}_2 := \{(P_0, f) \mid P_0^2(x) = x \text{ for some } x \in f^{-1}(1)\}. \quad (90)$$

In addition, let  $\text{irr} := \text{irr}_1 \cup \text{irr}_2$  and  $\text{reg} := \text{Der}(\{0, 1\}^n) \times \text{Func}(\{0, 1\}^n, \{0, 1\}) \setminus \text{irr}$  ( $\text{irr}$  and  $\text{reg}$  means irregular and regular, respectively).

Intuitively,  $\text{irr}_1$  is an irregular event that “ $\Phi$  does not matches  $\Phi'$ ”. That is,  $\Phi(P_0, f) = \Phi'(P_0, f)$  holds if  $(P_0, f) \notin \text{irr}_1$  by definition of  $\Phi$  and  $\Phi'$ . In addition,  $\text{irr}_2$  is an irregular event that “ $\Phi(P_0, f)$  have unexpected fixed points”<sup>2</sup>. That is, the number of fixed points of  $\Phi(P_0, f)$  becomes larger than  $|f^{-1}(1)|$  if  $(P_0, f) \in \text{irr}_2$ . If  $(P_0, f) \notin \text{irr}_1 \vee \text{irr}_2$ , which is equivalent to  $(P_0, f) \in \text{reg}$ , then the number of fixed points of  $\Phi(P_0, f)$  (and  $\Phi'(P_0, f)$ ) matches  $|f^{-1}(1)|$ .

Now we have

$$\begin{aligned} & \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{\sigma \sim U'_1} [\sigma = P] - \frac{1}{2^{n!}} \right| \\ &= \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P] - \frac{1}{2^{n!}} \right| \\ &\leq \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^{n!}} \right| \\ &+ \frac{1}{2} \sum_{P \in \text{good}} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{irr}] \\ &= \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^{n!}} \right| \\ &+ \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}]. \end{aligned} \quad (91)$$

From inequality (87), (88), and (91) we have,

$$\begin{aligned} \Delta(U, U'_1) &\leq \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^{n!}} \right| \\ &+ \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}] + \frac{1+e}{n!}. \end{aligned} \quad (92)$$

Next, we upper bound the term  $\frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^{n!}} \right|$ . We use the following claim.

<sup>2</sup> In fact  $\text{irr}_2$  is just the event that  $P_0$  have 2-cycles, though, it holds that there is no unexpected fixed point if the event  $\neg(\text{irr}_1 \vee \text{irr}_2)$  occurs.

*Claim.* For each fixed  $P \in \text{Perm}(\{0, 1\}^n)$ ,

$$\Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] = \frac{(2^n - \lambda)! \cdot (2^n - \lambda)}{2^n! \cdot 2^n \cdot (2^n - 2\lambda)!} \quad (93)$$

holds, where  $\lambda$  is the number of fixed points of  $P$ .

*Proof (of Claim).* If  $(P_0, f) \in \text{reg}$ , then the set of fixed points of  $\Phi(P_0, f)$  matches the set  $f^{-1}(1)$ , by definition of  $\text{reg}$  and  $\Phi$ . Thus, for each fixed permutation  $P$ , there exists unique function  $f_P$  that satisfies  $\Phi(P_0, f_P) = P \wedge (P_0, f_P) \in \text{reg}$  for some derangement  $P_0$ .

Now we count the number of derangement  $P_0$  that satisfies  $\Phi(P_0, f_P) = P \wedge (P_0, f_P) \in \text{reg}$ . Let  $x_1, \dots, x_\lambda$  be the fixed points of  $P$ . In the graph  $G_P$  which is associated to  $P$ , let us choose an ordered sequence of  $\lambda$  edges  $S = (y_1 \rightarrow y'_1, \dots, y_\lambda \rightarrow y'_\lambda)$  which are not self loops. Then, modify the graph  $G_P$  as follows: for each  $1 \leq i \leq \lambda$ , remove the edges  $x_i \rightarrow x_i, y_i \rightarrow y'_i$  and add new edges  $y_i \rightarrow x_i, x_i \rightarrow y'_i$ . Then the new graph has no self loop and corresponds to a derangement  $P_{0,S}$ , and this derangement  $P_{0,S}$  satisfies  $\Phi(P_{0,S}, f_P) = P \wedge (P_{0,S}, f_P) \in \text{reg}$ . This mapping  $S \mapsto P_{0,S}$  gives a one-to-one correspondence between ordered sequence  $S$  of  $\lambda$  edges which are not self loops, and derangement  $P_0$  that satisfies  $\Phi(P_0, f_P) = P \wedge (P_0, f_P) \in \text{reg}$ . Therefore, the number of derangement  $P_0$  that satisfies  $\Phi(P_0, f_P) = P \wedge (P_0, f_P) \in \text{reg}$  is

$$\binom{2^n - \lambda}{\lambda} \cdot \lambda!. \quad (94)$$

Hence we have

$$\begin{aligned} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] &= \sum_{\substack{(P_0, f): \Phi(P_0, f) = P \\ \wedge (P_0, f) \in \text{reg}}} \Pr_{\substack{\sigma \sim U_0 \\ F \sim \Lambda}} [\sigma = P_0 \wedge F = f] \\ &= \sum_{\substack{P_0: \Phi(P_0, f_P) = P \\ \wedge (P_0, f_P) \in \text{reg}}} \Pr_{\substack{\sigma \sim U_0 \\ F \sim \Lambda}} [\sigma = P_0 \wedge F = f_P] \\ &= \binom{2^n - \lambda}{\lambda} \cdot \lambda! \cdot \frac{1}{2^n!} \frac{!(2^n - \lambda)}{2^n!} \\ &= \frac{(2^n - \lambda)! \cdot (2^n - \lambda)}{2^n! \cdot 2^n \cdot (2^n - 2\lambda)!}, \end{aligned}$$

which completes the proof of the claim.  $\square$

Note that the number of permutations that has exactly  $\lambda$  fixed points is  $\binom{2^n}{\lambda} \cdot (2^n - \lambda)!$ . From the above claim, now we have

$$\begin{aligned}
& \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^n!} \right| \\
&= \frac{1}{2} \sum_{\lambda < 2n} \left( \sum_{P: \lambda \text{ fixed points}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^n!} \right| \right) \\
&= \frac{1}{2} \sum_{\lambda < 2n} \left| \frac{(2^n - \lambda)! \cdot (2^n - \lambda)}{2^n! \cdot !2^n \cdot (2^n - 2\lambda)!} - \frac{1}{2^n!} \right| \cdot \binom{2^n}{\lambda} \cdot (2^n - \lambda) \\
&= \frac{1}{2} \sum_{\lambda < 2n} \frac{\binom{2^n}{\lambda} \cdot (2^n - \lambda)}{2^n!} \left| 1 - \frac{(2^n - \lambda)! \cdot (2^n - \lambda)}{!2^n \cdot (2^n - 2\lambda)!} \right| \\
&= \frac{1}{2} \sum_{\lambda < 2n} \frac{!(2^n - \lambda)}{\lambda! \cdot (2^n - \lambda)!} \left| 1 - \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)! \cdot (2^n - \lambda)!}{2^n! \cdot (2^n - 2\lambda)!} \right| \quad (95)
\end{aligned}$$

Note that

$$1 - \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)! \cdot (2^n - \lambda)!}{2^n! \cdot (2^n - 2\lambda)!} \geq 0 \quad (96)$$

holds for  $\lambda < 2n$  and  $n \geq 32$ , since

$$\begin{aligned}
& \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)! \cdot (2^n - \lambda)!}{2^n! \cdot (2^n - 2\lambda)!} \\
&= \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)(2^n - \lambda - 1) \cdots (2^n - 2\lambda + 1)}{2^n(2^n - 1) \cdots (2^n - \lambda + 1)} \\
&= \frac{2^n!}{[2^n!/e + 1/2]} \cdot \frac{\lfloor (2^n - \lambda)!/e + 1/2 \rfloor}{(2^n - \lambda)!} \cdot \prod_{i=0}^{\lambda-1} \left( 1 - \frac{\lambda}{2^n - i} \right) \\
&\leq \frac{2^n!}{2^n!/e} \cdot \frac{(2^n - \lambda)!/e + 1}{(2^n - \lambda)!} \cdot \left( 1 - \frac{\lambda}{2^n} \right) \\
&= \left( 1 + \frac{e}{(2^n - \lambda)!} \right) \cdot \left( 1 - \frac{\lambda}{2^n} \right) \leq 1 - \frac{\lambda}{2^n} + \frac{e}{(2^n - \lambda)!} \\
&\leq 1 - \frac{1}{2^n} + \frac{e}{(2^n - 2n)!} \leq 1. \quad (97)
\end{aligned}$$



Furthermore, for  $\lambda < 2n$  and  $n \geq 32$  we have

$$\begin{aligned}
& \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)! \cdot (2^n - \lambda)!}{2^n! \cdot (2^n - 2\lambda)!} \\
&= \frac{2^n!}{[2^n!/e + 1/2]} \cdot \frac{[(2^n - \lambda)!/e + 1/2]}{(2^n - \lambda)!} \cdot \prod_{i=0}^{\lambda-1} \left(1 - \frac{\lambda}{2^n - i}\right) \\
&\geq \frac{2^n!}{2^n!/e + 1} \cdot \frac{(2^n - \lambda)!/e - 1}{(2^n - \lambda)!} \cdot \left(1 - \frac{\lambda}{2^n - \lambda + 1}\right)^\lambda \\
&\geq \frac{(2^n - \lambda)!}{(2^n - \lambda)!/e + 1} \cdot \frac{(2^n - \lambda)!/e - 1}{(2^n - \lambda)!} \cdot \left(1 - \frac{\lambda^2}{2^n - \lambda + 1}\right) \\
&\geq \left(1 - \frac{2e}{(2^n - \lambda)! + e}\right) \cdot \left(1 - \frac{\lambda^2}{2^n - \lambda + 1}\right) \\
&\geq 1 - \frac{2e}{(2^n - \lambda)! + e} - \frac{\lambda^2}{2^n - \lambda + 1},
\end{aligned}$$

which implies that

$$1 - \frac{2^n!}{!2^n} \cdot \frac{!(2^n - \lambda)}{(2^n - \lambda)!} \cdot \frac{(2^n - \lambda)! \cdot (2^n - \lambda)!}{2^n! \cdot (2^n - 2\lambda)!} \leq \frac{2e}{(2^n - \lambda)! + e} + \frac{\lambda^2}{2^n - \lambda + 1}. \quad (98)$$

From inequalities (95), (96), and (98), we have

$$\begin{aligned}
& \frac{1}{2} \sum_{P \in \text{good}} \left| \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) = P \wedge (P_0, f) \in \text{reg}] - \frac{1}{2^n!} \right| \\
&\leq \frac{1}{2} \sum_{\lambda < 2n} \frac{!(2^n - \lambda)}{\lambda! \cdot (2^n - \lambda)!} \cdot \left( \frac{2e}{(2^n - \lambda)! + e} + \frac{\lambda^2}{2^n - \lambda + 1} \right) \\
&\leq \frac{1}{2} \sum_{\lambda < 2n} \frac{(2^n - \lambda)!/e + 1}{(2^n - \lambda)!} \cdot \left( \frac{2e}{(2^n - \lambda)! + e} + \frac{\lambda^2}{2^n - \lambda + 1} \right) \\
&= \frac{1}{2} \sum_{\lambda < 2n} \left( \frac{1}{e} + \frac{1}{(2^n - \lambda)!} \right) \cdot \frac{\lambda^2}{2^n - \lambda + 1} \cdot \left( \frac{2e}{(2^n - \lambda)! + e} \cdot \frac{2^n - \lambda + 1}{\lambda^2} + 1 \right) \\
&\leq \frac{1}{2} \sum_{\lambda < 2n} \frac{2\lambda^2}{2^n - \lambda + 1} \leq 2n \cdot \frac{(2n)^2}{2^n - 2n + 1} = \frac{8n^3}{2^n - 2n + 1}. \quad (99)
\end{aligned}$$

From inequalities (92) and (99), now we have

$$\Delta(U, U'_1) \leq \frac{8n^3}{2^n - 2n + 1} + \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}] + \frac{1+e}{n!}. \quad (100)$$

Next, we upper bound  $\frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}]$ . Note that

$$\Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}] \leq \Pr_{P_0 \sim U_0, f \sim \Lambda} [|f^{-1}(1)| < 2n \wedge (P_0, f) \in \text{irr}]$$

holds, since the number of fixed points of  $\Phi(P_0, f)$  is always greater than or equal to  $|f^{-1}(1)|$ . Thus we have

$$\begin{aligned} & \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}] \\ & \leq \frac{1}{2} \sum_{\lambda < 2n} \Pr_{P_0 \sim U_0, f \sim \Lambda} [|f^{-1}(1)| = \lambda \wedge (P_0, f) \in \text{irr}]. \end{aligned}$$

Moreover, for a fixed  $\lambda$ , we have

$$\begin{aligned} & \Pr_{P_0 \sim U_0, f \sim \Lambda} [|f^{-1}(1)| = \lambda \wedge (P_0, f) \in \text{irr}] \\ & \leq \sum_{f: |f^{-1}(1)| = \lambda} \Pr_{P_0 \sim U_0, F \sim \Lambda} [F = f \wedge (P_0, f) \in \text{irr}] \\ & = \sum_{f: |f^{-1}(1)| = \lambda} \Pr_{P_0 \sim U_0} [(P_0, f) \in \text{irr}] \Pr_{F \sim \Lambda} [F = f]. \end{aligned}$$

Hence

$$\begin{aligned} & \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}] \\ & \leq \sum_{\lambda < 2n} \sum_{f: |f^{-1}(1)| = \lambda} \Pr_{P_0 \sim U_0} [(P_0, f) \in \text{irr}] \Pr_{F \sim \Lambda} [F = f] \quad (101) \end{aligned}$$

holds. For a fixed  $f$ , let  $x_1, \dots, x_\lambda$  be the preimages of 1 by  $f$ . Then we have

$$\begin{aligned} \Pr_{P_0 \sim U_0} [(P_0, f) \in \text{irr}] & \leq \Pr_{P_0 \sim U_0} [(P_0, f) \in \text{irr}_1] + \Pr_{P_0 \sim U_0} [(P_0, f) \in \text{irr}_2] \\ & \leq \sum_{i \neq j} \Pr_{P_0 \sim U_0} [P_0(x_i) = x_j] + \sum_i \Pr_{P_0 \sim U_0} [P_0^2(x_i) = x_i] \\ & \leq \sum_{i \neq j} \Pr_{P_0 \sim U_0} [P_0(x_i) = x_j] \\ & \quad + \sum_i \sum_{y \neq x_i} \Pr_{P_0 \sim U_0} [P_0(x_i) = y \wedge P_0(y) = x_i] \\ & \leq \sum_{i \neq j} \frac{!(2^n - 1)}{!2^n} + \sum_i \sum_{y \neq x_i} \frac{!(2^n - 2)}{!2^n} \\ & \leq \lambda^2 \cdot \frac{(2^n - 1)!/e + 1}{2^n!/e - 1} + \lambda \cdot (2^n - 1) \cdot \frac{(2^n - 2)!/e + 1}{2^n!/e - 1} \\ & = \frac{\lambda^2}{2^n} \cdot \frac{1 + e/(2^n - 1)!}{1 - e/2^n!} + \frac{\lambda}{2^n} \cdot \frac{1 + e/(2^n - 2)!}{1 - e/2^n!} \\ & = \frac{2\lambda^2}{2^n} + \frac{2\lambda}{2^n} \leq \frac{4\lambda^2}{2^n} \leq \frac{16n^2}{2^n} \quad (102) \end{aligned}$$

for  $n \geq 32$ . From inequality (101) and (102), we have

$$\begin{aligned}
& \frac{1}{2} \Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \mathbf{good} \wedge (P_0, f) \in \mathbf{irr}] \\
& \leq \frac{1}{2} \sum_{\lambda < 2n} \sum_{f: |f^{-1}(1)=\lambda} \frac{16n^2}{2^n} \Pr_{F \sim \Lambda} [F = f] \\
& \leq \frac{16n^3}{2^n}.
\end{aligned} \tag{103}$$

Then, from inequality (100) and (103), we have

$$\Delta(U, U'_1) \leq \frac{8n^3}{2^n - 2n + 1} + \frac{16n^3}{2^n} + \frac{1 + e}{n!},$$

which completes the proof of Lemma 4.4.  $\square$

## H Proof of Lemma 4.5

*Proof (of Lemma 4.5).* In this proof, we define subsets  $\mathbf{good}, \mathbf{bad} \subset \text{Func}(\{0, 1\}^n, \{0, 1\}^n)$  by  $F \in \mathbf{good}$  if and only if the number of fixed points of  $F$  is less than  $2n$ , and  $\mathbf{bad} := \text{Func}(\{0, 1\}^n, \{0, 1\}^n) \setminus \mathbf{good}$ . Let  $\mathbf{irr}_1$  be the same set defined in the proof of Lemma 4.4 (see (89)), and define  $\mathbf{reg}_1$  by  $\mathbf{reg}_1 = \text{Der}(\{0, 1\}^n) \times \text{Func}(\{0, 1\}^n, \{0, 1\}^n) \setminus \mathbf{irr}_1$ . By definition of  $\Phi$  and  $\Phi'$ ,  $\Phi(P_0, f) = \Phi'(P_0, f)$  holds

for any  $(P_0, f) \in \text{reg}_1$ . Thus we have

$$\begin{aligned}
\Delta(U_1^\pm, U_2') &= \frac{1}{2} \sum_F \left| \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) = F] - \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) = F] \right| \\
&\leq \frac{1}{2} \sum_{F \in \text{good}} \left| \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) = F] - \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) = F] \right| \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}], \\
&\leq \frac{1}{2} \sum_{F \in \text{good}} \left| \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) = F \wedge (P_0, f) \in \text{reg}_1] \right. \\
&\quad \left. - \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) = F \wedge (P_0, f) \in \text{reg}_1] \right| \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}] \\
&= \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}]. \quad (104)
\end{aligned}$$

Note that the number of fixed points of  $\Phi'(P_0, f)$  is less than or equal to that of  $\Phi(P_0, f)$  by definition of  $\Phi$  and  $\Phi'$ . Thus we have

$$\begin{aligned}
\Delta(U_1^{\pm}, U_2') &\leq \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}], \\
&= \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \\
&\quad + \frac{1}{2} \left( \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] \right. \\
&\quad \left. + \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{good} \wedge \Phi(P_0, f) \in \text{bad} \wedge (P_0, f) \in \text{irr}_1] \right) \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}], \\
&\leq \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] \\
&\quad + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{bad}] + \frac{1}{2} \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi'(P_0, f) \in \text{bad}].
\end{aligned}$$

Note that the number of fixed points of  $\Phi(P_0, f)$  and  $\Phi'(P_0, f)$  are less than  $2n$  if  $|f^{-1}(1)| < n$ . Thus, each of  $\Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi(P_0, f) \in \text{bad}]$  and  $\Pr_{P_0 \sim U_0, f \sim \Lambda} [\Phi'(P_0, f) \in \text{bad}]$  is upper bounded by  $\Pr_{f \sim \Lambda} [|f^{-1}(1)| \geq n] = \Pr_{\lambda \sim D_{num}} [\lambda \geq n]$ . Therefore, from inequality (28) and (103) we have

$$\begin{aligned}
\Delta(U_1^{\pm}, U_2') &\leq \Pr_{\substack{P_0 \sim U_0, \\ f \sim \Lambda}} [\Phi(P_0, f) \in \text{good} \wedge (P_0, f) \in \text{irr}_1] + 2 \Pr_{\lambda \sim D_{num}} [\lambda \geq n] \\
&\leq \frac{32n^3}{2^n} + \frac{2(e+1)}{n!},
\end{aligned}$$

which completes the proof of Lemma 4.5.  $\square$

## I Proof of Lemma 5.2

To prove Lemma 5.2, we need the following lemma.

**Lemma I.1.** *Let  $N_P := |\{x | P(x) = x\}|$  for  $P \in \text{Perm}(\{0, 1\}^n)$ . Then, the following inequalities hold.*

1.  $|\mathbf{E}_{P \sim U}[N_P] - 1| \leq \frac{e}{n!} + \frac{e^2}{(2^n - n)!}$ ,
2.  $\mathbf{Var}_{P \sim U}[N_P] \leq 2n$ .

*Proof (of Lemma I.1).* Note that  $N_P$  follows the distribution  $D_{num}$ . Now we show the first inequality

$$|\mathbf{E}_{\lambda \sim D_{num}}[\lambda] - 1| = |\mathbf{E}_{P \sim U}[N_P] - 1| \leq \frac{e}{n!} + \frac{e^2}{(2^n - n)!}. \quad (105)$$

We have

$$\begin{aligned} \mathbf{E}_{\lambda \sim D_{num}}[\lambda] &= \sum_{\lambda=0}^{2^n} \lambda \cdot \frac{\binom{2^n}{\lambda} \cdot (2^n - \lambda)!}{2^n!} = \sum_{\lambda=1}^{2^n} \frac{1}{(\lambda - 1)!} \sum_{i=0}^{2^n - \lambda} \frac{(-1)^i}{i!} \\ &= \sum_{\lambda=0}^{2^n - 1} \frac{1}{\lambda!} \sum_{i=0}^{2^n - \lambda - 1} \frac{(-1)^i}{i!}. \end{aligned} \quad (106)$$

Hence, from Lemma A.1 we have

$$\begin{aligned} \mathbf{E}_{\lambda \sim D_{num}}[\lambda] &= \sum_{\lambda=0}^{2^n - 1} \frac{1}{\lambda!} \sum_{i=0}^{2^n - \lambda - 1} \frac{(-1)^i}{i!} \\ &\leq \sum_{\lambda=0}^{n-1} \frac{1}{\lambda!} \sum_{i=0}^{2^n - \lambda - 1} \frac{(-1)^i}{i!} + \sum_{\lambda=n}^{2^n - 1} \frac{1}{\lambda!} \sum_{i=0}^{2^n - \lambda - 1} \frac{(-1)^i}{i!} \\ &\leq \sum_{\lambda=0}^{n-1} \frac{1}{\lambda!} \left( \frac{1}{e} + \frac{e}{(2^n - \lambda)!} \right) + \sum_{\lambda=n}^{2^n - 1} \frac{1}{\lambda!} \cdot e \\ &\leq \sum_{\lambda=0}^{n-1} \frac{1}{\lambda!} \left( \frac{1}{e} + \frac{e}{(2^n - n + 1)!} \right) + \sum_{\lambda=n}^{\infty} \frac{1}{\lambda!} \cdot e \\ &\leq 1 + \frac{e^2}{(2^n - n + 1)!} + \frac{e}{n!} \\ &\leq 1 + \frac{e^2}{(2^n - n)!} + \frac{e}{n!} \end{aligned} \quad (107)$$

for  $n \geq 32$ . In addition, since  $\frac{1}{e} - \frac{1}{(2^n - \lambda)!} \geq 0$  holds for  $\lambda \leq 2^n - 5$ , we have

$$\begin{aligned}
\mathbf{E}_{\lambda \sim D_{num}}[\lambda] &= \sum_{\lambda=0}^{2^n-1} \frac{1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&= \sum_{\lambda=0}^{2^n-5} \frac{1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&\quad + \frac{1}{(2^n-1)!} \cdot 1 + \frac{1}{(2^n-2)!} \cdot 0 + \frac{1}{(2^n-3)!} \cdot \frac{1}{2} + \frac{1}{(2^n-4)!} \cdot \frac{1}{3} \\
&\geq \sum_{\lambda=0}^{2^n-5} \frac{1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&\geq \sum_{\lambda=0}^{2^n-5} \frac{1}{\lambda!} \left( \frac{1}{e} - \frac{1}{(2^n-\lambda)!} \right) \geq \sum_{\lambda=1}^n \frac{1}{\lambda!} \left( \frac{1}{e} - \frac{1}{(2^n-\lambda)!} \right) \\
&\geq \sum_{\lambda=0}^n \frac{1}{\lambda!} \left( \frac{1}{e} - \frac{1}{(2^n-n)!} \right) \\
&\geq \left( e - \frac{e}{(n+1)!} \right) \left( \frac{1}{e} - \frac{1}{(2^n-n)!} \right) \\
&\geq 1 - \frac{1}{(n+1)!} - \frac{e}{(2^n-n)!} \geq 1 - \frac{e}{n!} - \frac{e^2}{(2^n-n)!}. \tag{108}
\end{aligned}$$

From inequality (107) and (108), we obtain the first inequality (105).

Next, we show the second inequality

$$\mathbf{Var}_{P \sim U}[N_P] = \mathbf{Var}_{\lambda \sim D_{num}}[\lambda] \leq 2n. \tag{109}$$

For  $n \geq 32$ , we have

$$\begin{aligned}
\mathbf{Var}_{\lambda \sim D_{num}}[\lambda] &= \mathbf{E}_{\lambda \sim D_{num}}[\lambda^2] - (\mathbf{E}_{\lambda \sim D_{num}}[\lambda])^2 \\
&\leq \mathbf{E}_{\lambda \sim D_{num}}[\lambda^2] \\
&= \sum_{\lambda=0}^{2^n-1} \frac{\lambda+1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&= \sum_{\lambda=0}^{n-1} \frac{\lambda+1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} + \sum_{\lambda=n}^{2^n-1} \frac{\lambda+1}{\lambda!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&\leq \sum_{\lambda=0}^{n-1} \frac{n}{\lambda!} \left( \frac{1}{e} + \frac{1}{(2^n - \lambda)!} \right) + \sum_{\lambda=n}^{2^n-1} \frac{1 + \frac{1}{\lambda}}{(\lambda-1)!} \sum_{i=0}^{2^n-\lambda-1} \frac{(-1)^i}{i!} \\
&\leq \sum_{\lambda=0}^{n-1} \frac{n}{\lambda!} \left( \frac{1}{e} + \frac{1}{(2^n - n + 1)!} \right) + \sum_{\lambda=n}^{2^n-1} \frac{1 + \frac{1}{e}}{(\lambda-1)!} \cdot e \\
&\leq n \left( 1 + \frac{e^2}{(2^n - n + 1)!} \right) + \frac{e(e+1)}{(n-1)!} \\
&\leq 2n. \tag{110}
\end{aligned}$$

Therefore the second inequality (109) holds, and Lemma I.1 follows.  $\square$

Next, we prove Lemma 5.2.

*Proof (of Lemma 5.2).* First, we show that

$$\sum_{E \in \text{Ciph}(m,n)} \left| \Pr_{(e,y) \sim V} [(e,y) = (E,Y)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e,y) = (E,Y)] \right| \tag{111}$$

does not depend on  $Y \in \{0,1\}^n$ . Define  $\phi_Y : \text{Ciph}(m,n) \rightarrow \text{Ciph}(m,n)$  by  $\phi_Y(E)(z,x) := E_z(x) \oplus Y$  for each  $Y \in \{0,1\}^n$ . Then we have

$$\Pr_{(e,y) \sim V} [(e,y) = (E,Y)] = \Pr_{(e,y) \sim V} [(e,y) = (\phi_Y(E), 0)],$$

by definition of  $V$ . Therefore

$$\begin{aligned}
&\sum_{E \in \text{Ciph}(m,n)} \left| \Pr_{(e,y) \sim V} [(e,y) = (E,Y)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e,y) = (E,Y)] \right| \\
&= \sum_{E \in \text{Ciph}(m,n)} \left| \Pr_{(e,y) \sim V} [(e,y) = (\phi_Y(E), 0)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e,y) = (\phi_Y(E), 0)] \right| \\
&= \sum_{E \in \text{Ciph}(m,n)} \left| \Pr_{(e,y) \sim V} [(e,y) = (E, 0)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e,y) = (E, 0)] \right| \tag{112}
\end{aligned}$$



holds for any  $Y$ . Hence (111) does not depend on  $Y$ . Now we have

$$\begin{aligned}
& \Delta(V, (U_E, U_n)) \\
&= \frac{1}{2} \cdot \sum_Y \sum_{E \in \text{Ciph}(m, n)} \left| \Pr_{(e, y) \sim V} [(e, y) = (E, Y)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e, y) = (E, Y)] \right| \\
&= \frac{2^n}{2} \cdot \sum_{E \in \text{Ciph}(m, n)} \left| \Pr_{(e, y) \sim V} [(e, y) = (E, 0)] - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e, y) = (E, 0)] \right| \\
&= \frac{2^n}{2} \cdot \sum_{E \in \text{Ciph}(m, n)} \left| \Pr_{(e, y) \sim V} [(e, y) = (E, 0) | e = E] \right. \\
&\quad \left. - \Pr_{\substack{e \sim U_E, \\ y \sim U_n}} [(e, y) = (E, 0) | e = E] \right| \cdot \Pr_{e \sim U_E} [e = E] \\
&= \frac{2^n}{2} \cdot \sum_{E \in \text{Ciph}(m, n)} \left| \frac{N_E}{2^{n+m}} - \frac{1}{2^n} \right| \cdot \Pr_{e \sim U_E} [e = E] \\
&= \frac{1}{2} \cdot \sum_{E \in \text{Ciph}(m, n)} \left| \frac{N_E}{2^m} - 1 \right| \cdot \Pr_{e \sim U_E} [e = E], \tag{113}
\end{aligned}$$

where  $N_E := |\{(z, x) | E_z(x) = x\}|$  for  $E \in \text{Ciph}(m, n)$ .

Now, define the subsets  $\text{bad}_0, \text{bad}_1, \dots, \text{bad}_n, \text{good} \subset \text{Ciph}(m, n)$  by

$$\text{bad}_0 := \left\{ E \left| \frac{N_E}{2^m} < 1 - \frac{1}{2^{(1-\epsilon)m}} \text{ or } 1 + \frac{1}{2^{(1-\epsilon)m}} < \frac{N_E}{2^m} \leq 2 \right. \right\},$$

and

$$\text{bad}_i := \left\{ E \left| 1 + 2^{i-1} < \frac{N_E}{2^m} \leq 1 + 2^i \right. \right\}$$

for  $1 \leq i \leq n$ , and

$$\text{good} := \text{Ciph}(m, n) \setminus \left( \bigcup_i \text{bad}_i \right) = \left\{ E \left| \left| \frac{N_E}{2^m} - 1 \right| \leq \frac{1}{2^{(1-\epsilon)m}} \right. \right\},$$

where  $\epsilon$  is a parameter such that  $1/m < \epsilon < 1$ .

From Lemma I.1, now we have

$$\mathbf{Var}_{E \sim U_E} [N_E] = \mathbf{Var}_{E \sim U_E} \left[ \sum_{z \in \{0, 1\}^m} N_{E_z} \right] = 2^m \mathbf{Var}_{P \sim U} [N_P] \leq 2^{m+1} n, \tag{114}$$

since  $E_z$  are independently sampled according to the distribution  $U$  for each  $z \in \{0, 1\}^m$ . In addition, we have

$$|\mathbf{E}_{E \sim U_E} [N_E] - 2^m| = |2^m \mathbf{E}_{P \sim U} [N_P] - 2^m| \leq \frac{2^m e^2}{n!} + \frac{2^m e^2}{(2^n - n)!} < 1 \tag{115}$$

for  $n \geq 32$  and  $m \leq n^2$ . In particular, for any fixed cipher  $E$  and a parameter  $1/m < \epsilon < 1$ ,  $|N_E - 2^m| > 2^{\epsilon m}$  holds if and only if  $|N_E - \mathbf{E}_{E \sim U_E}[N_E]| > 2^{\epsilon m}$  holds. Hence, by Chebyshev's inequality we have

$$\begin{aligned}
& \Pr_{E \sim U_E} [E \in \text{bad}_0] \\
&= \Pr_{E \sim U_E} \left[ \frac{N_E}{2^m} < 1 - \frac{1}{2^{(1-\epsilon)m}} \text{ or } 1 + \frac{1}{2^{(1-\epsilon)m}} < \frac{N_E}{2^m} \leq 2 \right] \\
&\leq \Pr_{E \sim U_E} \left[ \left| \frac{N_E}{2^m} - 1 \right| > \frac{1}{2^{(1-\epsilon)m}} \right] \\
&= \Pr_{E \sim U_E} [|N_E - 2^m| > 2^{\epsilon m}] \\
&= \Pr_{E \sim U_E} [|N_E - \mathbf{E}_{E \sim U_E}[N_E]| > 2^{\epsilon m}] \\
&\leq \frac{\mathbf{Var}_{E \sim U_E}[N_E]}{(2^{\epsilon m})^2} \leq \frac{2n}{2^{(2\epsilon-1)m}}. \tag{116}
\end{aligned}$$

In addition, for an integer  $i$  such that  $1 \leq i \leq n$ , similarly we have

$$\begin{aligned}
& \Pr_{E \sim U_E} [E \in \text{bad}_i] \\
&= \Pr_{E \sim U_E} \left[ 1 + 2^{i-1} < \frac{N_E}{2^m} \leq 1 + 2^i \right] \\
&\leq \Pr_{E \sim U_E} \left[ \left| \frac{N_E}{2^m} - 1 \right| > 2^{i-1} \right] \\
&= \Pr_{E \sim U_E} [|N_E - 2^m| > 2^{m+i-1}] \\
&= \Pr_{E \sim U_E} [|N_E - \mathbf{E}_{E \sim U_E}[N_E]| > 2^{m+i-1}] \\
&\leq \frac{\mathbf{Var}_{E \sim U_E}[N_E]}{(2^{m+i-1})^2} \leq \frac{2n}{2^{m+2i-2}}, \tag{117}
\end{aligned}$$

where we used inequality (114) again.

Thus, from inequality (113), (116), and (117) we have

$$\begin{aligned}
\Delta(V, (U_E, U_n)) &\leq \frac{1}{2} \sum_{E \in \text{Ciph}(m, n)} \left| \frac{N_E}{2^m} - 1 \right| \cdot \Pr_{e \sim U_E} [e = E] \\
&= \frac{1}{2} \sum_{E \in \text{good}} \left| \frac{N_E}{2^m} - 1 \right| \cdot \Pr_{e \sim U_E} [e = E] \\
&\quad + \frac{1}{2} \sum_{E \in \text{bad}_0} \left| \frac{N_E}{2^m} - 1 \right| \cdot \Pr_{e \sim U_E} [e = E] \\
&\quad + \frac{1}{2} \sum_{1 \leq i \leq n} \sum_{E \in \text{bad}_i} \left( \frac{N_E}{2^m} - 1 \right) \Pr_{e \sim U_E} [e = E] \\
&\leq \frac{1}{2} \sum_{E \in \text{good}} \frac{1}{2^{(1-\epsilon)m}} \cdot \Pr_{e \sim U_E} [e = E] \\
&\quad + \frac{1}{2} \Pr_{E \sim U_E} [E \in \text{bad}_0] \\
&\quad + \frac{1}{2} \sum_{1 \leq i \leq n} \sum_{E \in \text{bad}_i} 2^i \Pr_{e \sim U_E} [e = E] \\
&\leq \frac{1}{2^{(1-\epsilon)m+1}} + \frac{n}{2^{(2\epsilon-1)m}} + \frac{1}{2} \sum_{1 \leq i \leq n} 2^i \Pr_{e \sim U_E} [e \in \text{bad}_i] \\
&\leq \frac{1}{2^{(1-\epsilon)m+1}} + \frac{2n}{2^{(2\epsilon-1)m+1}} + \sum_{1 \leq i \leq n} 2^i \cdot \frac{n}{2^{m+2i-2}} \\
&\leq \frac{1}{2^{(1-\epsilon)m+1}} + \frac{2n}{2^{(2\epsilon-1)m+1}} + \frac{n^2}{2^{m-2}}.
\end{aligned}$$

Setting  $\epsilon$  as  $\epsilon = 2/3$ , we obtain the desired bound, and Lemma 5.2 follows.  $\square$

## J Lower Bound of $\Delta(U, U'_1)$

This section gives the lower bound of  $\Delta(U, U'_1)$ , i.e., we show the following claim:

*Claim.*  $\Delta(U, U'_1)$  is lower bounded by  $1/(4e \cdot 2^n)$ . That is,

$$\Delta(U, U'_1) \geq \frac{1}{4e \cdot 2^n} \tag{118}$$

holds.

*Proof.* Let  $\text{Fix}_1 \subset \text{Perm}(\{0,1\}^n)$  be the set of permutations with exactly one fixed point. Then we have

$$\begin{aligned}
\Delta(U, U'_1) &= \frac{1}{2} \sum_{P \in \text{Perm}(\{0,1\}^n)} \left| \Pr_{\sigma \sim U}[\sigma = P] - \Pr_{\sigma \sim U'_1}[\sigma = P] \right| \\
&\geq \frac{1}{2} \sum_{P \in \text{Fix}_1} \left| \Pr_{\sigma \sim U}[\sigma = P] - \Pr_{\sigma \sim U'_1}[\sigma = P] \right| \\
&\geq \frac{1}{2} \left| \sum_{P \in \text{Fix}_1} \left( \Pr_{\sigma \sim U}[\sigma = P] - \Pr_{\sigma \sim U'_1}[\sigma = P] \right) \right| \\
&\geq \frac{1}{2} \left| \Pr_{\sigma \sim U}[\sigma \in \text{Fix}_1] - \Pr_{\sigma \sim U'_1}[\sigma \in \text{Fix}_1] \right|. \tag{119}
\end{aligned}$$

Remember that  $\text{irr}_2$  is defined by  $(P_0, f) \in \text{irr}_2$  if and only if  $P_0^2(x) = x$  for some  $x \in f^{-1}(1)$  (see (90)). Define  $\text{reg}_1 := \text{Der}(\{0,1\}^n) \times \text{Func}(\{0,1\}^n, \{0,1\}) \setminus \text{irr}_1$ . Since  $\Pr_{\sigma \sim U}[\sigma \in \text{Fix}_1] = \Pr_{f \sim \Lambda}[|f^{-1}(1)| = 1]$  and  $\Pr_{\sigma \sim U'_1}[\sigma \in \text{Fix}_1] = \Pr_{f \sim \Lambda, P_0 \sim U_0}[|f^{-1}(1)| = 1 \wedge (P_0, f) \in \text{reg}_2]$ , we have

$$\begin{aligned}
\Delta(U, U'_1) &\geq \frac{1}{2} \left| \Pr_{\sigma \sim U}[\sigma \in \text{Fix}_1] - \Pr_{\sigma \sim U'_1}[\sigma \in \text{Fix}_1] \right| \\
&= \frac{1}{2} \left| \Pr_{f \sim \Lambda}[|f^{-1}(1)| = 1] - \Pr_{f \sim \Lambda, P_0 \sim U_0}[|f^{-1}(1)| = 1 \wedge (P_0, f) \in \text{reg}_2] \right| \\
&= \frac{1}{2} \Pr_{f \sim \Lambda}[|f^{-1}(1)| = 1] \cdot \left| 1 - \Pr_{f \sim \Lambda, P_0 \sim U_0}[(P_0, f) \in \text{reg}_2 | |f^{-1}(1)| = 1] \right| \\
&= \frac{1}{2} \Pr_{f \sim \Lambda}[|f^{-1}(1)| = 1] \cdot \Pr_{f \sim \Lambda, P_0 \sim U_0}[(P_0, f) \in \text{irr}_2 | |f^{-1}(1)| = 1] \tag{120}
\end{aligned}$$

Now we have

$$\Pr_{f \sim \Lambda}[|f^{-1}(1)| = 1] = \frac{\binom{2^n}{1} \cdot (2^n - 1)!}{(2^n)!} = \frac{(2^n - 1)!}{(2^n - 1)!} = \sum_{i=1}^{2^n-1} \frac{(-1)^i}{i!} \geq \frac{1}{e} - \frac{e}{2^n!}, \tag{121}$$

and

$$\begin{aligned}
\Pr_{f \sim \Lambda}[(P_0, f) \in \text{irr}_2 | |f^{-1}(1)| = 1] &= \Pr_{f \sim \Lambda}[P_0^2(f^{-1}(1)) = f^{-1}(1) | |f^{-1}(1)| = 1] \\
&= \frac{(2^n - 1)! \cdot (2^n - 2)!}{!(2^n)} \geq \frac{(2^n - 1) \cdot (2^n - 2)! / e}{(2^n)! / e + 1} \\
&= \frac{(2^n - 1)!}{(2^n)! + e} \geq \frac{1}{2} \frac{(2^n - 1)!}{(2^n)!} = \frac{1}{2 \cdot 2^n} \tag{122}
\end{aligned}$$

Therefore

$$\Delta(U, U'_1) \geq \left( \frac{1}{e} - \frac{e}{2^n!} \right) \cdot \frac{1}{2 \cdot 2^n!} \geq \frac{1}{4e \cdot 2^n} \tag{123}$$

holds for  $n \geq 32$ , which completes the proof of the claim.  $\square$