Quantum Collision-Resistance of Non-uniformly Distributed Functions: Upper and Lower Bounds

Ehsan Ebrahimi and Dominique Unruh

University of Tartu, Estonia

{ehsan.ebrahimi.targhi, unruh}@ut.ee

June 15, 2017

Abstract

We study the quantum query complexity of finding a collision for a function f whose outputs are chosen according to a non-uniform distribution $\mathcal{D}$. We derive some upper bounds and lower bounds depending on the min-entropy and the collision-entropy of $\mathcal{D}$. In particular, we improve the previous lower bound in [TTU16] from $\Omega(2^{k/9})$ to $\Omega(2^{k/5})$ where k is the min-entropy of $\mathcal{D}$.

Keywords: Quantum, Collision, Non-uniform distribution, Query complexity.

1 Introduction

We study the quantum query complexity of finding a collision for a function f whose outputs are chosen according to a non-uniform distribution $\mathcal{D}$. We derive some upper bounds and lower bounds depending on the min-entropy and the collision-entropy of $\mathcal{D}$. Recall that a collision for function f consists of two distinct inputs x_1 and x_2 such that $f(x_1) = f(x_2)$. By Lemma A.9 in [KL07], $\Omega(2^{k/2})$ classical queries are necessary to find a collision with constant probability where k is the min-entropy of $\mathcal{D}$, and Theorem 3 in [Wie05] shows that $O(2^{k/2})$ classical queries are sufficient to find a collision with constant probability where k is the collision-entropy of $\mathcal{D}$. However, in the quantum query model this number of queries may be too high since one quantum query may contain the whole input-output values of the function. The quantum collision problem for a non-uniform distribution has been studied in [TTU16]. They prove an $\Omega(2^{k/9})$ lower bound for the problem where k is the min-entropy of $\mathcal{D}$. We improve their bound to $\Omega(2^{k/5})$ in this paper, and additionally prove upper and lower bounds for a number of cases (see below).

One motivation for studying the quantum collision problem for a non-uniform distribution is the interest in proving the security of classical cryptographic schemes against quantum adversaries. Hash functions are crucial cryptographic primitives that are used to construct many encryption schemes and cryptographic schemes and they may be distributed non-uniformly. Even though in the random oracle model [BR93] they are modeled as random functions, they are often used as a composition with other functions. Therefore the output of the combination of a function f and a random function H may not be distributed uniformly at random and finding a collision for this non-uniformly distributed $f \circ H$ may break the security of the scheme. For example the well-known Fujisaki-Okamoto construction [FO99] uses a random function H to produce the randomness for an encryption scheme f. The security relies on the fact that the

Lower bound			Upper bound		
Quantifier	$H_\infty(\mathcal{D})$	$H_2(\mathcal{D})$	Quantifier	$H_\infty(\mathcal{D})$	$H_2(\mathcal{D})$
$\forall\,\mathcal{A}\,\forall\,\mathcal{D}$	$\Omega(2^{H_\infty(\mathcal{D})/5})$	$\Omega(2^{H_2(\mathcal{D})/9})$	$\exists\,\mathcal{A}\,\exists\,\mathcal{D}$	$O(2^{H_\infty(\mathcal{D})/3})$	$O(2^{H_2(\mathcal{D})/4})$
$\forall\,\mathcal{A}\,\exists\,\mathcal{D}$	$\Omega(2^{\frac{H_\infty(\mathcal{D})}{2}})$	$\Omega(2^{\frac{H_2(\mathcal{D})}{3}})$	$\exists\,\mathcal{A}\,\forall\,\mathcal{D}$	$O(2^{\frac{2H_\infty(\mathcal{D})}{3}})$	$O(2^{\frac{H_2(\mathcal{D})}{3}})$
$\exists\,\mathcal{D}\,\forall\,\mathcal{A}$	$\Omega(2^{\frac{H_\infty(\mathcal{D})}{2}})$	$\Omega(2^{\frac{H_2(\mathcal{D})}{3}})$	$\forall\,\mathcal{D}\,\exists\,\mathcal{A}$	$O(2^{\frac{H_\infty(\mathcal{D})}{2}})$	$O(2^{\frac{H_2(\mathcal{D})}{3}})$

Figure 1: Summary of the bounds achieved in this paper. The columns marked $H_\infty(\mathcal{D})$, $H_2(\mathcal{D})$ give lower/upper bounds on the number of queries needed for finding a collision in terms of the min-entropy and the collision-entropy, respectively. The "quantifier" column indicates for what quantification of collision-finding algorithm $\mathcal{A}$ and distribution $\mathcal{D}$ the respective bound is achieved. For example, a lower bound $\Omega(B)$ with quantifiers $\forall\,\mathcal{A}\,\exists\,\mathcal{D}$ means that for any adversary $\mathcal{A}$, there exists a distribution $\mathcal{D}$ such that $\mathcal{A}$ needs at least $\Omega(B)$ queries to find a collision.

adversary can not find two inputs of the random function that lead to the same ciphertext. This is roughly equivalent to saying that $f \circ H$ is collision-resistant. In fact, the quantum collision-resistance of a non-uniform distributed function has been used for analyzing a variant of the Fujisaki-Okamoto construction in the quantum setting [TU16].

The table in Figure 1 summarizes our results. We give twelve bounds altogether, depending on whether it is an upper or a lower bound, whether it is expressed in terms of the min-entropy or the collision-entropy, and depending on the order of the quantification of the collision-finding algorithm and the non-uniform distribution.

Why twelve bounds? Our work gives twelve different bounds (see Figure 1). Why do we need this number of bounds? First, when considering non-uniform distributions, there are a number of entropy measures that can quantify the deviation from the uniform distribution. At least in a cryptographic context, the min-entropy H_∞ and the collision-entropy H_2 are often of particular interest. Thus we state our bounds in terms of these two entropy measures. (Other entropy measures may be of interest in this context, too. We leave this as possible future work.) Then, there are three main applications that we might be interested in:

- *A cryptographic bound.* For example, in the context of [TU16], we have a cryptosystem which can be broken iff the adversary can find a collision in a function with non-uniform distribution $\mathcal{D}$ (see above). All we know about $\mathcal{D}$ is a lower bound on its entropy. Thus we need a bound of the following form: "for every algorithm $\mathcal{A}$, and every distribution $\mathcal{D}$ of entropy $\geq k$, finding a collision takes at least $\Omega(\dots)$ queries." Such lower bounds are given in Figure 1 in the row marked $\forall\,\mathcal{A}\,\forall\,\mathcal{D}$.
- *An algorithmic upper bound.* On the other hand, our analysis might be motivated by algorithmic considerations. We wish to implement an algorithm that finds collisions in functions with non-uniform distribution. That is, we look for a statement of the kind: "for every distribution $\mathcal{D}$ of entropy $\leq k$, there is an algorithm $\mathcal{A}$ that finds a collision in at most $O(\dots)$ queries." Such upper bounds are given in Figure 1 in the row marked $\forall\,\mathcal{D}\,\exists A$.
- *An algorithmic upper bound for universal algorithms.* Still in the algorithmic setting, we might not be content with an algorithm that depends on the distribution at hand. We may search for an algorithm that works without needing to know what the distribution is. That is, we look for a universal algorithm. (This might also be relevant if the distribution is known, but too hard to describe.) In this case, we look for a statement of the form: "there is an single algorithm $\mathcal{A}$ that finds a collision in at most $O(\dots)$ queries for any distribution $\mathcal{D}$ of entropy $\leq k$ (without needing to know from which distribution we draw)." Such upper bounds are given in Figure 1 in the row marked $\exists A\forall\,\mathcal{D}$.

And finally, in each of these three cases we may ask whether the bound is tight (and if not, how loose it is). For example, to see whether the $\forall\,\mathcal{A}\forall D$ lower bound is tight, we need to find an $\exists\,\mathcal{A}\exists D$ upper bound. (Our bounds are not matching yet, we leave this as future work).

Similarly, to understand whether the $\exists\mathcal{A}\forall D$ upper bound is tight, we need a matching $\forall\mathcal{A}\exists D$ lower bound. (We achieve a tight bound for the collision-entropy case here.) And finally, to show that the $\forall\mathcal{D}\exists\mathcal{A}$ upper bound is tight, we need an $\exists\mathcal{D}\forall\mathcal{A}$ lower bound. (We have matching bounds in this case.)

Related works. The quantum collision problem has been studied in various previous works. In the following, we mention the existing results on the number of queries that are necessary to find a collision. In [TTU16], they prove an $\Omega(2^{k/9})$ lower bound for the quantum query complexity of the function f whose output are chosen according to a distribution with the min-entropy k. An $\Omega\big((N/r)^{1/3}\big)$ lower bound for an r-to-one function f is given by Aaronson and Shi [AS04] where N is the domain size and $r|N$. Yuen [Yue14] proves an $\Omega(N^{1/5}/\text{polylog}N)$ lower bound for the quantum collision problem for a random function f with same domain and co-domain of size N. They reduce the distinguishing between a random function and a random permutation problem to the distinguishing between a function with r-to-one part and a function without r-to-one part. Their proof combines the r-to-one lower bound from [AS04] and the quantum adversary method [Amb00]. Zhandry [Zha15] improves Yuen's bound to the $\Omega(N^{1/3})$ and also removes the same size domain and co-domain constraint (N is the size of co-domain here.). They use the existing result from [Zha12] to prove their bound.

The sufficient number of quantum queries to find a collision is given in the following works. A quantum algorithm that requires $O\big((N/r)^{1/3}\big)$ quantum queries and finds a collision for any r-to-one function f is given by Brassard, Høyer and Tapp [BHT97]. Ambainis [Amb07] gives a quantum algorithm that requires $O(N^{2/3})$ queries to find two equal elements among N given elements and therefore it is an algorithm for finding a collision in an arbitrary function f with the domain of size N given the promise that f has at least one collision. Yuen [Yue14] shows that the collision-finding algorithm from [BHT97] is able to produce a collision for a random function with the same domain and co-domain using $O(N^{1/3})$ queries. Zhandry shows that $O(M^{1/3})$ queries are sufficient to find a collision for a random function $f : [N] \to [M]$ where $N = \Omega(M^{1/2})$. They use Ambainis's element distinctness algorithm [Amb07] as a black box in their proof.

Organization of the paper. In Section 2, we present some definitions and known results that are needed in this paper. In Section 3, we present upper bounds for the collision problem and cover both the collision-entropy and the min-entropy. Section 4 is devoted to lower bounds for the collision problem and we divide the section into two subsections for the case of the collision-entropy, (Subsection 4.1), and the min-entropy, (Subsection 4.2).

2 Preliminaries

In this section, we present some definitions and existing results that are needed in this paper. We show the restriction of the function f to the set S by notation $f|_S$. We represent the set $\{1,\ldots,m\}$ by $[m]$. Notation $x \xleftarrow{\$} X$ shows that x is chosen uniformly at random from set X. If $\mathcal{D}$ is a distribution over X, then notation $x \leftarrow \mathcal{D}$ shows that x is chosen at random according to the distribution $\mathcal{D}$. $\Pr[P : G]$ is the probability that the predicate P holds true where free variables in P are assigned according to the program in G. We say that the quantum algorithm $\mathcal{A}$ has quantum access to the oracle $O : \{0,1\}^{n_0} \to \{0,1\}^{n_1}$, denoted by $\mathcal{A}^O$, where $\mathcal{A}$ can submit queries in superposition and the oracle O answers to the queries by a unitary transformation that maps $|x,\ y\rangle$ to $|x,\ y \oplus O(x)\rangle$.

Definition 1. *Let $\mathcal{D}$ be a distribution on a set X. The min-entropy and collision-entropy of the*

distribution $\mathcal{D}$ is defined as the following, respectively.

$$H_\infty(\mathcal{D}) = -\log \max_{x \in X} \mathcal{D}(x), \quad H_2(\mathcal{D}) = -\log \sum_{x \in X} \mathcal{D}(x)^2.$$

Lemma 1 (Theorem 7 [Zha15])**.** *Let $h : X \to Y$ be a random function. Then any quantum algorithm making q queries to h outputs a collision for h with probability at most $\frac{C(q+2)^3}{|Y|}$ where C is a universal constant.*

Definition 2 (Universal Hash Function [CW79])**.** *A family of functions $H = \{h : \{0,1\}^n \to \{0,1\}^m\}$ is called a family of universal hash function if for all distinct $x, y \in \{0,1\}^n$:*

$$\Pr[h(x) = h(y) : h \xleftarrow{\$} H] \leq 1/2^m.$$

Definition 3. *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be distributions on a set X. The statistical distance between $\mathcal{D}_1$ and $\mathcal{D}_2$ is*

$$\mathrm{SD}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2} \sum_{x \in X} \big| \Pr[\mathcal{D}_1(x)] - \Pr[\mathcal{D}_2(x)] \big|.$$

Lemma 2 (Leftover Hash Lemma [HILL93])**.** *Let $\mathcal{D}$ be a distribution with collision-entropy k and e be a positive integer. Let $h : \{0,1\}^m \times \{0,1\}^n \to \{0,1\}^{k-2e}$ be a universal hash function. Then,*

$$\mathrm{SD}\Big(\big(h(y,x), y\big), \big(z, y\big)\Big) \leq 2^{-e-1}$$

where $x \xleftarrow{\mathcal{D}} \{0,1\}^n, y \xleftarrow{\$} \{0,1\}^m$ and $z \xleftarrow{\$} \{0,1\}^{k-2e}$.

Lemma 3 (Proof of Theorem 1 in [HRS16])**.** *Let $F : X \to \{0,1\}$ be a function such that $F(x) := 1$ with probability γ, and $F(x) := 0$ otherwise. Then for any oracle algorithm $\mathcal{A}$ making q queries,*

$$|\Pr[b = 1 : b \leftarrow \mathcal{A}^F] - \Pr[b = 1 : b \leftarrow \mathcal{A}^N]| \leq 8q^2\gamma,$$

where N is the zero function on X.

Lemma 4 ([Zha12])**.** *Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be efficiently sampleable distributions over some set Y, and let X be some other set. For $i = 1, 2$, let $\mathcal{D}_i^X$ be the distributions of functions F_i from X to Y where for each $x \in X$, $F_i(x)$ is chosen at random according to the distribution $\mathcal{D}_i$. Then if A be a quantum algorithm that makes q queries and distinguish $\mathcal{D}_1^X$ from $\mathcal{D}_2^X$ with non-negligible probability ϵ, we can construct a quantum algorithm $\mathcal{B}$ that distinguishes samples from $\mathcal{D}_1$ and $\mathcal{D}_2$ with probability at least $\frac{3\epsilon^2}{64\pi^2 q^3}$.*

3 Upper bounds

We divide this section to three following subsections according to the possible quantifiers. We denote the set of natural numbers by $\mathbb{N}$.

3.1 Quantifier order $\exists\ \mathcal{A}\ \exists\ \mathcal{D}$

Theorem 1. *There exist a quantum algorithm $\mathcal{A}$ and a distribution $\mathcal{D}$ with the min-entropy k such that $\mathcal{A}$ returns a collision for the function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ with constant probability using $O(2^{k/3})$ queries.*

Proof. Theorem 6 in [Zha15] presents a quantum algorithm that outputs a collision with constant probability using $O(2^{k/3})$ queries where $k := H_\infty(\mathcal{D})$ and $\mathcal{D}$ is an uniform distribution. By repeating their algorithm on distinct subset of $\mathbb{N}$, we can amplify the success probability arbitrary close to 1. □

We improve the bound above to $O(2^{k/4})$ for the collision-entropy in the following theorem.

Theorem 2. *There exists a quantum algorithm $\mathcal{A}$ and a distribution $\mathcal{D}$ with $H_2(\mathcal{D}) \geq k$ such that $\mathcal{A}$ returns a collision for the function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ with constant probability using $O(2^{k/4})$ queries.*

Proof. Let n be some integer such that $\frac{2^n-1}{(1-\gamma)^2} \geq 2^{k+1}$ for $\gamma := 1/2^{(k+1)/2}$. Let $\mathcal{D}$ be a distribution over $\{0,1\}^n$ such that $\mathcal{D}(0) = \gamma$ and $\mathcal{D}(y) = \frac{1-\mathcal{D}(0)}{2^n-1}$ for any non-zero $y \in \{0,1\}^n$. Simple calculation shows $H_2(\mathcal{D}) \geq k$. We define the boolean function $g : N_1 \to \{0,1\}$ as:

$$g(n) := \begin{cases} 1 & \text{if } f(n) = 0 \\ 0 & \text{otherwise} \end{cases},$$

where N_1 is an arbitrary subset of $\mathbb{N}$ and of size $\lceil 2/\gamma \rceil$ and $f \leftarrow \mathcal{D}^{\mathbb{N}}$. Note that every query to g can be implemented by 2 queries to f. Let $\mathcal{X} := \sum_{n \in N_1} g(n)$. Therefore $\mu := \mathbb{E}(\mathcal{X}) \geq 2$. We use Chernoff inequality, Theorem 4.5 in [MU05], that states for any $0 < \delta < 1$:

$$\Pr[\mathcal{X} > (1-\delta)\mu] > 1 - e^{-\frac{\delta^2 \mu}{2}}.$$

By choosing $\delta = 1/2$, there exists at least one pre-image for 1 with probability at least $1 - e^{-1/4}$. Hence, Grover's algorithm [Gro96, BBHT98] returns input $\hat{n}$ such that $g(\hat{n}) = 1$ using $O(\sqrt{|N_1|})$ queries with constant probability C. We choose an arbitrary subset $N_2 \subset \mathbb{N} \setminus N_1$ and of size $\lceil 2/\gamma \rceil$ and define the function $g' : N_2 \to \{0,1\}$ similar to g and invoke Grover's algorithm to obtain n' such that $g'(n') = 1$. Note that $(\hat{n}, n')$ is a collision for f and the success probability of the algorithm is $C^2(1 - e^{-1/4})^2$. By repeating this procedure on distinct subset of the domain of f, we can amplify the success probability arbitrary close to 1. □

3.2 Quantifier order $\exists\ \mathcal{A}\ \forall\ \mathcal{D}$

Theorem 3. *There exists a quantum algorithm $\mathcal{A}$ such that for any distribution $\mathcal{D}$ with $H_2(\mathcal{D}) \leq k$ outputs a collision for the function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ with constant probability using $O(2^{k/3})$ queries.*

Proof. The proof follows by the reduction technique used in [AS04, Zha15] and using Ambainis's algorithm [Amb07] for the element distinctness as a black box. Let S be a random subset of $\mathbb{N}$ of size $2^{k/2} + 1$. By Theorem 3 in [Wie05], $f' := f|_S$ has at least one collision with probability at least $1 - 2/e$. Now, invoking Ambainis's algorithm [Amb07] for f' returns a collision with bounded error. The query complexity of Ambainis's algorithm is $O(|S|^{2/3}) = O(2^{k/3})$. By repeating this procedure on distinct subset of the domain of f, we can amplify the success probability arbitrary close to 1. □

Theorem 4. *There exists a quantum algorithm $\mathcal{A}$ such that for any distribution $\mathcal{D}$ with $H_\infty(\mathcal{D}) \leq k$ outputs a collision for function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ using $O(2^{2k/3})$ queries and with constant probability.*

Proof. The proof follows by the theorem above and the inequality $H_2(\mathcal{D}) \leq 2H_\infty(\mathcal{D})$. □

3.3 Quantifier order $\forall\ \mathcal{D}\ \exists\ \mathcal{A}$

Theorem 5. *For any distribution $\mathcal{D}$ with $H_\infty(\mathcal{D}) \leq k$, there exists a quantum algorithm $\mathcal{A}$ that finds a collision for the function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ with constant probability using $O(2^{k/2})$ queries.*

Proof. Let $Y := \sup(D)$. Fix some $y \in Y$ such that $\mathcal{D}(y) \geq 1/2^k$. We define the boolean function $g : N_1 \to Y$ as follows:

$$g(n) := \begin{cases} 1 & \text{if } f(n) = y \\ 0 & \text{otherwise} \end{cases},$$

where N_1 is a random subset of $\mathbb{N}$ and of size 2^{k+1}. Note that every query to g can be implemented by 2 queries to f. Using Chernoff bound and similar to the Theorem 2, we can show that there exists at least one pre-image for 1 with probability at least $1-e^{-1/4}$. Hence, Grover's algorithm [Gro96, BBHT98] returns input $\hat{n}$ such that $g(\hat{n})=1$ using $O(\sqrt{|N_1|})$ queries with constant probability. We choose a random subset $N_2 \subset \mathbb{N} \setminus N_1$ and of size 2^{k+1} and define the function $g' : N_2 \to \{0,1\}$ similar to g and invoke Grover's algorithm to obtain n' such that $g'(n')=1$. Note that $(\hat{n}, n')$ is a collision for f. By repeating this procedure on distinct subset of the domain of f, we can amplify the success probability arbitrary close to 1.

□

Theorem 6. *For any distribution $\mathcal{D}$ with $H_2(\mathcal{D}) \leq k$, there exists a quantum algorithm $\mathcal{A}$ that outputs a collision for the function $f \leftarrow \mathcal{D}^{\mathbb{N}}$ with constant probability using $O(2^{k/3})$ queries.*

Proof. By Theorem 3. □

4 Lower bounds

In this section, we prove the lower bounds on number of queries needed to output a collision. We present them in two subsections based on the collision-entropy and min-entropy.

4.1 Collision-entropy

4.1.1 Quantifier orders $\exists\ \mathcal{D}\ \forall\ \mathcal{A}$ and $\forall\ \mathcal{A}\ \exists\ \mathcal{D}$.

Theorem 7. *There exist a distribution $\mathcal{D}$ with $H_2(\mathcal{D}) = k$ such that for any quantum algorithm $\mathcal{A}$ making q queries to $f \leftarrow \mathcal{D}^X$ returns a collision for f with probability at most $\frac{C(q+2)^3}{2^k}$ where C is a universal constant. That is,*

$$\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x, x') \leftarrow \mathcal{A}^f] \leq \frac{C(q+2)^3}{2^k}.$$

Proof. If we consider the uniform distribution $\mathcal{D}$ over $\{0,1\}^k$ then Lemma 1 shows the result for the reason that $H_2(\mathcal{D}) = k$. □

Theorem 8. *For any quantum algorithm $\mathcal{A}$, there exist a distribution $\mathcal{D}$ with $H_2(\mathcal{D}) = k$ such that $\mathcal{A}$ making q queries to $f \leftarrow \mathcal{D}^X$ returns a collision for f with probability at most $\frac{C(q+2)^3}{2^k}$ where C is a universal constant. That is,*

$$\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x, x') \leftarrow \mathcal{A}^f] \leq \frac{C(q+2)^3}{2^k}.$$

Proof. It follows immediately from the above theorem. □

4.1.2 Quantifier order $\forall\ \mathcal{A}\ \forall\ \mathcal{D}$

In the following, we prove that for any quantum algorithm $\mathcal{A}$ and for any distribution $\mathcal{D}$, $\Omega(2^{k/9})$ quantum queries are needed to output a collision.

Lemma 5. *Let $\mathcal{D}$ be a distribution over $\{0,1\}^{n_1}$. Let $f : \{0,1\}^{n_1} \to \{0,1\}^{n_2}$ be a public function and $X = \{0,1\}^{n_0}$. If $\mathcal{A}$ is a quantum algorithm that makes q queries to the function O drawn from distribution $\mathcal{D}^X$ and finds a collision for $f \circ O$ with some probability, then there exists a quantum algorithm $\mathcal{B}$ that makes $2q$ queries to $f \circ O$ and outputs a collision for $f \circ O$ with the same probability.*

Proof. Let $S_y = f^{-1}(\{y\})$ for $y \in \operatorname{Im} f$. Let $Y_0 := \{y \in \operatorname{Im} f : \mathcal{D} \text{ is zero on } S_y\}$. For any $y \in \operatorname{Im} f \setminus Y_0$, we define the distribution $\mathcal{D}_y$ over S_y as

$$\mathcal{D}_y(z) := \frac{\mathcal{D}(z)}{\sum_{z \in S_y} \mathcal{D}(z)}.$$

Let $\mathcal{D}'$ be the distribution of functions $F : \{0,1\}^{n_0} \times (\operatorname{Im} f \setminus Y_0) \to \{0,1\}^{n_1}$ where for any $x \in \{0,1\}^{n_0}$ and $y \in \operatorname{Im} f \setminus Y_0$, $F(x,y)$ is chosen at random in S_y according to the distribution $\mathcal{D}_y$. Let $(F \odot g)(x) := F\big(x, g(x)\big)$. We show that output of $F \odot (f \circ O)$ are chosen independently according to the distribution $\mathcal{D}$ when F is chosen according to distribution $\mathcal{D}'$ and this shows that O and $F \odot (f \circ O)$ have the same distribution. For every $x \in \{0,1\}^{n_0}$ and $z \in \{0,1\}^{n_1} \setminus \cup_{y \in Y_0} S_y$:

$$\begin{aligned}
&\Pr[\big(F \odot (f \circ O)\big)(x) = z : \ O \leftarrow \mathcal{D}^X, \ F \leftarrow \mathcal{D}'] \\
&\quad = \Pr[F\Big(x, \ f\big(O(x)\big)\Big) = z : \ O \leftarrow \mathcal{D}^X, \ F \leftarrow \mathcal{D}'] \\
&\quad = \Pr[F\big(x, \ f(z')\big) = z : \ z' \leftarrow \mathcal{D}, \ F \leftarrow \mathcal{D}'] \\
&\quad = \Pr[z'' = z : \ z' \leftarrow \mathcal{D}, \ z'' \leftarrow \mathcal{D}_{f(z')}] \\
&\quad \overset{(*)}{=} \Pr[z'' = z \wedge z' \in S_{f(z)} : \ z' \leftarrow \mathcal{D}, \ z'' \leftarrow \mathcal{D}_{f(z')}] \\
&\quad \overset{(**)}{=} \Pr[z' \in S_{f(z)} : \ z' \leftarrow \mathcal{D}] \Pr[z'' = z : \ z'' \leftarrow \mathcal{D}_{f(z)}] \\
&\quad = \Big(\sum_{z' \in S_{f(z)}} \mathcal{D}(z')\Big) \cdot \frac{\mathcal{D}(z)}{\sum_{z' \in S_{f(z)}} \mathcal{D}(z')} = \mathcal{D}(z),
\end{aligned}$$

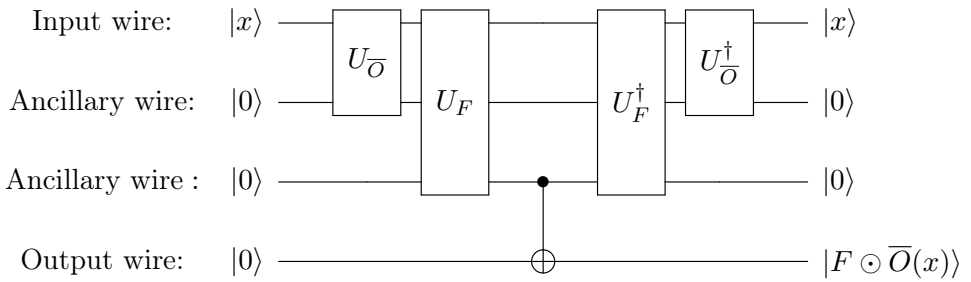
where $(*)$ holds for the reason that if $z'' = z$ be true, then $f(z) = f(z')$ and z' will be in the set $S_{f(z)}$, and $(**)$ holds because z' and z'' are independent. Note that when $z \in \cup_{y \in Y_0} S_y$, then

$$\Pr[\big(F \odot (f \circ O)\big)(x) = z : \ O \leftarrow \mathcal{D}^X, \ F \leftarrow \mathcal{D}'] = 0 = \mathcal{D}(z).$$

Let $\bar{O} := f \circ O$. As a result:

$$\begin{aligned}
&\Pr[\bar{O}(x) = \bar{O}(x') \wedge x \neq x' : O \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^O] \\
&\quad = \Pr[f \circ (F \odot \bar{O})(x) = f \circ (F \odot \bar{O})(x') \wedge x \neq x' : O \leftarrow \mathcal{D}^X, F \leftarrow \mathcal{D}', (x,x') \leftarrow \mathcal{A}^{F \odot \bar{O}}].
\end{aligned}$$

Now, we construct a quantum algorithm $\mathcal{B}$. The algorithm $\mathcal{B}$ with oracle access to $\bar{O}$ runs $\mathcal{A}^{F \odot \bar{O}}$ with $F \leftarrow \mathcal{D}'$. The way that the quantum algorithm $\mathcal{B}$ handles quantum queries is shown in the following circuit.



The algorithm $\mathcal{B}$ returns the output of $\mathcal{A}$, say (x,x'), as a collision for $\bar{O}$ after $2q$ queries to oracle $\bar{O}$. Note that from $f \circ (F \odot \bar{O})(x) = f \circ (F \odot \bar{O})(x')$, we can deduce $\bar{O}(x) = \bar{O}(x')$ because $F \odot \bar{O}(x) \in f^{-1}(\{\bar{O}(x)\})$ and $F \odot \bar{O}(x') \in f^{-1}(\{\bar{O}(x')\})$. Therefore, we prove the existence of the quantum algorithm $\mathcal{B}$ stated in the lemma. □

Theorem 9. *Let $\mathcal{D}$ be a distribution with $H_2(\mathcal{D}) \geq k$ over set $\{0,1\}^{n_1}$. Let f be a function drawn from distribution $\mathcal{D}^X$. Then any quantum algorithm $\mathcal{A}$ making q queries to f returns a collision for f with probability at most $\frac{C'(q+2)^{9/5}}{2^{k/5}}$ where C' is a universal constant. That is,*

$$\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^f] \leq \frac{C'(q+2)^{9/5}}{2^{k/5}}.$$

Let $h : \{0,1\}^m \times \{0,1\}^{n_1} \to \{0,1\}^{k-2e}$ be a universal hash function. Lemma 2 implies that:

$$\mathrm{SD}\left((h_y(d), y), (z, y)\right) \leq 2^{-e-1}, \tag{1}$$

where $h_y(d) := h(y,d)$, $d \leftarrow \mathcal{D}$, $y \xleftarrow{\$} \{0,1\}^m$ and $z \xleftarrow{\$} \{0,1\}^{k-2e}$. Let $\mathcal{U}$ be the uniform distribution over $\{0,1\}^{k-2e}$. The upper bound can be concluded by the following steps:

$$\begin{aligned}
&\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^f] \\
&\overset{(i)}{\leq} \Pr[h_y \circ f(x) = h_y \circ f(x') \wedge x \neq x' : y \xleftarrow{\$} \{0,1\}^m, f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^f] \\
&= \sum_{y \in \{0,1\}^m} \frac{1}{2^m} \Pr[h_y \circ f(x) = h_y \circ f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^f] \\
&\overset{(ii)}{=} \sum_{y \in \{0,1\}^m} \frac{1}{2^m} \Pr[h_y \circ f(x) = h_y \circ f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{B}^{h_y \circ f}(y)] \\
&\overset{(iii)}{\leq} \Pr[f^*(x) = f^*(x') \wedge x \neq x' : f^* \leftarrow \mathcal{U}^X, (x,x') \leftarrow \mathcal{B}^{f^*}] + \sqrt{64\pi^2 q^3 2^{-e-1}/3} \\
&\overset{(iv)}{\leq} \frac{C(q+2)^3}{(2^{k-2e})} + \sqrt{\frac{64\pi^2 q^3}{3(2^{e+1})}}
\end{aligned}$$

where

(i) follows from the fact that collisions for f will also be collisions for $h_y \circ f$;
(ii) follows from Lemma 5 that implies the existence of quantum algorithm $\mathcal{B}$;
(iii) can be seen as follows: Let

$$\epsilon_y := \Big| \Pr[h_y \circ f(x) = h_y \circ f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{B}^{h_y \circ f}(y)] - \Pr[f^*(x) = f^*(x') \wedge x \neq x' : f^* \leftarrow \mathcal{U}^X, (x,x') \leftarrow \mathcal{B}^{f^*}] \Big|.$$

Let for any $y \in \{0,1\}^m$, $\mathcal{D}_{1y}$ be a distribution over $\{0,1\}^{k-2e}$ where for any $\ell \in \{0,1\}^{k-2e}$, $\mathcal{D}_{1y}(\ell) = \Pr[h_y(d) = \ell : d \leftarrow \mathcal{D}]$. Using Lemma 4, we can conclude that there exists an adversary $\mathcal{A}$ such that

$$\mathrm{SD}(\mathcal{D}_{1y}, \mathcal{U}) \geq |\Pr[\mathcal{A}(x) = 1 : x \leftarrow \mathcal{D}_{1y}] - \Pr[\mathcal{A}(x) = 1 : x \leftarrow \mathcal{U}]| \geq \frac{3\epsilon_y^2}{64\pi^2 q^3},$$

and Equation 1 implies that

$$\sum_{y \in \{0,1\}^m} \frac{1}{2^m} \mathrm{SD}(\mathcal{D}_{1y}, \mathcal{U}) \leq 2^{-e-1}.$$

Therefore, we can deduce

$$\frac{1}{2^m} \sum_{y \in \{0,1\}^m} \frac{3\epsilon_y^2}{64\pi^2 q^3} \leq 2^{-e-1}.$$

Using Jensen's inequality [Jen06], we can obtain $\sum_{y \in \{0,1\}^m} \frac{1}{2^m} \epsilon_y \leq \sqrt{64\pi^2 q^3 2^{-e-1}/3}$.

Let

$$\epsilon := \Big| \sum_{y\in\{0,1\}^m} \frac{1}{2^m} \Pr[h_y \circ f(x) = h_y \circ f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{B}^{h_y\circ f}(y)] - \Pr[f^*(x) = f^*(x') \wedge x \neq x' : f^* \leftarrow \mathcal{U}^X, (x,x') \leftarrow \mathcal{B}^{f^*}]\Big|.$$

By triangle inequality, $\epsilon \leq \sum_{y\in\{0,1\}^m} \frac{1}{2^m}\epsilon_y \leq \sqrt{64\pi^2 q^3 2^{-e-1}/3}$.
(iv) follows from applying Lemma 1 to the random function f^*.
So far, we have the upper bound

$$\eta_e := \frac{2^{2e}\mu}{2^k} + \frac{\nu}{2^{e/2}}, \qquad \text{where } \mu := C(q+2)^3 \text{ and } \nu := \frac{8\pi q^{3/2}}{\sqrt{6}}.$$

It is minimized by choosing

$$e = \frac{2}{5}k + \frac{2}{5}\log\frac{\nu}{4\mu}.$$

Substituting this value of e gives us

$$\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X, (x,x') \leftarrow \mathcal{A}^f] \leq \frac{2^{2/5}\mu^{1/5}\nu^{4/5}}{2^{k/5}} \leq \frac{C'(q+2)^{9/5}}{2^{k/5}}.$$

.

4.2 Min-entropy

4.2.1 Quantifier orders $\exists\ \mathcal{D}\ \forall\ \mathcal{A}$ and $\forall\ \mathcal{A}\ \exists\ \mathcal{D}$.

Theorem 10. *For any k, there exists a distribution $\mathcal{D}$ with min-entropy k such that for every adversary $\mathcal{A}$ making q queries,*

$$\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f] \leq O\big(\max\{q^2/2^k, q^3/2^{3k/2}\}\big).$$

Proof. Let $n \geq 3k/2$. Let $\mathcal{D}$ be a distribution over $\{0,1\}^n \cup \{a\}$ such that $\mathcal{D}(a) = 1/2^k$ and $\mathcal{D}(y) = (1-1/2^k)/2^n$ for $y \in \{0,1\}^n$. Let $\mathcal{A}$ be a quantum adversary that makes q queries to the function $f \leftarrow \mathcal{D}^X$ and outputs a collision with probability ϵ.

$$\begin{aligned}\epsilon &:= \Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f] \\ &= \Pr[f(x) = f(x') = a \wedge x \neq x' : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f] (:= \epsilon_1) \\ &+ \Pr[f(x) = f(x') \neq a \wedge x \neq x' : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f] (:= \epsilon_2).\end{aligned}$$

First we obtain an upper bound for ϵ_1. Let $\mathcal{B}$ be a quantum adversary that want to distinguish between the function

$$g(x) := \begin{cases} 1 & \text{with probability } 1/2^k \\ 0 & \text{otherwise} \end{cases}$$

and the zero-function F on X. $\mathcal{B}$ picks a random function $f_0 : X \to \{0,1\}^n$ and the constant function $f_1 : X \to \{a\}$ and runs the adversary $\mathcal{A}$. It answers to $\mathcal{A}$'s queries by function $f'(x) := f_{O(x)}(x)$ where O is either g or F. Simple calculation shows that f and f' has the same distribution when $\mathcal{B}$ has oracle access to g. To simulate a quantum query to f', $\mathcal{B}$ makes two queries to its oracle (a quantum circuit that simulates the queries has been presented in the proof of Lemma 7 that uses the same idea). The adversary $\mathcal{B}$ outputs 1 if $\mathcal{A}$ outputs $x \neq x'$ such that $f'(x) = f'(x') = a$ and 0 otherwise (this includes when $\mathcal{A}$ halts without an output). Note that since the element 'a' is not in image of f_F, $\mathcal{B}$ outputs 0 with probability 1 when it interacts

with F. Therefore, the adversary $\mathcal{B}$ distinguishes g from a zero-function with probability ϵ_1 and making at most $2q+2$ queries (two extra queries needed to check that x, x' are pre-image of a). By Lemma 3, we can conclude $\epsilon_1 \leq 8(2q+2)^2/2^k$.
To prove an upper bound for ϵ_2, let $\mathcal{B}$ be a quantum adversary that has oracle access to a random function $f' : X \to \{0,1\}^n$. The adversary $\mathcal{B}$ picks function $\tilde{f}$ as the following,

$$\tilde{f}(x) := \begin{cases} a & \text{with probability } 1/2^k \\ f'(x) & \text{otherwise} \end{cases},$$

and runs $\mathcal{A}^{\tilde{f}}$. Note that f and $\tilde{f}$ has the same distribution, hence, after q queries $\mathcal{A}$ returns $x \neq x'$ such that $f(x) = f(x') \neq a$ with probability ϵ_2. By Lemma 1, $\epsilon_2 \leq \frac{C(q+2)^3}{2^n}$ and therefore $\epsilon \leq O\big(\max\{q^2/2^k, q^3/2^{3k/2}\}\big)$ since $n \geq 3k/2$. □

The theorem above implies that the adversary needs $\Omega(2^{k/2})$ queries in order to find a collision with constant probability.

Theorem 11. *For any k and any quantum adversary $\mathcal{A}$ there exists a a distribution $\mathcal{D}$ with min-entropy k such that $\Omega(2^{k/2})$ quantum queries are needed to output a collision with constant probability.*

Proof. The proof follows from the theorem above. □

4.2.2 Quantifier order $\forall\ \mathcal{D}\ \forall\ \mathcal{A}$.

It is left to prove a lower bound for any adversary $\mathcal{A}$ and any distribution $\mathcal{D}$. One could try to use the fact that every distribution of min-entropy k is a convex combination of some flat distributions over subsets of size at least 2^k [CG88] and reduces the collision problem for a non-uniform function to the uniform case . However, when the number of the flat distributions are exponentially large, the derived bound might not be suitable. To circumvent this problem, we use a similar idea used in [FRS16], namely leveling approach, and work with some nearly flat distributions. This gives us an $\Omega(2^{k/5})$ lower bound.

Leveling approach. Let $\mathcal{D}$ be a distribution with $H_\infty(\mathcal{D}) \geq k$ over $Y := \{0,1\}^n$. For $i = k, \ldots, m$, we define the distributions $\tilde{\mathcal{D}}_i$ as the following:

$$\tilde{\mathcal{D}}_i(y) := \frac{\mathcal{D}_i(y)}{\sum_{y\in\{0,1\}^n} \mathcal{D}_i(y)},$$

where for $i = k, \ldots, m-1$,

$$\mathcal{D}_i(y) := \begin{cases} \mathcal{D}(y), & \text{if } \mathcal{D}(y) \in \left(2^{-(i+1)}, 2^{-i}\right] \\ 0, & \text{otherwise} \end{cases}$$

and

$$\mathcal{D}_m(y) := \begin{cases} \mathcal{D}(y), & \text{if } \mathcal{D}(y) \in (0, 2^{-m}] \\ 0, & \text{otherwise} \end{cases}.$$

We define the distribution $\alpha(i) := \sum_{y\in\{0,1\}^n} \mathcal{D}_i(y)$ over $\{k, \ldots, m\}$. It is clear that the distribution $\mathcal{D}$ is equivalent to the distribution $\mathcal{D}''$ obtained by choosing i according to the distribution α and then picking an element according to the distribution $\tilde{\mathcal{D}}_i$. (Note that there are no values i with $D(i) > 2^{-k}$ since $H_\infty(\mathcal{D}) \geq k$.) For $i = k, \ldots, m$, let $Y_i := \{y \in \{0,1\}^n : \tilde{\mathcal{D}}_i(y) \neq 0\}$. Note that the sets of $\{Y_k, \ldots, Y_m\}$ is a partition of the set $\sup(\mathcal{D}) := \{y \in \{0,1\}^n; \mathcal{D}(y) \neq 0\}$.

Lemma 6. *For $i = k, \ldots, m-1$ with $Y_i \neq \emptyset$, any set X and any quantum adversary $\mathcal{A}$ making q queries to the function $\tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X$,*

$$\Pr[\tilde{f}(x) = \tilde{f}(x') \wedge x \neq x' : \tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X,\ (x, x') \leftarrow \mathcal{A}^{\tilde{f}}] \leq O\big(\frac{q^3}{2^i \alpha(i)}\big).$$

Proof. In order to prove the lemma, we need to define the new distributions $\mathcal{D}_i^* : Y_i \cup \{\perp\} \to [0,1]$ as the following:

$$\mathcal{D}_i^*(y) := \begin{cases} \frac{2^i}{|Y_i|}\,\mathcal{D}_i(y), & \text{if } y \in Y_i \\ 1 - \frac{2^i}{|Y_i|}\alpha(i), & \text{if } y = \perp \end{cases}.$$

It is easy to show that $\mathcal{D}_i^*$ is a distribution for $i = k, \ldots, m-1$. Then we prove that the function $f^* \leftarrow \mathcal{D}_i^{*X}$ is collision-resistant in the claim below and use it later on to show the lemma.

Claim 1. *For $i = k, \ldots, m-1$ with $Y_i \neq \emptyset$, and any quantum adversary $\mathcal{A}$ making q queries to the function $f^* \leftarrow \mathcal{D}_i^{*X}$:*

$$\epsilon_i^* := \Pr[f^*(x) = f^*(x') \wedge x \neq x' \wedge f^*(x) \in Y_i : f^* \leftarrow \mathcal{D}_i^{*X},\ (x, x') \leftarrow \mathcal{A}^{f^*}] \leq O\big(\frac{q^3}{2^i \alpha(i)}\big).$$

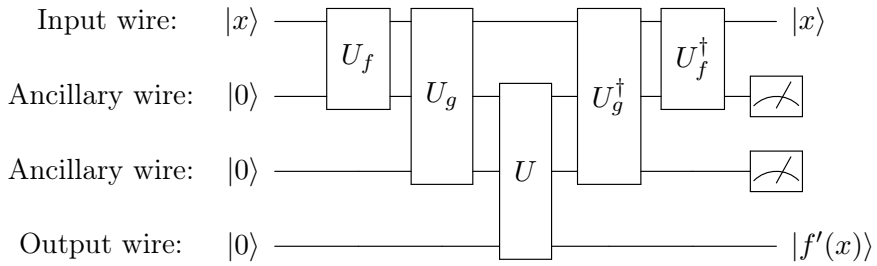
Proof. Let U_i be the uniform distribution over Y_i and $\mathcal{B}$ be a quantum adversary that has access to the function $f \leftarrow U_i^X$. The adversary $\mathcal{B}$ picks the function $g : X \times Y_i \to \{0,1\}$,

$$g(x,y) := \begin{cases} 1, & \text{with probability } 2^i\,\mathcal{D}_i(y) \\ 0, & \text{otherwise} \end{cases}$$

and answers to $\mathcal{A}$'s queries by the following function:

$$f'(x) := \begin{cases} f(x), & \text{if } g(x, f(x)) = 1 \\ \perp, & \text{otherwise} \end{cases},$$

and returns $\mathcal{A}$'s output. The quantum circuit corresponding to the function f' is drawn below.



where U is an unitary operator such that

$$U\,|y, b, z\rangle := \begin{cases} |y, b, z \oplus \perp\rangle\,, & \text{if } b = 0 \\ |y, b, z \oplus y\rangle\,, & \text{if } b = 1 \end{cases}.$$

We show that the outputs of f' are chosen independently at random according to the distribution D_i^*. Therefore, f' and f^* have the same distribution. For any $x \in X$ and $y \in Y_i$,

$$\begin{aligned} &\Pr[f'(x) = y] \\ &\quad = \Pr[g(x, f(x)) = 1 \wedge f(x) = y] \\ &\quad = \Pr[g(x, f(x)) = 1 \mid f(x) = y] \cdot \Pr[f(x) = y] \\ &\quad = 2^i\,\mathcal{D}_i(y) \cdot (1/|Y_i|) = \mathcal{D}_i^*(y), \end{aligned}$$

and consequently for any $x \in X$, $\Pr[f'(x) = \perp] = 1 - \frac{2^i}{|Y_i|}\alpha(i)$ and it is easy to see that $f'(x)$ are independent. Now since f^* and f' have the same distribution, given f' the adversary $\mathcal{A}$ outputs $x \neq x'$ such that $f'(x) = f'(x') \in Y_i$ with probability ϵ_i^*. The adversary $\mathcal{B}$ returns the output of $\mathcal{A}$ as a collision for function f and by Lemma 1, $\epsilon_i^* \leq O(\frac{q^3}{|Y_i|}) \leq O(\frac{q^3}{2^i\alpha(i)})$ where the last inequality follows for the reason that $|Y_i| \geq 2^i\alpha(i)$. □

Now using Claim 1 we prove Lemma 6, i.e.,

$$\tilde{\epsilon}_i := \Pr[\tilde{f}(x) = \tilde{f}(x') \wedge x \neq x' : \tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X,\ (x, x') \leftarrow \mathcal{A}^{\tilde{f}}] \leq O\big(\frac{q^3}{2^i\alpha(i)}\big).$$

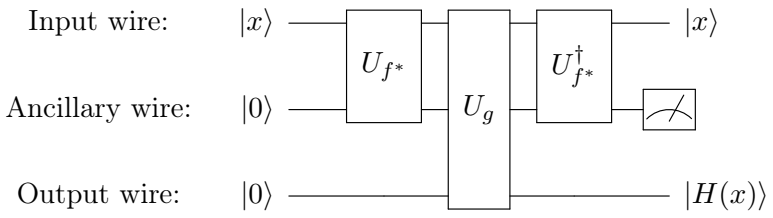
Let $\mathcal{A}$ be a quantum adversary that makes q queries to the oracle $\tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X$ and outputs a collision with probability $\tilde{\epsilon}_i$. Let $\mathcal{B}$ be a quantum adversary that has oracle access to the function $f^* \leftarrow \mathcal{D}_i^{*X}$ and does the following. The adversary $\mathcal{B}$ picks function $g : X \times (Y_i \cup \{\perp\}) \to Y_i$,

$$g(x, y) := \begin{cases} y, & \text{if } y \neq \perp \\ y \leftarrow \tilde{\mathcal{D}}_i, & \text{if } y = \perp \end{cases},$$

(let say $\mathcal{B}$ picks the function g according to the distribution $\mathcal{D}_g$) runs the adversary $\mathcal{A}$ and answers to its queries by function $H(x) := g(x, f^*(x))$ and returns $\mathcal{A}$'s output. Let $\mathcal{D}_{f^*}$ be the distribution of this function. We show that $\tilde{f}$ and H have the same distribution. For any $x \in X,\ y \in Y_i$:

$$\begin{aligned}
&\Pr[H(x) = y] \\
&\quad = \Pr[g(x, f^*(x)) = y] \\
&\quad = \sum_{y' \in Y_i} \Pr[g(x, f^*(x)) = y \wedge f^*(x) = y'] + \Pr[g(x, f^*(x)) = y \wedge f^*(x) = \perp] \\
&\quad = \sum_{y' \in Y_i} \Pr[g(x, f^*(x)) = y \mid f^*(x) = y'] \cdot \Pr[f^*(x) = y'] + \Pr[g(x, f^*(x)) = y \mid f^*(x) = \perp] \cdot \Pr[f^*(x) = \perp] \\
&\quad = \sum_{y' \in Y_i} \Pr[y' = y] \cdot \Pr[f^*(x) = y'] + \Pr[y \leftarrow \tilde{\mathcal{D}}_i] \cdot \Pr[f^*(x) = \perp] \\
&\quad = \Pr[f^*(x) = y] + \frac{\mathcal{D}_i(y)}{\alpha(i)}(1 - \frac{2^i\alpha(i)}{|Y_i|}) \\
&\quad = \frac{2^i\,\mathcal{D}_i(y)}{|Y_i|} + \frac{\mathcal{D}_i(y)}{\alpha(i)}(1 - \frac{2^i\alpha(i)}{|Y_i|}) = \frac{\mathcal{D}_i(y)}{\alpha(i)} = \tilde{\mathcal{D}}_i(y).
\end{aligned}$$

It is easy to see that $H(x)$ are independent. The quantum circuit corresponding to the function H is drawn below.



The quantum adversary $\mathcal{B}^{f^*}$ returns the output of $\mathcal{A}$ as a collision for f^*. Let $\epsilon_{\mathcal{B}}$ be the probability that $\mathcal{B}$ finds a collision. Note that $H(x) = H(x')$ implies $f^*(x) = f^*(x')$ when $f^*(x) \neq \perp$ and $f^*(x') \neq \perp$. We show that

$$\epsilon_{\mathcal{B}} = \Pr[f^*(x) = f^*(x') \wedge x \neq x' \wedge f^*(x) \in Y_i : f^* \leftarrow \mathcal{D}_i^{*X}, (x, x') \leftarrow \mathcal{B}^{f^*}] \geq \frac{1}{4}\tilde{\epsilon}_i.$$

Let $\mathcal{D}_i^*(\perp) := \delta$, and $\tilde{f}^*$ be a function with the following distribution, called $\mathcal{D}_{\tilde{f}}$,

$$\tilde{f}^*(x) := \begin{cases} \tilde{f}(x), & \text{with probability } 1-\delta \\ \perp, & \text{with probability } \delta \end{cases}.$$

We show that $\tilde{f}^*$ and f^* have the same distribution. For any $x \in X$ and $y \in Y$,

$$\Pr[\tilde{f}^*(x) = y] = \Pr[\tilde{f}(x) = y] \cdot (1-\delta) = \frac{\mathcal{D}_i(y)}{\alpha(i)} \cdot \frac{2^i}{|Y_i|}\alpha(i) = \mathcal{D}_i^*(y),$$

and it is easy to see that $\tilde{f}^*(x)$ are independent. It is clear that (H, f^*) and $(\tilde{f}, \tilde{f}^*)$ have the same distribution and $\delta \leq 1/2$. Therefore,

$$\begin{aligned}
&\Pr[f^*(x) = f^*(x') \wedge x \neq x' \wedge f^*(x) \neq \perp : f^* \leftarrow \mathcal{D}_i^{*X}, (x,x') \leftarrow \mathcal{B}^{f^*}] \\
&= \Pr[f^*(x) = f^*(x') \wedge x \neq x' \wedge f^*(x) \neq \perp : f^* \leftarrow \mathcal{D}_i^{*X}, g \leftarrow \mathcal{D}_g, H(x) := g(x, f^*(x)), (x,x') \leftarrow \mathcal{A}^H] \\
&\geq \Pr[H(x) = H(x') \wedge x \neq x' \wedge f^*(x) \neq \perp \wedge f^*(x') \neq \perp : f^* \leftarrow \mathcal{D}_i^{*X}, H \leftarrow \mathcal{D}_{f^*}, (x,x') \leftarrow \mathcal{A}^H] \\
&= \Pr[\tilde{f}(x) = \tilde{f}(x') \wedge x \neq x' \wedge \tilde{f}^*(x) \neq \perp \wedge \tilde{f}^*(x') \neq \perp : \tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X, \tilde{f}^* \leftarrow \mathcal{D}_{\tilde{f}}^X, (x,x') \leftarrow \mathcal{A}^{\tilde{f}}] \\
&= \Pr[\tilde{f}(x) = \tilde{f}(x') \wedge x \neq x' \wedge \tilde{f}^*(x) \neq \perp \wedge \tilde{f}^*(x') \neq \perp : \tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X, (x,x') \leftarrow \mathcal{A}^{\tilde{f}}, \tilde{f}^* \leftarrow \mathcal{D}_{\tilde{f}^*}] \\
&= \Pr[\tilde{f}(x) = \tilde{f}(x') \wedge x \neq x' : \tilde{f} \leftarrow \tilde{\mathcal{D}}_i^X, (x,x') \leftarrow \mathcal{A}^{\tilde{f}}](1-\delta)^2 \\
&\geq \frac{1}{4}\tilde{\epsilon}_i.
\end{aligned}$$

And by Claim 1, $\tilde{\epsilon}_B \leq O(\frac{q^3}{2^i\alpha(i)})$ and thus $\tilde{\epsilon}_i \leq O(\frac{q^3}{2^i\alpha(i)})$.

□

Lemma 7. *Let $\mathcal{D}$ be a distribution with $H_\infty(\mathcal{D}) \geq k$ over $\{0,1\}^n$ and X be some other set. Then for any $i \in [m]$, any quantum algorithm $\mathcal{A}$ making q queries to the function $f \leftarrow \mathcal{D}^X$ outputs $x \in X$ such that $f(x) \in Y_i$ with probability at most $8(2q+1)^2\alpha(i)$.*

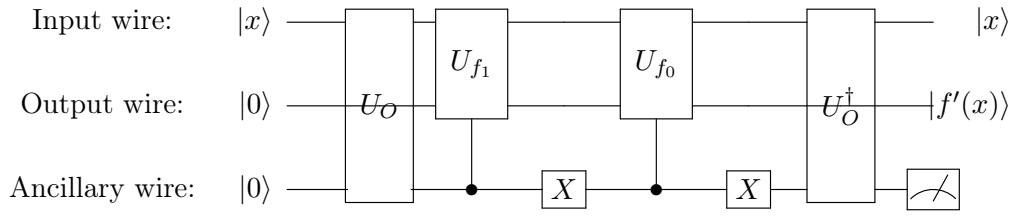
Proof. Let $\delta_i := \Pr[f(x) \in Y_i : f \leftarrow \mathcal{D}^X, x \leftarrow \mathcal{A}^f]$. Let $\mathcal{B}$ be a quantum adversary that has oracle access either to the zero function over X or the function $g : X \to \{0,1\}$,

$$g(x) := \begin{cases} 1, & \text{with probability } \alpha(i) \\ 0, & \text{otherwise} \end{cases}.$$

$\mathcal{B}$ chooses the functions $f_1 : X \to Y_i$ and $f_0 : X \to \{0,1\}^n \setminus Y_i$ as follows:

$$\begin{aligned} f_1(x) &:= y \text{ with probability } \mathcal{D}(y)/\alpha(i) \\ f_0(x) &:= y \text{ with probability } \mathcal{D}(y)/(1-\alpha(i)) \end{aligned} \tag{2}$$

The adversary $\mathcal{B}^O$ runs $\mathcal{A}$ and answers to its queries by function $f'(x) := f_{O(x)}(x)$. The following circuit depicts how the queries are simulated.



At the end, $\mathcal{B}$ outputs 1 if $f'(x) \in Y_i$ where x is the output of $\mathcal{A}$, and it outputs 0 otherwise (this includes when $\mathcal{A}$ halts without any outputs). Simple calculation shows that f and f' has the

same distribution when $\mathcal{B}$ is interacting with g, and therefore $\mathcal{A}$ outputs x such that $f'(x) \in Y_i$ with probability δ_i in this case. In the other hand, when $\mathcal{B}$ is interacting with the zero-function never output 1. Therefore $\mathcal{B}^O$ can distinguish g from the zero-function over X making at most $2q+1$ queries (an extra query needed to check $f'(x) \in Y_i$) with probability δ_i. By Lemma 3, $\delta_i \leq 8(2q+1)^2\alpha(i)$. □

Theorem 12. *Let $\mathcal{D}$ be a distribution with $H_\infty(\mathcal{D}) \geq k$ over $\{0,1\}^n$ and X be some other set. Then any quantum algorithm $\mathcal{A}$ making q queries to the function $f \leftarrow \mathcal{D}^X$ outputs a collision for f with probability at most $O\big(n\frac{q^{5/2}}{2^{k/2}}\big)$.*

Proof. Let

$$\beta_i := \Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f]$$

for any $i \in [m]$. By Lemma 7, it is clear that $\beta_i \leq 8(2q+1)^2\alpha(i)$. We obtain another upper bound for β_i when $i < m$. Note that the distribution $\mathcal{D}$ is equivalent to the distribution $\mathcal{D}''$ obtained by choosing i according to the distribution α and then picking an element according to the distribution $\tilde{\mathcal{D}}_i$. We define distribution $\mathcal{D}'$ over partitions of the set X into subset $X_k, \ldots, X_m$ as follows: for every $x \in X$ we put x in the set X_i with probability $\alpha(i)$. Let P be the set of all such partitions. We can conclude:

$$\begin{aligned}
&\Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : f \leftarrow \mathcal{D}^X,\ (x,x') \leftarrow \mathcal{A}^f] \\
&\quad = \Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : f \leftarrow \mathcal{D}''^X,\ (x,x') \leftarrow \mathcal{A}^f] \\
&\quad = \Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : (X_k,\ldots,X_m) \leftarrow \mathcal{D}', \forall j \in \{k,\cdots,m\} \\
&\qquad\qquad f_j \leftarrow \tilde{\mathcal{D}}_j^{X_j}, f = \bigcup f_j,\ (x,x') \leftarrow \mathcal{A}^f] \\
&\quad = \sum_{(X_k,\ldots,X_m)\in P} \mathcal{D}'\big((X_k,\ldots,X_m)\big) \cdot \Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : \forall j \in \{k,\cdots,m\} \\
&\qquad\qquad f_j \leftarrow \tilde{\mathcal{D}}_j^{X_j}, f = \bigcup f_j,\ (x,x') \leftarrow \mathcal{A}^f] \\
&\quad \leq \sum_{(X_k,\ldots,X_m)\in P} \mathcal{D}'\big((X_k,\ldots,X_m)\big) \cdot \sum_{j=k}^{m} \Pr[f_j(x) = f_j(x') \wedge x \neq x' \wedge f_j(x) \in Y_i : \forall j \in \{k,\cdots,m\} \\
&\qquad\qquad f_j \leftarrow \tilde{\mathcal{D}}_j^{X_j}, f = \bigcup f_j,\ (x,x') \leftarrow \mathcal{A}^f] \\
&\quad \overset{(*)}{=} \sum_{(X_k,\ldots,X_m)\in P} \mathcal{D}'\big((X_k,\ldots,X_m)\big) \cdot \sum_{j=k}^{m} \Pr[f_j(x) = f_j(x') \wedge x \neq x' \wedge f_j(x) \in Y_i : \forall j \in \{k,\cdots,m\} \\
&\qquad\qquad f_j \leftarrow \tilde{\mathcal{D}}_j^{X_j}, f = \bigcup f_j,\ (x,x') \leftarrow \mathcal{B}^{f_j}_{j,X_k,\ldots,X_m}] \\
&\quad \overset{(**)}{=} \sum_{(X_k,\ldots,X_m)\in P} \mathcal{D}'\big((X_k,\ldots,X_m)\big) \cdot \Pr[f_i(x) = f_i(x') \wedge x \neq x' \wedge f_i(x) \in Y_i : f_i \leftarrow \tilde{\mathcal{D}}_i^{X_i}, \\
&\qquad\qquad (x,x') \leftarrow \mathcal{B}^{f_i}_{i,X_k,\ldots,X_m}] \\
&\quad \overset{(***)}{\leq} O\big(\frac{q^3}{2^i\alpha(i)}\big),
\end{aligned}$$

where $(*)$ holds because we define $\mathcal{B}^{f_j}_{j,X_k,\ldots,X_m}$ to be a quantum adversary that picks $f_\ell \leftarrow \tilde{\mathcal{D}}_\ell^{X_\ell}$ for $\ell \in \{k,\cdots,m\}$ and $\ell \neq j$, runs $\mathcal{A}^{\bigcup f_i}$ and returns $\mathcal{A}$'s output. The equality labeled by $(**)$ holds since for any $x \in X_j$, $f_j(x) \notin Y_i$ when $j \neq i$. Finally, $(***)$ is obtained by Lemma 6.

Using the bounds derived above:

$$\begin{aligned}
&\Pr[f(x) = f(x') \wedge x \neq x' : f \leftarrow \mathcal{D}^X,\ (x, x') \leftarrow \mathcal{A}^f] \\
&= \sum_{i=k}^{m} \Pr[f(x) = f(x') \wedge x \neq x' \wedge f(x) \in Y_i : f \leftarrow \mathcal{D}^X,\ (x, x') \leftarrow \mathcal{A}^f] \\
&\leq \sum_{i=k}^{m-1} O\Big(\min\big\{\frac{q^3}{2^i\alpha(i)}, q^2\alpha(i)\big\}\Big) + O\big(q^2\alpha(m)\big) \\
&\overset{(*)}{\leq} (m-k-1)O\Big(\frac{q^{5/2}}{2^{k/2}}\Big) + O\big(q^2\alpha(m)\big)
\end{aligned}$$

where $(*)$ holds because $\min\big\{\frac{q^3}{2^i\alpha(i)}, q^2\alpha(i)\big\}$ will be maximised when $\frac{q^3}{2^i\alpha(i)} = q^2\alpha(i)$ and the maximum value is $q^{5/2}/2^{i/2}$ (Note that the function $f(x) := \frac{q^3}{2^i x}$ is strictly decreasing and the function $g(x) := q^2 x$ is strictly increasing.). Choosing $m = n + k$ results in $q^2\alpha(m) \leq q^2|Y_m|/2^m \leq q^2 2^{n-m} \leq q^2/2^k$ and this proves the bound stated in the theorem. □

5 Acknowledgment

We would like to thank Will Perkins, Gelo Noel Tabia and Tore Vincent Carstens for helpful discussion about this work. This work was supported by institutional research funding IUT2-1 of the Estonian Ministry of Education and Research, by the Estonian ICT program 2011-2015 (3.2.1201.13-0022).

References

[Amb00] Andris Ambainis. Quantum lower bounds by quantum arguments. In F. Frances Yao and Eugene M. Luks, editors, *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 636–643. ACM, 2000.

[Amb05] Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005.

[Amb07] Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007.

[AS04] Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, July 2004.

[BBHT98] Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4-5):493–505, 1998.

[BHT97] Gilles Brassard, Peter Høyer, and Alain Tapp. Quantum algorithm for the collision problem. *ACM SIGACT News (Cryptology Column)*, 28:14–19, 1997.

[BR93] Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In Dorothy E. Denning, Raymond Pyle, Ravi Ganesan, Ravi S. Sandhu, and Victoria Ashby, editors, *CCS '93, Proceedings of the 1st ACM Conference on Computer and Communications Security, Fairfax, Virginia, USA, November 3-5, 1993.*, pages 62–73. ACM, 1993.

[CG88] Benny Chor and Oded Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988.

[CW79] Larry Carter and Mark N. Wegman. Universal classes of hash functions. *J. Comput. Syst. Sci.*, 18(2):143–154, 1979.

[FO99] Eiichiro Fujisaki and Tatsuaki Okamoto. Secure integration of asymmetric and symmetric encryption schemes. In *Proceedings of the 19th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '99, pages 537–554, London, UK, UK, 1999. Springer-Verlag.

[FRS16] Benjamin Fuller, Leonid Reyzin, and Adam D. Smith. When are fuzzy extractors possible? In Jung Hee Cheon and Tsuyoshi Takagi, editors, *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 I*, volume 10031 of *Lecture Notes in Computer Science*, pages 277–306, 2016.

[Gro96] Lov K. Grover. A fast quantum mechanical algorithm for database search. In Gary L. Miller, editor, *Proceedings of the Twenty-Eighth Annual ACM Symposium on the Theory of Computing, Philadelphia, Pennsylvania, USA, May 22-24, 1996*, pages 212–219. ACM, 1996.

[HILL93] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Construction of a pseudo-random generator from any one-way function. *SIAM Journal on Computing*, 28:12–24, 1993.

[HRS16] Andreas Hülsing, Joost Rijneveld, and Fang Song. Mitigating multi-target attacks in hash-based signatures. In Chen-Mou Cheng, Kai-Min Chung, Giuseppe Persiano, and Bo-Yin Yang, editors, *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*, volume 9614 of *Lecture Notes in Computer Science*, pages 387–416. Springer, 2016.

[Jen06] Johan L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, 30(1):175–193, 1906.

[KL07] Jonathan Katz and Yehuda Lindell. *Introduction to Modern Cryptography*. Chapman and Hall/CRC Press, 2007.

[MU05] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, New York, NY, USA, 2005.

[TTU16] Ehsan Ebrahimi Targhi, Gelo Noel Tabia, and Dominique Unruh. Quantum collision-resistance of non-uniformly distributed functions. In Tsuyoshi Takagi, editor, *Post-Quantum Cryptography - 7th International Workshop, PQCrypto 2016, Fukuoka, Japan, February 24-26, 2016, Proceedings*, volume 9606 of *Lecture Notes in Computer Science*, pages 79–85. Springer, 2016.

[TU16] Ehsan Ebrahimi Targhi and Dominique Unruh. Post-quantum security of the Fujisaki-Okamoto and OAEP transforms. In Martin Hirt and Adam D. Smith, editors, *Theory of Cryptography - 14th International Conference, TCC 2016-B, Beijing, China, October 31 - November 3, 2016, Proceedings, Part II*, volume 9986 of *Lecture Notes in Computer Science*, pages 192–216, 2016.

[Wie05] Michael J. Wiener. Bounds on birthday attack times. *IACR Cryptology ePrint Archive*, 2005:318, 2005.

[Yue14] Henry Yuen. A quantum lower bound for distinguishing random functions from random permutations. *Quantum Information & Computation*, 14(13-14):1089–1097, 2014.

[Zha12] Mark Zhandry. 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*, pages 679–687. IEEE Computer Society, 2012.

[Zha15] Mark Zhandry. A note on the quantum collision and set equality problems. *Quantum Information & Computation*, 15(7&8):557–567, 2015.