Finding Random Collisions for Random Degree-2 Functions

Xinyu Mao
University of Southern California
xinyumao.tcs@gmail.com

May 24, 2026

Abstract

We study distributional collision resistance for random degree-2 functions over prime fields. Let p be a prime and let

$$h : \mathbb{F}_p^N \to \mathbb{F}_p^M,\ x \mapsto (h_1(x), \ldots, h_M(x)) \qquad M < N,$$

be a random polynomial map where each coordinate h_i is independently, uniformly chosen at random from all polynomials of degree at most 2 over $\mathbb{F}_p$. The ideal collision distribution is obtained by choosing $x \leftarrow \mathbb{F}_p^N$ and $y \leftarrow h^{-1}(h(x))$ uniformly at random and output (x, y). We give an efficient algorithm whose output distribution is statistically close to this ideal collision distribution whenever p^{N-M} is superpolynomial. In other words, we show that random degree-2 functions are not distributionally collision-resistant if p^{N-M} is superpolynomial, resolving an open question posed by Bitansky, Haitner, Komargodski, and Yogev (Eurocrypt 2019) in this parameter regime.

1 Introduction

Collision-resistant hash function (CRH) is a central primitive in cryptography. A hash family $\mathcal{H}$ is collision resistant if, for a random function $h \leftarrow \mathcal{H}$, it is computationally infeasible to find distinct inputs $x \neq y$ with $h(x) = h(y)$. A particularly clean algebraic candidate is a random low-degree polynomial map over a finite field. Degree two is the first nonlinear case and is closely tied to the multivariate-quadratic, or MQ, problem. For a prime p, the natural quadratic hash function family $\mathcal{Q}_{p,M,N}$ consists of maps

$$h : \mathbb{F}_p^N \to \mathbb{F}_p^M, \qquad h(x)_i = \gamma_i + \ell_i^\top x + \sum_{1 \le a \le b \le N} q_{i,a,b} x_a x_b, \qquad i \in [M], \tag{1}$$

where $\gamma_i \in \mathbb{F}_p,\ \ell_i \in \mathbb{F}_p^N,\ q_{i,a,b} \in \mathbb{F}_p$ for all a, b, and i. To sample a random hash function (key), one simply chooses all coefficients $(\gamma_i, \ell_i, q_{i,a,b})_{i,a,b}$ uniformly at random.

The ordinary collision-resistance status of this family is well understood. Applebaum, Haramaty, Ishai, Kushilevitz, and Vaikuntanathan [AHI+17] showed that the random degree-2 family is target-collision resistant, equivalently a universal one-way hash family, under the corresponding MQ assumption. They also formalized a simple attack due to Ding and Yang [DY08], showing that the same random family is not collision-resistant. The attack chooses a nonzero difference Δ and solves

$$h(x + \Delta) = h(x).$$

Since h is quadratic, this is a linear system in the unknown x. With constant probability, the system is solvable and yields a collision $(x, x+\Delta)$.

This leaves open the possibility that random degree-2 functions satisfy other weaker notions of collision-resistance. For example, *multi-collision resistance* asks that it be hard to find many distinct inputs that are mapped to the same value, even if finding a colliding pair is easy [BDRV18, BKP18, KNY18, RV24, MZ25]. Zhandry's quantum lightning scheme used a non-affine multi-collision resistance assumption for random quadratic maps over $\mathbb{F}_2$, motivated by the fact that quadratic maps admit structured affine multi-collisions in some parameter regimes [Zha21]. Another notion called *distributional collision resistance* asks for a different kind of hardness: it should be hard to sample a collision pair that is statistically close to the ideal random-collision distribution. This notion originates in the work of Dubrov and Ishai and was developed further by Komargodski and Yogev and by Bitansky, Haitner, Komargodski, and Yogev [DI06, KY18, BHKY19].

For a fixed shrinking map $h : \mathbb{F}_p^N \to \mathbb{F}_p^M$, the ideal collision distribution, denoted Col_h is defined as follows:

$$x \leftarrow \mathbb{F}_p^N, \qquad y \leftarrow h^{-1}(h(x)), \qquad \text{output } (x, y).$$

At first sight, the Ding–Yang attack need not break distributional collision resistance. The main point of this paper is that, for random quadratic maps with sufficient shrinkage, the differential attack can produce a collision pair close to the ideal distribution.

1.1 Main Result

We present a simple collision-finding algorithm for the function family $\mathcal{Q}_{p,N,M}$, denoted Diff, and prove the following upper bound for the statistical distance between the output distribution of $\mathsf{Diff}(h)$ and Col_h.

Theorem 1.1. *Let p be a prime and let $M < N$. Let $\sigma = N - M$ denote the shrinkage. For $h \leftarrow\!\!\$\ \mathcal{Q}_{p,N,M}$, it holds that*

$$\mathsf{SD}((h, \mathsf{Diff}(h)) \ ; \ (h, \mathsf{Col}_h)) \leq \frac{1}{2}p^{-\sigma/2} + p^{-\sigma} + 3 \cdot p^{-(\sigma+1)}.$$

Corollary 1.2 (Random degree-2 functions are not dCRH)**.** *Let $\{(p_\lambda, N_\lambda, M_\lambda)\}_{\lambda\in\mathbb{N}}$ be an efficiently computable parameter sequence with p_λ prime and $M_\lambda < N_\lambda$. If $p_\lambda^{N_\lambda - M_\lambda}$ is superpolynomial in λ, then the function family ensemble $(\mathcal{Q}_{p_\lambda,N_\lambda,M_\lambda})_{\lambda\in\mathbb{N}}$ is not distributional collision resistant.*

1.2 Discussion and Related Work

Multi-collision resistance. Zhandry observed that the differential attack can be generalized to find multi-collisions for random degree-2 functions, and these multi-collisions form an affine space. Though the number of collisions could be large, the dimension of the affine space containing them is relatively small. Based on this observation, Zhandry defined *non-affine multi-collision resistance (NAMCR)* and constructed quantum lightning assuming NAMCR of $\mathcal{Q}_{p,N,M}$ [Zha21]. More precisely, we say $\mathcal{Q}_{p,N,M}$ satisfies K-NAMCR if, for a random $h \leftarrow \mathcal{Q}_{p,N,M}$, it is infeasible to find K inputs $x_1, \ldots, x_K$ such that (1) they do not lie in any $(K-1)$-dimension affine space, and (2) $h(x_1) = \cdots = h(x_K)$. Investigating K-NAMCR of $\mathcal{Q}_{p,N,M}$ is an intriguing question.

Distributional CRH vs. CRH. It is natural to ask whether distributional CRH (dCRH) implies CRH. Komargodski and Yogev showed that dCRH can be constructed from the average hardness of **SZK** [KY18]. Therefore, if one aims to study the relationship between CRH and the average hardness of **SZK**, it is a good starting point to clarify the relationship between dCRH and CRH first.

2 Preliminaries

2.1 Notation

All vector spaces are over a prime field $\mathbb{F}_p$. For $n \in \mathbb{N}$, we write $[n] = \{1, \ldots, n\}$. We use $\leftarrow$ to denote sampling from a distribution, choosing an element from a set uniformly at random, or collecting the output of a randomized algorithm.

A function $\nu : \mathbb{N} \to [0, 1]$ is said to be negligible if for every $c \in \mathbb{N}$, $\nu(\lambda) \leq 1/(c\lambda^c)$ for sufficiently large λ.

Throughout this paper, we consider functions from $\mathbb{F}_p^N$ to $\mathbb{F}_p^M$ where p is prime and $M < N$. Write $\sigma \stackrel{\mathsf{def}}{=} N - M$ and $\mu \stackrel{\mathsf{def}}{=} p^\sigma$.

Definition 2.1. For a function $h : \mathcal{X} \to \mathcal{Y}$, and $y \in \mathcal{Y}$, we call $h^{-1}(y) = \{x \in \mathcal{X} : h(x) = y\}$ a *fiber* of h, define the *fiber size* $F_{h,y}$ as $F_{h,y} \stackrel{\mathsf{def}}{=} |h^{-1}(y)|$.

2.2 Degree-2 Functions

Definition 2.2 (Random degree-2 function family over $\mathbb{F}_p$)**.** For a prime p and integers $M < N$, the family $\mathcal{Q}_{p,N,M}$ consists of functions $h : \mathbb{F}_p^N \to \mathbb{F}_p^M$ defined as follows. Each function $h \in \mathcal{Q}_{p,N,M}$ is specified by coefficients

$$\gamma_i \in \mathbb{F}_p, \qquad \ell_i = (\ell_{i,1}, \ldots, \ell_{i,N}) \in \mathbb{F}_p^N, \qquad q_{i,a,b} \in \mathbb{F}_p \quad (1 \leq a \leq b \leq N), \quad i \in [M],$$

and the i-th coordinate of $h(x)$ is given by

$$h(x)_i = \gamma_i + \sum_{j=1}^{N} \ell_{i,j} x_j + \sum_{1 \leq a \leq b \leq N} q_{i,a,b} x_a x_b.$$

Note that the constants γ_i do not affect any collision relation; hence, in the following, we sometimes omit the constants.

Remark 2.3 (An equivalent matrix formulation)**.** The definition above is equivalent to the usual upper-triangular matrix notation where h is specified by

$$\mathbf{Q} = (Q_1, \ldots, Q_M) \in (\mathbb{F}_p^{N \times N})^M, \ \mathbf{L} = (L_1, \ldots, L_M) \in (\mathbb{F}_p^N)^M, \ \gamma_1, \ldots, \gamma_M \in \mathbb{F}_p$$

and the i-th coordinate of $h(x)$ is given by

$$h(x)_i = x^\top Q_i x + L_i^\top x + \gamma_i.$$

That is, the entries of Q_i encode the coefficients $q_{i,a,b}$ with $a \leq b$.

2.3 Distributional Collisions

Definition 2.4 (Ideal collision distribution)**.** For a function $h : X \to Y$, define Col_h as the distribution over ordered pairs obtained by

$$x \leftarrow \mathbb{F}_p^N, \qquad x' \leftarrow h^{-1}(h(x)), \qquad \text{output } (x, x').$$

The pair may be *diagonal*, i.e., $x = x'$.

Definition 2.5 (Uniform collision-pair distributions)**.** For a function $h : \mathbb{F}_p^N \to \mathbb{F}_p^M$, define

$$\mathsf{Pairs}(h) = \left\{(x, y) \in (\mathbb{F}_p^N)^2 \,:\, h(x) = h(y)\right\}$$

and

$$\mathsf{Pairs}_{\neq}(h) = \left\{(x, y) \in (\mathbb{F}_p^N)^2 \,:\, x \neq y \wedge h(x) = h(y)\right\}.$$

Let AllCol_h be the uniform distribution on $\mathsf{Pairs}(h)$, and let UCol_h be the uniform distribution on $\mathsf{Pairs}_{\neq}(h)$. Since $M < N$, $\mathsf{Pairs}_{\neq}(h)$ is nonempty for every h.

Definition 2.6 (Distributional collision resistance)**.** Let $\mathcal{H} = \{\mathcal{H}_\lambda\}_{\lambda \in \mathbb{N}}$ be an efficiently samplable ensemble of function families. We say that $\mathcal{H}$ is *distributional collision resistant* if for every PPT algorithm $\mathcal{A}$, the statistical distance between

$$(h, \mathcal{A}(h)) \qquad \text{and} \qquad (h, \mathsf{Col}_h), \qquad h \leftarrow \mathcal{H}_\lambda,$$

is not negligible. Equivalently, no efficient algorithm can produce a distribution negligibly close to the ideal random-collision experiment.

3 The Differential Attack

For $\Delta \in \mathbb{F}_p^N \setminus \{0\}$ and $h \in \mathcal{Q}_{p,N,M}$, define the derivative of h in direction Δ by

$$D_\Delta h(x) = h(x + \Delta) - h(x).$$

Since h has degree at most two, $D_\Delta h$ is affine-linear. Thus there are $A_{h,\Delta} \in \mathbb{F}_p^{M \times N}$ and $b_{h,\Delta} \in \mathbb{F}_p^M$ such that

$$D_\Delta h(x) = A_{h,\Delta} x + b_{h,\Delta}. \tag{2}$$

The pair $(x, x + \Delta)$ is a collision pair if and only if

$$A_{h,\Delta} x = -b_{h,\Delta}.$$

If we use the matrix notation as in Remark 2.3, we can explicitly write

$$A_{h,\Delta} = \begin{pmatrix} \Delta^\top (Q_1 + Q_1^\top) \\ \Delta^\top (Q_2 + Q_2^\top) \\ \vdots \\ \Delta^\top (Q_M + Q_M^\top) \end{pmatrix} \in \mathbb{F}_p^{M \times N}, \quad b_{h,\Delta} = h(\Delta) = \begin{pmatrix} \Delta^\top Q_1 \Delta + L_1^\top \Delta \\ \Delta^\top Q_2 \Delta + L_2^\top \Delta \\ \vdots \\ \Delta^\top Q_M \Delta + L_M^\top \Delta \end{pmatrix} \in \mathbb{F}_p^M.$$

Definition 3.1 (Differential attack)**.** The algorithm Diff, on input $h \in \mathcal{Q}_{p,N,M}$, operates as follows.

1. Sample $\Delta \leftarrow\!\!\$\ \mathbb{F}_p^N \backslash \{0\}$, compute $A_{h,\Delta}, b_{h,\Delta}$ as in Equation (2), and test whether $\mathrm{rank}(A_{h,\Delta}) = M$.
2. If $\mathrm{rank}(A_{h,\Delta}) = M$, sample x uniformly from the affine solution space

$$\mathsf{Sol}_h(\Delta) \stackrel{\mathsf{def}}{=} \left\{ x \in \mathbb{F}_p^N \ : \ A_{h,\Delta} x = -b_{h,\Delta} \right\}$$

   and output $(x, x + \Delta)$.
3. Otherwise, output $(0, 0)$.

Definition 3.2 (Good difference)**.** For $h \in \mathcal{Q}_{p,M,N}$, call $\Delta \in \mathbb{F}_p^N \backslash \{0\}$ good w.r.t. h if $\mathrm{rank}(A_{h,\Delta}) = M$. Write

$$\mathsf{Good}_h \stackrel{\mathsf{def}}{=} \left\{ \Delta \in \mathbb{F}_p^N \setminus \{0\} \ : \ \mathrm{rank}(A_{h,\Delta}) = M \right\}.$$

It is easy to see that $\mathsf{Diff}(h)$ runs in time polynomial in N, M, and $\log p$ using Gaussian elimination to test rank and to sample a uniform solution. When a good difference is found, the output is always a nontrivial collision pair.

The proof of Theorem 1.1 is presented in two parts. In the rest of this section, we shall prove that the distribution of $(h, \mathsf{Diff}(h))$ is statistically close to (h, UCol_h):

Lemma 3.3 (From $\mathsf{Diff}(h)$ to UCol_h)**.** *For $h \leftarrow \mathcal{Q}_{p,N,M}$, it holds that*

$$\mathsf{SD}((h, \mathsf{Diff}(h)); (h, \mathsf{UCol}_h)) \le 3p^{-(\sigma+1)}.$$

Then, Section 4 shows that (h, UCol_h) is statistically close to (h, Col_h).

3.1 The Rank Failure

We first show that the rank test in $\mathsf{Diff}(h)$ fails with tiny probability. We start by characterizing the distribution of $A_{h,\Delta}$ for fixed $\Delta \neq 0$. The situation is slightly different for $p = 2$. Observe that the rows of $A_{h,\Delta}$ and b_Δ are independently distributed. Hence, it suffices to study the distribution of one row.

Lemma 3.4 (Row distribution)**.** *Fix $\Delta \in \mathbb{F}_p^N \setminus \{0\}$ and one output coordinate $i \in [M]$, and consider the random choice of $h \leftarrow \mathcal{Q}_{p,N,M}$. Let $a_{i,\Delta} \in \mathbb{F}_p^N$ be the i-th row of $A_{h,\Delta}$, so that*

$$h(x + \Delta)_i - h(x)_i = a_{i,\Delta}^\top x + b_{i,\Delta}.$$

Then the following hold.

(a) If p is odd, then $a_{i,\Delta}$ is uniform over $\mathbb{F}_p^N$.

(b) If $p = 2$, then $a_{i,\Delta}$ is uniform over the hyperplane $\Delta^\perp = \left\{ a \in \mathbb{F}_2^N \ : \ a^\top \Delta = 0 \right\}$.

(c) In both cases, $b_{i,\Delta}$ is uniform over $\mathbb{F}_p$ even conditioned on $a_{i,\Delta}$.

Proof. Fix $i \in [M]$ and omit it from the subscripts. For one output coordinate, write the quadratic part as

$$q(x) = \sum_{1 \le a \le b \le N} q_{a,b} x_a x_b.$$

The coefficient of x_j in $q(x+\Delta)-q(x)$ is

$$2q_{j,j}\Delta_j + \sum_{a<j} q_{a,j}\Delta_a + \sum_{b>j} q_{j,b}\Delta_b. \tag{3}$$

For odd p, the variables $2q_{j,j}$ are uniform because 2 is invertible. Equivalently, a_Δ can be viewed as $a_\Delta = S\Delta$, where S is a uniformly random symmetric $N \times N$ matrix over $\mathbb{F}_p$. We claim that

Claim 3.5. *The linear map $T_\Delta : \mathrm{Sym}_N(\mathbb{F}_p) \to \mathbb{F}_p^N$, $T_\Delta(S) \mapsto S\Delta$, is surjective, where $\mathrm{Sym}_N(\mathbb{F}_p)$ denotes the set of symmetric $N \times N$ matrix over $\mathbb{F}_p$.*

Proof of claim. Choose k such that $\Delta_k \neq 0$. Given any $u \in \mathbb{F}_p^N$, set

$$S_{j,k} = S_{k,j} = u_j/\Delta_k \quad (j \neq k),$$

set all entries not involving k to zero, except possibly $S_{k,k}$, and then choose

$$S_{k,k} = \frac{u_k - \sum_{j\neq k} S_{k,j}\Delta_j}{\Delta_k}.$$

Then it is easy to see that S is symmetric and $S\Delta = u$. □

A linear map sends the uniform distribution on a finite vector space to the uniform distribution on its image, so a_Δ is uniformly distributed over $\mathbb{F}_p^N$.

For $p = 2$, the diagonal term in (3) vanishes. We can view a_Δ as $S\Delta$, where S is a uniformly random symmetric matrix with zero diagonal. Such matrices are alternating, so $(S\Delta)^\top\Delta = \Delta^\top S\Delta = 0$. Thus a_Δ always lies in $\Delta^\perp$. Conversely, the linear map $S \mapsto S\Delta$ is onto $\Delta^\perp$. The argument is similar to Claim 3.5. Pick k with $\Delta_k = 1$. Given any $u \in \Delta^\perp$, set $S_{j,k} = u_j$ for $j \neq k$, set all other entries not forced by symmetry to zero, and set $S_{k,k} = 0$. Then, for $j \neq k$, $(S\Delta)_j = u_j$; for the k-th coordinate,

$$(S\Delta)_k = \sum_{j\neq k} u_j\Delta_j = u_k,$$

where the last equality uses $u^\top\Delta = 0$ and $\Delta_k = 1$. Hence a_Δ is uniform over $\Delta^\perp$.

Finally, the affine offset b_Δ contains the term $\ell^\top\Delta$. Since ℓ is uniform and $\Delta \neq 0$, $\ell^\top\Delta$ is uniform over $\mathbb{F}_p$ and independent of all quadratic coefficients. Therefore b_Δ is uniform even conditioned on the a_Δ. □

Lemma 3.6 (Rank failure)**.** *For all $\Delta \in \mathbb{F}_p^N \setminus \{0\}$, it holds that*

$$\Pr_{h\leftarrow\mathcal{Q}_{p,N,M}}[\mathrm{rank}(A_\Delta) < M] < p^{-(\sigma+1)}.$$

Proof. By Lemma 3.4, after choosing a basis for the row-supporting subspace, A_Δ is an $M \times (N - \kappa_p)$ matrix with independent uniformly random rows, where

$$\kappa_p = \begin{cases} 1, & p = 2, \\ 0, & p \text{ odd}, \end{cases}.$$

The probability that such a matrix has full row rank is

$$\prod_{t=0}^{M-1}(1-p^{t-N+\kappa_p}) \geq \prod_{t=0}^{M-1}(1-p^{t-N}) = \prod_{j=N-M+1}^{N}(1-p^{-j}).$$

Using $1-\prod_j(1-a_j) \leq \sum_j a_j$ for $a_j \in [0,1]$, we have

$$\Pr_{h\leftarrow\mathcal{Q}_{p,N,M}}[\mathrm{rank}(A_\Delta) < M] \leq 1 - \prod_{j=N-M+1}^{N}(1-p^{-j}) \leq \sum_{j=N-M+1}^{N} p^{-j} \leq \frac{p^{-(N-M)}}{p-1} < p^{-(\sigma+1).}$$

□

3.2 Diff Produces Uniform Collision Pairs

We now show that the output distribution of the differential attack algorithm Diff is close to UCol_h. To this end, we first show that, conditioned on the difference being good w.r.t. h, the output distribution of $\mathsf{Diff}(h)$ is exactly UCol_h under the same conditioning.

Lemma 3.7 (Collision with good difference)**.** *Fix $h \in \mathcal{Q}_{p,N,M}$ such that $\mathsf{Good}_h \neq \emptyset$. For all $\Delta \in \mathbb{F}_p^N \setminus \{0\}$, if $\Delta \in \mathsf{Good}_h$, then $|\mathsf{Sol}_h(\Delta)| = \mu$. Consequently, the following two distributions are identical:*

- *$\mathcal{D}_0$: Sample $\Delta \leftarrow \mathsf{Good}_h$, $x \leftarrow \mathsf{Sol}_h(\Delta)$, and output $(x, x+\Delta)$.*
- *$\mathcal{D}_1$: Sample $(x,y) \leftarrow \mathsf{UCol}_h$ conditioned on $y - x \in \mathsf{Good}_h$.*

Proof. If $\Delta \in \mathsf{Good}_h$, then $A_{h,\Delta}$ has rank M. Hence the affine system $A_{h,\Delta}x = -b_{h,\Delta}$ has $p^{N-M} = \mu$ solutions. These solutions are exactly the x's such that $(x, x+\Delta)$ is a collision pair.

Write $C_h(\Delta) \stackrel{\mathsf{def}}{=} \{(x, x+\Delta) : x \in \mathsf{Sol}_h(\Delta)\}$. Note that $\{C_h(\Delta)\}_{\Delta\in\mathbb{F}_p^N\setminus\{0\}}$ forms a partition of $\mathsf{Pairs}_{\neq}(h)$. Hence, sampling from $\mathcal{D}_1$ is equivalent to choosing a pair uniformly at random from the set

$$\bigcup_{\Delta\in\mathsf{Good}_h} C_h(\Delta),$$

where the union is disjoint. Since $|C_h(\Delta)| = |\mathsf{Sol}_h(\Delta)| = \mu$ for all $\Delta \in \mathsf{Good}_h$, this is equvalent to $\mathcal{D}_0$. □

Proof of Lemma 3.3. Let (H, X, Y) denote the output distribution of $\mathsf{UCol}(H)$ where H is the uniform distribution over $\mathcal{Q}_{p,N,M}$. In the process of $\mathsf{Diff}(h)$, conditioned on the event that $\Delta \in \mathsf{Good}_h$, the output distribution is identical to $(X,Y)|_{Y-X\in\mathsf{Good}_H}$ according to Lemma 3.7. Therefore, for $h \leftarrow \mathcal{Q}_{p,N,M}$,

$$\mathsf{SD}((h,\mathsf{Diff}(h)); (H,(X,Y))|_{Y-X\in\mathsf{Good}_h}) \leq \Pr_{h,\Delta}[\Delta \notin \mathsf{Good}_h] \leq p^{-(\sigma+1)},$$

where the last inequality follows from Lemma 3.6. Meanwhile, by definition,

$$\mathsf{SD}((h,\mathsf{UCol}(h))\ ;\ (H,(X,Y))|_{Y-X\in\mathsf{Good}_h}) \leq \Pr_{h\leftarrow\mathcal{Q}_{p,N,M},(x,y)\leftarrow\mathsf{UCol}_h}[y-x \notin \mathsf{Good}_h].$$

Then by triangle inequality,

$$\mathsf{SD}((h,\mathsf{Diff}(h))\ ;\ (H,\mathsf{UCol}(H)))$$

$$\leq \mathsf{SD}((h,\mathsf{Diff}(h))\ ;\ (H,(X,Y))|_{Y-X\in\mathsf{Good}_H}) + \mathsf{SD}((h,\mathsf{UCol}(h))\ ;\ (H,(X,Y))|_{Y-X\in\mathsf{Good}_H})$$
$$\leq p^{-(\sigma+1)} + \mathop{\mathbf{Pr}}_{h\leftarrow\mathcal{Q}_{p,N,M},(x,y)\leftarrow\mathsf{UCol}_h}[y-x\notin\mathsf{Good}_h],$$

and thus it suffices to prove the following claim:

Claim 3.8. $\mathbf{Pr}_{h\leftarrow\mathcal{Q}_{p,N,M},(x,y)\leftarrow\mathsf{UCol}_h}[y-x\notin\mathsf{Good}_h]\leq 2p^{-(\sigma+1)}$.

Proof of claim. For a fixed h,

$$\mathop{\mathbf{Pr}}_{(x,y)\leftarrow\mathsf{UCol}_h}[y-x\notin\mathsf{Good}_h] = \frac{\sum_{\Delta\in\mathbb{F}_p^M\setminus\{0\}:\Delta\notin\mathsf{Good}_h}|\mathsf{Sol}_h(\Delta)|}{\sum_{\Delta\in\mathbb{F}_p^M\setminus\{0\}}|\mathsf{Sol}_h(\Delta)|}.$$

The denominator is $|\mathsf{Pairs}_{\neq}(h)| = \sum_y F_{h,y}(F_{h,y}-1)$. Since $\sum_y F_{h,y} = p^N$, we have

$$\sum_y F_{h,y}^2 \geq p^M\mu^2 = p^N\mu,$$

and therefore

$$\sum_y F_{h,y}(F_{h,y}-1) = \sum_y F_{h,y}^2 - p^N \geq p^N(\mu-1). \tag{4}$$

We next compute the expectation of the numerator over a random h. Fix $\Delta\neq 0$. Observe that the condition $\Delta\notin\mathsf{Good}_h$ is determined by A_Δ. By Lemma 3.4, $b_{h,\Delta}$ is uniform over $\mathbb{F}_p^M$ even conditioned on $A_{h,\Delta}$. If $\mathrm{rank}(A_{h,\Delta}) = r$, then the system $A_{h,\Delta}x = -b_{h,\Delta}$ has p^{N-r} solutions with probability p^{r-M}, and has no solutions with probability $1-p^{r-M}$. Hence,

$$\mathop{\mathbf{E}}_h[|\mathsf{Sol}_h(\Delta)|\mid A_{h,\Delta}=A] = p^{N-r}p^{r-M} = p^{N-M} = \mu$$

for every fixed matrix A. It follows that for every $\Delta\in\mathbb{F}_p^N\setminus\{0\}$,

$$\mathop{\mathbf{E}}_h[\mathbf{1}[\Delta\notin\mathsf{Good}_h]\cdot|\mathsf{Sol}_h(\Delta)|] = \mu\mathop{\mathbf{Pr}}_h[\Delta\notin\mathsf{Good}_h] = \mu\cdot p^{-(\sigma+1)}.$$

Summing over p^N-1 nonzero differences yields

$$\mathop{\mathbf{E}}_h\left[\sum_{\Delta\in\mathbb{F}_p^N\setminus\{0\}:\Delta\notin\mathsf{Good}_h}|\mathsf{Sol}_h(\Delta)|\right] = \mathop{\mathbf{E}}_h\left[\sum_{\Delta\in\mathbb{F}_p^N\setminus\{0\}}\mathbf{1}[\Delta\notin\mathsf{Good}_h]\cdot|\mathsf{Sol}_h(\Delta)|\right] = (p^N-1)\cdot\mu p^{-(\sigma+1)}.$$

Combining this with the lower bound on the denominator (Equation (4)) gives

$$\begin{aligned}\mathop{\mathbf{Pr}}_{h\leftarrow\mathcal{Q}_{p,N,M},(x,y)\leftarrow\mathsf{UCol}_h}[y-x\notin\mathsf{Good}_h] &= \mathop{\mathbf{E}}_h\left[\mathop{\mathbf{Pr}}_{(x,y)\leftarrow\mathsf{UCol}_h}[y-x\notin\mathsf{Good}_h]\right]\\ &\leq \frac{(p^N-1)\cdot\mu p^{-(\sigma+1)}}{p^N(\mu-1)}\\ &\leq \frac{\mu}{\mu-1}p^{-(\sigma+1)}\\ &\leq 2p^{-(\sigma+1)}.\end{aligned}$$

□

□

4 From Ideal Collisions to Uniform Nontrivial Collisions

This section compares the ideal distribution Col_h with UCol_h. We shall prove that

Lemma 4.1. *For $h \leftarrow \mathcal{Q}_{p,N,M}$, it holds that* $\mathsf{SD}((h, \mathsf{Col}_h)\ ;\ (h, \mathsf{UCol}_h) \leq \frac{1}{2\sqrt{\mu}} + \frac{1}{\mu}$.

The crux is that since $\mathcal{Q}_{p,N,M}$ is a pairwise independent hash function family, a random $h \leftarrow \mathcal{Q}_{p,N,M}$ has some regular property in expectation. Ideally, for a regular function h, one has $\mathsf{Col}_h \equiv \mathsf{AllCol}_h$, and that Col_h and AllCol_h differs only on the diagonal pairs.

4.1 Regularity and Normalized Fiber Variance

We define a quantity $\nu(h)$ that captures how regular h is. Recall that $F_{h,y} \stackrel{\mathsf{def}}{=} |h^{-1}(y)|$.

Definition 4.2 (Normalized fiber variance)**.** For a function $h : \mathbb{F}_p^N \to \mathbb{F}_p^M$, define

$$\nu(h) = \frac{1}{p^M} \sum_{y \in \mathbb{F}_p^M} \left(\frac{F_{h,y}}{\mu} - 1 \right)^2 .$$

The next lemma relates $\nu(h)$ and the statistical distance between Col_h and UCol_h.

Lemma 4.3 (From ideal collisions to uniform nontrivial collisions)**.** *For every function $h : \mathbb{F}_p^N \to \mathbb{F}_p^M$, it holds that*

$$\mathsf{SD}(\mathsf{Col}_h, \mathsf{UCol}_h) \leq \frac{1}{2}\sqrt{\nu(h)} + \frac{1}{\mu}. \tag{5}$$

Proof. We first compare Col_h and AllCol_h, and then remove the diagonal pairs to pass from AllCol_h to UCol_h. Both Col_h and AllCol_h are mixtures over fibers and, conditioned on a fiber with value $y \in \mathbb{F}_p^M$, both are uniform over the $F_{h,y}^2$ ordered pairs in that fiber. Thus their statistical distance is the distance between their fiber-index distributions.

Let $R = p^M$, $r_y = F_{h,y}/\mu$, and observe that

$$\frac{1}{R}\sum_y r_y = 1, \qquad \frac{1}{R}\sum_y r_y^2 = 1 + \nu(h).$$

The distribution Col_h chooses fiber y with probability r_y/R. The distribution AllCol_h chooses fiber y with probability $r_y^2/(R(1+\nu(h)))$. Therefore

$$\begin{aligned}\mathsf{SD}(\mathsf{Col}_h, \mathsf{AllCol}_h) &= \frac{1}{2} \cdot \frac{1}{R} \sum_y \left| r_y - \frac{r_y^2}{1+\nu(h)} \right| \\ &= \frac{1}{2(1+\nu(h))} \cdot \frac{1}{R} \sum_y r_y \left| 1 + \nu(h) - r_y \right| .\end{aligned}$$

By Cauchy–Schwarz,

$$\begin{aligned}\frac{1}{R}\sum_y r_y \left| 1 + \nu(h) - r_y \right| &\leq \sqrt{\frac{1}{R}\sum_y r_y^2} \cdot \sqrt{\frac{1}{R}\sum_y (1+\nu(h)-r_y)^2} \\ &= \sqrt{1+\nu(h)} \cdot \sqrt{\nu(h)(1+\nu(h))} = (1+\nu(h))\sqrt{\nu(h)}.\end{aligned}$$

Hence $\mathsf{SD}(\mathsf{Col}_h, \mathsf{AllCol}_h) \le \frac{1}{2}\sqrt{\nu(h)}$.

It remains to remove diagonal pairs. The diagonal mass of AllCol_h is

$$\Pr_{(x,x')\leftarrow \mathsf{AllCol}_h}\left[x = x'\right] = \frac{p^N}{\sum_y F_{h,y}^2}.$$

Since $\sum_y F_{h,y} = p^N$, we have

$$\sum_y F_{h,y}^2 \ge p^M \left(\frac{p^N}{p^M}\right)^2 = p^N \mu.$$

Therefore, the diagonal mass is at most $1/\mu$. Conditioning AllCol_h on the complement of the diagonal gives UCol_h, and this conditioning changes statistical distance by $\Pr_{(x,x')\leftarrow \mathsf{AllCol}_h}[x = x']$. Therefore,

$$\mathsf{SD}(\mathsf{AllCol}_h, \mathsf{UCol}_h) \le \frac{1}{\mu}.$$

Then Equation (5) follows from the triangle inequality. □

4.2 The Case of $\mathcal{Q}_{p,M,N}$

Next, we show that a random degree-2 function has low fiber variance in expectation.

Lemma 4.4 (Expected fiber variance)**.** $\mathbf{E}_{h\leftarrow \mathcal{Q}_{p,N,M}}[\nu(h)] = \frac{1}{\mu} - \frac{1}{p^N}$.

Proof. We first calculate the expectation and variance of $F_{h,y}$ for every fixed $y \in \mathbb{F}_p^M$, using the pairwise independence of $\mathcal{Q}_{p,M,N}$.

Lemma 4.5 (Pairwise independence)**.** *Let $h \leftarrow\!\!\$\ \mathcal{Q}_{p,N,M}$. For every distinct $x, x' \in \mathbb{F}_p^N$, the pair $(h(x), h(x'))$ is uniform over $(\mathbb{F}_p^M)^2$. Consequently, for every $y \in \mathbb{F}_p^M$,*

$$\mathop{\mathbf{E}}_h[F_{h,y}] = \mu, \qquad \mathop{\mathbf{Var}}_h(F_{h,y}) = \mu - \frac{\mu}{p^M}.$$

Proof. It suffices to prove the claim for one output coordinate. Condition on all quadratic coefficients. The two values $h_i(x), h_i(x')$ are affine functions of (γ_i, ℓ_i), namely

$$(\gamma_i + \ell_i^\top x + c_x, \gamma_i + \ell_i^\top x' + c_{x'}),$$

where $c_x, c_{x'}$ are fixed by the conditioned quadratic part. The linear map

$$(\gamma_i, \ell_i) \mapsto (\gamma_i + \ell_i^\top x, \gamma_i + \ell_i^\top x')$$

has rank two because $(1, x) \ne (1, x')$. Hence, $h_i(x), h_i(x')$ is uniform over $\mathbb{F}_p^2$. Independence across output coordinates gives the claim for all M output coordinates.

For fixed y,

$$F_{h,y} = \sum_{x\in\mathbb{F}_p^N} \mathbf{1}[h(x) = y].$$

Each indicator has expectation p^{-M}, and by the linearity of expectation we have $\mathbf{E}[F_{h,y}] = p^{N-M} = \mu$. Moreover,

$$F_{h,y}^2 = \left(\sum_{x \in \mathbb{F}_p^N} \mathbf{1}[h(x) = y] \right)^2 = \sum_{x,x' \in \mathbb{F}_p^N} \mathbf{1}[h(x) = y] \cdot \mathbf{1}[h(x') = y].$$

Taking expectation, by the pairwise independence, we have

$$\mathbf{E}[F_{h,y}^2] = p^N(p^N - 1)p^{2M} + p^{N-M}$$

and thus

$$\mathbf{Var}(F_{h,y}) = \mathbf{E}[F_{h,y}^2] - (\mathbf{E}[F_{h,y}])^2 = \mu - \frac{\mu}{p^M}.$$

□

By the definition of $\nu(h)$ and Lemma 4.5, we have

$$\mathop{\mathbf{E}}_h[\nu(h)] = \frac{1}{\mu^2 p^M} \sum_{y \in \mathbb{F}_p^M} \mathop{\mathbf{Var}}_h[F_{h,y}] = \frac{\mu - \mu/p^M}{\mu^2} = \frac{1}{\mu} - \frac{1}{p^N}.$$

□

Then the proof of Lemma 4.1 is immediate.

Proof of Lemma 4.1. By Lemma 4.3 and Jensen's inequality,

$$\mathsf{SD}((h, \mathsf{Col}_h)\ ;\ (h, \mathsf{UCol}_h) = \mathop{\mathbf{E}}_h[\mathsf{SD}(\mathsf{Col}_h, \mathsf{UCol}_h)] \le \mathop{\mathbf{E}}_h\left[\frac{1}{2}\sqrt{\nu(h)} + \frac{1}{\mu}\right] \le \frac{1}{2}\sqrt{\mathop{\mathbf{E}}_h[\nu(h)]} + \frac{1}{\mu}$$

Meanwhile, Lemma 4.4 says that $\mathbf{E}_h[\nu(h)] = \frac{1}{\mu} - p^{-N} < \frac{1}{\mu}$. Thus, we conclude that

$$\mathsf{SD}\left((h, \mathsf{Col}_h)\ ;\ (h, \mathsf{UCol}_h)\right) \le \frac{1}{2\sqrt{\mu}} + \frac{1}{\mu}.$$

□

Proving the main theorem. Theorem 1.1 follows from Lemma 3.3 and Lemma 4.1.

References

[AHI+17] Benny Applebaum, Naama Haramaty-Krasne, Yuval Ishai, Eyal Kushilevitz, and Vinod Vaikuntanathan. Low-complexity cryptographic hash functions. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:31. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017. Full version: Cryptology ePrint Archive, Paper 2017/036.

[BDRV18] Itay Berman, Akshay Degwekar, Ron D. Rothblum, and Prashant Nalini Vasudevan. Multi collision resistant hash functions and their applications. In *Advances in Cryptology – EUROCRYPT 2018, Part II*, volume 10821 of *Lecture Notes in Computer Science*, pages 133–161. Springer, 2018. Full version: Cryptology ePrint Archive, Paper 2017/489.

[BHKY19] Nir Bitansky, Iftach Haitner, Ilan Komargodski, and Eylon Yogev. Distributional collision resistance beyond one-way functions. In *Advances in Cryptology – EUROCRYPT 2019, Part III*, volume 11478 of *Lecture Notes in Computer Science*, pages 667–695. Springer, 2019. Also available as Cryptology ePrint Archive, Paper 2019/115.

[BKP18] Nir Bitansky, Yael Tauman Kalai, and Omer Paneth. Multi-collision resistance: A paradigm for keyless hash functions. In *Proceedings of the 50th Annual ACM Symposium on Theory of Computing (STOC 2018)*, pages 671–684. ACM, 2018. Full version: Cryptology ePrint Archive, Paper 2017/488.

[DI06] Bella Dubrov and Yuval Ishai. On the randomness complexity of efficient sampling. In *Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC 2006)*, pages 711–720. ACM, 2006.

[DY08] Jintai Ding and Bo-Yin Yang. Multivariates polynomials for hashing. In *Post-Quantum Cryptography, Second International Workshop (PQCrypto 2008)*, volume 5299 of *Lecture Notes in Computer Science*, pages 358–371. Springer, 2008.

[KNY18] Ilan Komargodski, Moni Naor, and Eylon Yogev. Collision resistant hashing for paranoids: Dealing with multiple collisions. In *Advances in Cryptology – EUROCRYPT 2018, Part II*, volume 10821 of *Lecture Notes in Computer Science*, pages 162–194. Springer, 2018. Full version: Cryptology ePrint Archive, Paper 2017/486.

[KY18] Ilan Komargodski and Eylon Yogev. On distributional collision resistant hashing. In *Advances in Cryptology – CRYPTO 2018, Part II*, volume 10992 of *Lecture Notes in Computer Science*, pages 303–327. Springer, 2018.

[MZ25] Xinyu Mao and Jiapeng Zhang. Black-box separation between multi-collision resistance and collision resistance. *Cryptology ePrint Archive*, 2025.

[RV24] Ron D. Rothblum and Prashant Nalini Vasudevan. Collision resistance from multi-collision resistance. *Journal of Cryptology*, 37:14, 2024. Preliminary version in CRYPTO 2022; Cryptology ePrint Archive, Paper 2022/173.

[Zha21] Mark Zhandry. Quantum lightning never strikes the same state twice. or: Quantum money from cryptographic assumptions. *Journal of Cryptology*, 34(1):6, 2021. Preliminary version in EUROCRYPT 2019; Cryptology ePrint Archive, Paper 2017/1080.