On the Number of Restricted Solutions to Constrained Systems and their Applications

Benoît Cogliati[1], Jordan Ethan[2], Ashwin Jha[3], Mridul Nandi[4], and Abishanka Saha[4]

[1]Thales DIS France SAS, Meudon, France
benoit.cogliati@gmail.com
[2]CISPA Helmholtz Center for Information Security, Saarbrücken, Germany
jordan.ethan@cispa.de
[3]Ruhr-Universität Bochum, Bochum, Germany
letterstoashwin@gmail.com
[4]Indian Statistical Institute, Kolkata, India
mridul.nandi@gmail.com,sahaa.1993@gmail.com

Abstract. In this paper, we formulate a special class of systems of linear equations over finite fields and derive lower bounds on the number of solutions adhering to some predefined restrictions. We then demonstrate the applications of these lower bounds to derive tight PRF security (up to $2^{3n/4}$ queries) for single-keyed variants of the Double-block Hash-then-Sum (`DBHtS`) paradigm, specifically `PMAC+` and `LightMAC+`. Additionally, we show that the sum of r independent copies of the Even-Mansour cipher is a secure PRF up to $2^{\frac{r}{r+1}n}$ queries.

Keywords: `PMAC+`, `LightMAC+`, Sum of Even-Mansour, tight security

1 Introduction

For some $k \geq 2$, let $\Pi_1, \ldots, \Pi_k$ denote k mutually independent and uniform random permutations of $\{0,1\}^n$, and consider the function $\mathtt{F} : \{0,1\}^n \rightarrow \{0,1\}^n$ defined by the mapping

$$\mathtt{F}(x) := \Pi_1(x) \oplus \Pi_2(x) \oplus \ldots \oplus \Pi_k(x)$$

It is well-known [19,21] that $\mathtt{F}$ — the well-known sum of k permutations — is statistically indistinguishable from a length-preserving uniform random function, provided the permutations are secret and the number of queried points $q \leq 2^{n-1}$. Over the years several proof techniques [4,29,21,17,20,16,19] have been employed to prove this result, with varied degree of success. In particular, Patarin's mirror theory [33,34], has been the main tool to study the underlying combinatorial problem.

Suppose $k = 2$ and the adversary makes q queries to the oracle at hand. Let $Y_1^i := \Pi_1(x_i)$, $Y_2^i := \Pi_2(x_i)$, and λ_i denote the oracle output, for any $1 \leq i \leq q$. A typical mirror theory based proof studies the system of equations $\{Y_1^i \oplus Y_2^i = \lambda_i\}$ and aims to count all solutions $(y_1^i, y_2^i : i \in [q])$, such that $y_b^i \neq y_b^j$ for all $i \neq j$. In a more general setting, one can study a system of bivariate equations, endowed with a partition of the set of variables, such that any two variables in the same partition must be assigned distinct values. We call this structure, a *constrained system.* It is not difficult to see that for random outputs, the expected number of solutions is $(2^n)_q \times (2^n)_q / 2^{nq}$, where $(2^n)_q$. Dutta et al. and Cogliati and Patarin studied [20,16] the problem specific to the sum of permutations and showed a lower bound close to the expectation while $q \leq 2^n/24$ and the solution space is $\{0,1\}^{2n}$, and as a result a good bound on the advantage. While this approach works when the permutations are secret, it does not apply directly when the adversary has oracle access to the permutations.

This is, for instance, the case with the sum of Even-Mansour or SOEM construction [14] defined by the mapping

$$\mathsf{F}(x) \coloneqq \Pi_1(x \oplus K_1) \oplus \Pi_2(x \oplus K_2),$$

where (K_1, K_2) denotes the key. Since the adversary can now make primitive queries, certain solutions are forbidden for fresh permutation inputs for any construction query. More specifically, if $\mathcal{P}_1$ and $\mathcal{P}_2$ denote the set of primitive query outputs, then the solution space is restricted to $\{0,1\}^{2n} \smallsetminus \overline{\mathcal{P}_1 \times \mathcal{P}_2}$. As it turns out, the existing mirror theory approaches cannot be extended directly in this general setting. In fact, the best lower bounds [18,14,26] show that number of solutions are just $(1 - O(q^3/2^{2n}))$-close to the expectation, provided $q \leq 2^{2n/3}$.

A similar situation also arises in the secret permutations regime. For instance, all single-keyed attempts at DbHtS-based MACs: 1k-PMAC+, 1kf9, 1k-LightMAC+ and n1kf9 are shown to be secure up to $2^{2n/3}$ queries. The main bottleneck: a (possibly) sub-optimal lower bound for the number of solutions for the underlying constrained system. This motivates us to study the aforementioned combinatorial problem in its full generality.

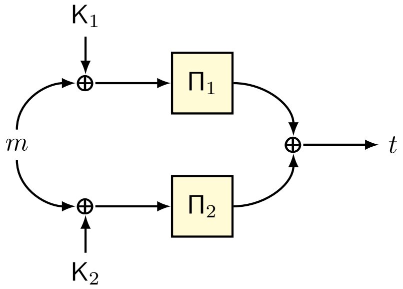
1.1 Related Works

Single-keyed DbHtS MAC. Most common constructions of MAC are either based on block ciphers, e.g., CBC-MAC [5], PMAC [10], OMAC [24], LightMAC [30], etc., or based on a cryptographic hash functions, e.g., HMAC [3]. At a high level these constructions come under the umbrella of UHF-then-PRF designs, where first a message is compressed to a short string by a universal hash function (UHF) and then a PRF is applied on this string to generate the tag. However, due to the detectable collision property, that any collision among the outputs of the UHF results in a tag collision, this design paradigm cannot overcome the birthday bound. This becomes a problem when many MAC constructions have been proposed with lightweight block ciphers, e.g., PRESENT [13], LED [22], GIFT [2].

To go beyond birthday bound, one possible way to improve upon the UHF-then-PRF design, is to replace the UHF by a hash function with double block-output, such that each block behaves like the output of a UHF and then apply the sum-of-permutations PRF on the blocks, i.e., passing each block through a block cipher, and the resulting pair of outputs being xored to get the tag. Such a design idea is bolstered by the fact that the XOR constructions are optimally secure. Dutta et al. [18] formalized this, naming the design *diblock hash-then-sum* (`DBHtS`). In this paper they proved that several constructions falling under the `DBHtS` design paradigm, e.g., `PolyMAC` [11,9,36], `SUM-ECBC` [37], `PMAC+` [38], `LightMAC+` [32] achieve $2n/3$-bit security. In [28], Leurent et al. presented a $3n/4$-bit attack against `DBHtS` schemes. Finally, Kim et al. [27] proved the $3n/4$-bit security of the above constructions, closing the gap.

Sum of Even-Mansour. All the PRF designs discussed till now are block cipher-based. Since we are designing functions, only the forward direction matters, and that is why using block ciphers for PRF constructions, seems superfluous to a degree, because block ciphers have the extraneous property of being efficient in the backward direction too. Instead we could instantiate PRFs based on *public random permutations*, e.g., `Keccac` [8], `Gimli` [6], `SPONGENT` [12], etc., which are designed to be very fast in the forward direction, but not necessarily in the backward direction. Public random permutation based constructions like keyed sponge [1,31], `Farfalle` [7], are variable length constructions. There is a scope of a more efficient/secure design for short fixed-length messages.

In [14] Chen et al. proposed the public random permutation-based PRF construction, called the *sum of Even-Mansour* ($\mathtt{SOEM}^2$), where the idea is to instantiate the block ciphers in the sum of permutations PRF construction, with the public-permutation based block cipher $\mathtt{EM}^{\Pi}(\mathsf{K}, m) = \Pi(\mathsf{K} \oplus m) \oplus \mathsf{K}$. Chen et al. showed that the sum of two Even Mansour constructions, $\mathtt{SOEM}^2_{\Pi_1,\Pi_2}(\mathsf{K}_1, \mathsf{K}_2, m) = \mathtt{EM}^{\Pi_1}(\mathsf{K}_1, m) \oplus \mathtt{EM}^{\Pi_2}(\mathsf{K}_2, m)$ is a $2n/3$-bit secure PRF only if Π_1 is independent of Π_2 and K_1 is independent of K_2. Any weaker assumption would restrict the security to birthday-bound. In [35], Sibleyras et al. showed that post-adding the keys as in Even-Mansour is redundant, achieving the same security with a more efficient design, *keyed sum of permutations*, $\mathsf{KSoP}_{\Pi_1,\Pi_2}(\mathsf{K}_1, \mathsf{K}_2, m) = \Pi_1(\mathsf{K}_1 \oplus m) \oplus \Pi_2(\mathsf{K}_2 \oplus m)$. The authors point out that the independence requirements between Π_1, Π_2 and $\mathsf{K}_1, \mathsf{K}_2$, remain same, in order to achieve said security.



1.2 Our Contributions

Our Contributions are threefold:

- In section 3, we formalize and study the general constrained systems problem over an arbitrary finite field. In section 4, we derive a lower bound on the number of solutions for a large class of constrained systems that encompasses all the known instances in literature.
- As an application we prove tight security bounds for several class of constructions:

  Tight bounds for single-keyed DbHtS: There remains one aspect where the DBHtS schemes can be made yet more efficient. In the general implementations of DBHtS, three keys are used, one for all the block cipher-calls corresponding to hash value evaluations, and one for each of the block ciphers constituting the sum-of-permutations PRF. Since rekeying is an expensive process, the obvious alternative is to use the same key for all the block ciphers, whether it be a part of the hash or the PRF, the design being called the 1k-DBHtS. In section 5, we prove that the single-keyed variant of 1k-PMAC+ and 1k-LightMAC+ achieve security up to $2^{3n/4}$ queries. In particular, in section 6, we show that the corresponding hash functions PHash and LightHash are diblock hash functions having the desired properties.
  - *The Sum of Even-Mansour:* In section 7, for $r \geq 2$, we define the sum of r Even-Mansour ciphers — an extension of the Sibleyras-Todo [35] variant of the sum of two Even-Mansour construction [14] by Chen et al. We show that this construction achieves security up to $2^{\frac{r}{r+1}n}$ queries, which can be shown to be tight by a simple key recovery argument. This directly generalizes the previous results, both in terms of design and security.

2 Preliminaries

For any prime power N, $\mathbb{F}_N$ denotes the finite field of order N. With a slight abuse of notation, we use $\oplus$ and $\cdot$ to denote the addition and multiplication operations in any finite field. For $m, n \in \mathbb{N}^+$, $\mathbb{F}_N^m$ and $\mathbb{F}_N^{m\times n}$ denote the m-dimensional vector space and the set of all $(m \times n)$-matrices over $\mathbb{F}_N$, respectively. For any $\boldsymbol{v} \in \mathbb{F}_N^m$, $\mathtt{H}(\boldsymbol{v})$ denotes the number of non-zero coordinates in $\boldsymbol{v}$.

For any $n \in \mathbb{N}^+$, we identify $\mathbb{F}_{2^n}$ with $\{0,1\}^n$, the set of all n-bit strings. We write $\{0,1\}^* := \cup_{n=0}^{\infty}\{0,1\}^n$. For any $k \leq n \in \mathbb{N}^+$, $(n)_k := n(n-1)\ldots(n-k+1)$ denotes the falling factorial, and $(n)_0 = 1$ by convention.

Sum Capture: For some $k \geq 2$, let $\boldsymbol{\alpha} \in \mathbb{F}_N^k$ and $\mathcal{A}, \mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k \subseteq \mathbb{F}_N$ such that $\mathtt{H}(\boldsymbol{\alpha}) = k$. Define

$$\mathcal{SC}_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}) := \left\{ b \in \mathcal{B}_1 \times \ldots \times \mathcal{B}_k : \bigoplus_{i=1}^{k} \alpha_i \cdot b_i \in \mathcal{A} \right\}, \tag{1}$$

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}) := \left|\mathcal{SC}_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B})\right|, \tag{2}$$

where $\mathcal{B} = (\mathcal{B}_{i_1}, \ldots, \mathcal{B}_{i_k})$ denotes an arbitrary ordering of the constituent sets. For $\alpha = (1, 1, \ldots, 1)$, $\mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}_{i_1}, \mathcal{B}_{i_2}, \ldots, \mathcal{B}_{i_k}) = \mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}_{j_1}, \mathcal{B}_{j_2}, \ldots, \mathcal{B}_{j_k})$ for any

two permutations $(i_1\, i_2\, \ldots\, i_k)$ and $(j_1\, j_2\, \ldots\, j_k)$ of $[k]$. We drop the mask from notation whenever $\boldsymbol{\alpha} = (1, 1, \ldots, 1)$.
For any $k \geq 2$ and $p \geq 0$, we define

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, p) := \max_{\substack{\mathcal{B}_1, \ldots, \mathcal{B}_k \subseteq \mathbb{F}_N \\ |\mathcal{B}_i| \leq p}} \mu_{\boldsymbol{\alpha}}(\mathcal{A}, \mathcal{B}), \tag{3}$$

The following lemma is a restatement of [25, Theorem 1].

Lemma 1. *For all but an $O(N^{-1})$ fraction of (multi)sets $\mathcal{A} \subseteq \mathbb{F}_N$ such that $|\mathcal{A}| = q$ and any $\boldsymbol{\alpha} \in \mathbb{F}_N^k$ with $H(\boldsymbol{\alpha}) = k$, we have*

$$\mu_{\boldsymbol{\alpha}}(\mathcal{A}, p) \leq \left(\frac{qp^k}{N} + 4p^{k-1}\sqrt{\ln(N)q} \right).$$

Proposition 1. *For any real-valued random variable* $\mathtt{X}$, *we have*

$$\mathbb{E}\left(|\mathtt{X} - \mathbb{E}\left(\mathtt{X}\right)|\right) \leq \sqrt{\mathbb{V}\left(\mathtt{X}\right)}.$$

Proof. We have

$$\begin{aligned} \mathbb{E}\left(|\mathtt{X} - \mathbb{E}\left(\mathtt{X}\right)|\right) &= \sqrt{\mathbb{E}\left(|\mathtt{X} - \mathbb{E}\left(\mathtt{X}\right)|\right)^2} \\ &\leq \sqrt{\mathbb{E}\left((\mathtt{X} - \mathbb{E}\left(\mathtt{X}\right))^2\right)} = \sqrt{\mathbb{V}\left(\mathtt{X}\right)}, \end{aligned}$$

where the inequality also follows from Jensen's inequality among others. □

2.1 Hash Functions

A $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$-keyed function H is the function family $\{H_K : \{0,1\}^* \to \mathcal{Y}\}_{K \in \mathcal{K}}$.
We often call H a *diblock hash function*, if we can write $\mathcal{Y}$ as $\mathcal{Z}^2$ for some $\mathcal{Z}$. For any diblock hash function H, we write $(H_K^1(m), H_K^2(m)) := (z_1, z_2)$, where $z_1, z_2 \in \mathcal{Z}$, whenever $H_K(m) = y = (z_1, z_2)$.

Permutation-based Hash Functions. A $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$-hash function is said to be permutation-based if $\mathcal{K} \subseteq \mathcal{P}\left(n\right)^r$ for some $r \in \mathbb{N}$. For any such hash function H, the *block function*, $\beta_H : \mathcal{P}\left(n\right) \times \{0,1\}^* \to \mathbb{N}$, is defined by the mapping:

$$(\pi^r, m) \mapsto \beta_{(\pi^r, m)},$$

where $\pi^r = (\pi_1, \ldots, \pi_r)$ and $\beta_{(\pi^r, m)}$ denotes the minimum number of invocations[1] of π needed to compute $H_\pi(m)$.
In this paper, we fix $r = 1$, and make the following two plausible assumptions on β_H:

[1] Note that, there exists a circuit for H such that on every input, H makes (possibly) a large but bounded number of black-box calls to π^r. Thus, $\beta_{\pi^r, m}$ is well-defined for any π^r and m.

1. β_H is functionally independent of the permutation, whence we drop the permutation from the parameters.
2. there exists a constant $c \in \mathbb{R}^+$ such that for any $m \in \{0,1\}^*$, $\beta_H(m) := c\lceil |m|/n \rceil$. We refer to such an H a *rate-c^{-1}* hash function.

Note that, 1 follows from 2. We state it explicitly for brevity.

We remark that the underlying hash functions in almost all the popular constructions, including `LightMAC`, `PMAC`, `LightMAC+`, `PMAC+`, `3kf9` etc. are rate-1, and `SUM-ECBC` is rate-2^{-1}. Thus, the above assumption is indeed plausible, and $c \leq 2$ in most applications.

Coverfree Hash Functions. For any $(\mathcal{K}, \{0,1\}^*, \mathcal{Y}^2)$-diblock hash function H, any $r \geq 3$, $s \geq 2$, and any $\boldsymbol{m} := (m_1, \ldots, m_q) \in (\{0,1\}^*)_q$, we define the following events

$$\begin{aligned}
\mathsf{COLL1}_H(\boldsymbol{m})&: \ \exists^* i, j \in [q] \text{ such that } H^1_{\mathtt{K}}(m_i) = H^1_{\mathtt{K}}(m_j);\\
\mathsf{COLL2}_H(\boldsymbol{m})&: \ \exists^* i, j \in [q] \text{ such that } H^2_{\mathtt{K}}(m_i) = H^2_{\mathtt{K}}(m_j);\\
\mathsf{AP1}^r_H(\boldsymbol{m})&: \ \exists^* i_1, \ldots, i_r \in [q] \text{ such that}\\
&\quad H^1_{\mathtt{K}}(m_{i_1}) = H^1_{\mathtt{K}}(m_{i_2}), H^2_{\mathtt{K}}(m_{i_2}) = H^2_{\mathtt{K}}(m_{i_3}), \ldots, H^1_{\mathtt{K}}(m_{i_{r-1}}) = H^1_{\mathtt{K}}(m_{i_r});\\
\mathsf{AP2}^r_H(\boldsymbol{m})&: \ \exists^* i_1, \ldots, i_r \in [q] \text{ such that}\\
&\quad H^2_{\mathtt{K}}(m_{i_1}) = H^2_{\mathtt{K}}(m_{i_2}), H^1_{\mathtt{K}}(m_{i_2}) = H^1_{\mathtt{K}}(m_{i_3}), \ldots, H^2_{\mathtt{K}}(m_{i_{r-1}}) = H^2_{\mathtt{K}}(m_{i_r});\\
\mathsf{MC1}^s_H(\boldsymbol{m})&: \ \exists^* i_1, \ldots, i_s \in [q] \text{ such that}\\
&\quad H^1_{\mathtt{K}}(m_{i_1}) = H^1_{\mathtt{K}}(m_{i_2}) = \cdots = H^1_{\mathtt{K}}(m_{i_s});\\
\mathsf{MC2}^s_H(\boldsymbol{m})&: \ \exists^* i_1, \ldots, i_s \in [q] \text{ such that}\\
&\quad H^2_{\mathtt{K}}(m_{i_1}) = H^2_{\mathtt{K}}(m_{i_2}) = \cdots = H^2_{\mathtt{K}}(m_{i_s}),\\
\mathsf{COLL}_H(\boldsymbol{m})&: \ \exists^* i, j \in [q] \text{ such that } H_{\mathtt{K}}(m_i) = H_{\mathtt{K}}(m_j).
\end{aligned}$$

where the randomness is induced by $\mathtt{K} \twoheadleftarrow \mathcal{K}$.

Definition 1. *For some $\epsilon_1, \delta : \mathbb{N}^3 \to [0,1]$ and $\epsilon_2, \epsilon_3 : \mathbb{N}^4 \to [0,1]$, a $(\mathcal{K}, \{0,1\}^*, \mathcal{Y})$-diblock hash function H is said to be an $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-Coverfree Hash or CfH if and only if for any $\rho = (q, \ell, \sigma) \in \mathbb{N}^3$, any $\boldsymbol{m} = (m_1, \ldots, m_q) \in (\{0,1\}^{n\ell})_q$ containing at most σ blocks, any $r \geq 3$, and any $s \geq 2$, it satisfies*

$$\Pr(\mathsf{COLL1}_H(\boldsymbol{m})) \leq \epsilon_1(\rho), \quad \Pr(\mathsf{AP1}^r_H(\boldsymbol{m})) \leq \epsilon_2(\rho, r), \quad \Pr(\mathsf{MC1}^s_H(\boldsymbol{m})) \leq \epsilon_3(\rho, s),$$

$$\Pr(\mathsf{COLL2}_H(\boldsymbol{m})) \leq \epsilon_1(\rho), \quad \Pr(\mathsf{AP2}^r_H(\boldsymbol{m})) \leq \epsilon_2(\rho, r), \quad \Pr(\mathsf{MC2}^s_H(\boldsymbol{m})) \leq \epsilon_3(\rho, s),$$

and $\Pr(\mathsf{COLL}_H(\boldsymbol{m})) \leq \delta(\rho)$.

Double-block Hash-then-Sum. Let H be a $(\mathcal{K}, \{0,1\}^*, \{0,1\}^{2n})$-diblock hash function. The *DiBlock Hash-then-Sum* construction is a $(\mathcal{K} \times \mathcal{P}(n)^2, \{0,1\}^*, \{0,1\}^n)$-keyed function $\mathtt{DBHtS}_H$ defined by the mapping:

$$(K, \pi_1, \pi_2, m) \mapsto \pi_1(H^1_K(m)) \oplus \pi_2(H^2_K(m)) \tag{4}$$

Several beyond-the-birthday bound MAC constructions, including `SUM-ECBC` [37], `PMAC+` [38], `LightMAC+` [32] etc. follow this paradigm.

2.2 Security Definitions

In this paper, we assume that the distinguisher is non-trivial, i.e. it never makes a duplicate query, and it never makes a query for which the response is already known due to some previous query. Let $\mathbb{A}(q, \ell, \sigma, t)$ be the class of all non-trivial distinguishers limited to q oracle queries of each of length up to ℓ blocks and a total of σ blocks, and t computations. Any $\mathcal{A} \in \mathbb{A}(q, \ell, \sigma, t)$ is referred as a (q, ℓ, σ, t)-adversary.

In our analyses, especially security proofs, it will be convenient to work in the information-theoretic setting. Accordingly, we always skip the boilerplate hybrid steps and often assume that the adversary is computationally unbounded, i.e., $t = \infty$, and deterministic. A computational equivalent of all our security proofs can be easily obtained by a simple hybrid argument.

The advantage of any adversary $\mathcal{A}$ in distinguishing some oracle $\mathcal{O}_1$ from another oracle $\mathcal{O}_0$ is defined as

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) := \left|\Pr\left(\mathcal{A}^{\mathcal{O}_1} = 1\right) - \Pr\left(\mathcal{A}^{\mathcal{O}_0} = 1\right)\right|.$$

PRF Security: The PRF advantage of distinguisher $\mathcal{A}$ against a $(\mathcal{K}, \mathcal{X}, \mathcal{Y})$-keyed function $\mathtt{F}$ instantiated with a key $\mathtt{K} \twoheadleftarrow \mathcal{K}$ is defined as

$$\mathbf{Advt}^{\mathrm{prf}}_{\mathtt{F}}(\mathcal{A}) = \Delta_{\mathtt{F};\Gamma}(\mathcal{A}). \tag{5}$$

In this paper, we also consider the security model where the distinguisher is given oracle access to the internal primitives of the construction. More specifically, suppose $\mathtt{F}$ is constructed on top of k uniform random permutations $\Pi = (\Pi_1, \ldots, \Pi_k)$ of $\{0,1\}^n$, denoted $\mathtt{F}[\Pi]$. Then, the PRF advantage of $\mathcal{A}$ is defined as

$$\mathbf{Advt}^{\mathrm{prf}}_{\mathtt{F}[\Pi]}(\mathcal{A}) = \Delta_{(\mathtt{F}[\Pi],\Pi^{\pm});(\Gamma,\Pi^{\pm})}(\mathcal{A}), \tag{6}$$

where the superscript ± denotes a bidirectional access to Π.

2.3 The Expectation Method

Let $\mathcal{A}$ be a computationally unbounded and deterministic distinguisher that tries to distinguish between two oracles $\mathcal{O}_0$ and $\mathcal{O}_1$ via black box interaction with one of them. We denote the query-response tuple of $\mathcal{A}$'s interaction with its oracle by a transcript ω. This may also include any additional information the oracle chooses to reveal to the distinguisher at the end of the query-response phase of the game. We denote by Θ_{re} (res. Θ_{id}) the random transcript variable when $\mathcal{A}$ interacts with $\mathcal{O}_1$ (res. $\mathcal{O}_0$). The probability of realizing a given transcript ω in the security game with an oracle $\mathcal{O}$ is known as the *interpolation probability* of ω with respect to $\mathcal{O}$. Since $\mathcal{A}$ is deterministic, this probability depends only on the oracle $\mathcal{O}$ and the transcript ω. A transcript ω is said to be *attainable* if $\Pr(\Theta_{\mathrm{id}} = \omega) > 0$.

Lemma 2 (Fine-grained Expectation Method). *Let Ω be the set of all transcripts. For some $\varepsilon_{\text{bad}} \geq 0$ and $\varepsilon_{\text{ratio}} : \Omega \to \mathbb{R}$, suppose there is a set $\Omega_{\text{bad}} \subseteq \Omega$ satisfying the following conditions:*

- $\Pr(\Theta_{\text{id}} \in \Omega_{\text{bad}}) \leq \varepsilon_{\text{bad}}$,
- *$\varepsilon_{\text{ratio}}$ is non-negative on $\Omega_{\text{good}} = \Omega \setminus \Omega_{\text{good}}$,*
- *for any $\omega \in \Omega_{\text{good}}$, ω is attainable and* $\dfrac{\Pr(\Theta_{\text{re}} = \omega)}{\Pr(\Theta_{\text{id}} = \omega)} \geq 1 - \varepsilon_{\text{ratio}}(\omega)$.

Then for any distinguisher $\mathcal{A}$ trying to distinguish between $\mathcal{O}_1$ and $\mathcal{O}_0$, we have the following bound on its distinguishing advantage:

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) \leq \varepsilon_{\text{bad}} + \mathbb{E}_{\Theta_{\text{id}}}(1_{\text{good}}\varepsilon_{\text{ratio}}),$$

where 1_{good} denotes the indicator variable corresponding to Ω_{good}.

The expectation method due to Hoang and Tessaro [23] is a simple corollary of the above result, when $\varepsilon_{\text{ratio}}$ is non-negative over the entire transcript space.

Corollary 1 (Expectation Method). *Suppose there is a non-negative function $\varepsilon_{\text{ratio}} : \Omega \to [0, \infty)$ satisfying the following conditions:*

- $\Pr(\Theta_{\text{id}} \in \Omega_{\text{bad}}) \leq \varepsilon_{\text{bad}}$;
- *For any $\omega \notin \Omega_{\text{bad}}$, ω is attainable and* $\dfrac{\Pr(\Theta_{\text{re}} = \omega)}{\Pr(\Theta_{\text{id}} = \omega)} \geq 1 - \varepsilon_{\text{ratio}}(\omega)$.

Then for any distinguisher $\mathcal{A}$ trying to distinguish between $\mathcal{O}_1$ and $\mathcal{O}_0$, we have the following bound on its distinguishing advantage:

$$\Delta_{\mathcal{O}_1;\mathcal{O}_0}(\mathcal{A}) \leq \varepsilon_{\text{bad}} + \mathbb{E}_{\Theta_{\text{id}}}(\varepsilon_{\text{ratio}}).$$

3 Constrained Systems

System of Linear Equations: Fix some $q, r \leq N$. Any system of q linear equations in r variables, $A\boldsymbol{x} = \boldsymbol{\lambda}$, over $\mathbb{F}_N$ can be compactly represented by the augmented matrix $A|\boldsymbol{\lambda}$, where $A \in \mathbb{F}_N^{q\times r}$ and $\boldsymbol{\lambda} \in \mathbb{F}_N^q$.

System-graph and Components: It would be often convenient to look at a graph-theoretic representation of the system $A|\boldsymbol{\lambda}$. Formally, to any system $A|\boldsymbol{\lambda}$, we associate an undirected, labeled, bipartite graph $G(A|\boldsymbol{\lambda}) = (\text{row}(A|\boldsymbol{\lambda}), \text{col}(A), \mathcal{E})$ where $\text{row}(A|\boldsymbol{\lambda})$ and $\text{col}(A)$ denote the two disjoint sets of vertices, and

$$\mathcal{E} = \{(\{A_{i\bullet}|\lambda_i, A_{\bullet j}\}, A_{ij}) \ : \ (i,j) \in [q] \times [r], A_{i,j} \neq 0^n\}$$

denotes the edge-set. Each edge $e = (\{A_{i\bullet}|\lambda_i, A_{\bullet j}\}, A_{ij}) \in \mathcal{E}$ is often written in a more illustrative notation as $A_{i\bullet}|\lambda_i \overset{A_{ij}}{\text{———}} A_{\bullet j}$ or simply $i^- \text{———} j^|$ whenever convenient, where the superscripts $-$ and $|$ are used to differentiate row and column index, respectively. We call $G(A|\boldsymbol{\lambda})$ a *system-graph.*

In this context, we say that two rows $A_{i\bullet}|\lambda_i$ and $A_{i'\bullet}|\lambda_{i'}$ are *adjacent*, denoted $A_{i\bullet}|\lambda_i \sim A_{i'\bullet}|\lambda_{i'}$, if and only if there exists an $A_{\bullet j} \in \text{col}(A)$ such that $i^- \text{——} j^| \text{——} i'^-$.[2] The relation $\sim$ on $\text{row}(A|\boldsymbol{\lambda})$ is reflexive and symmetric, but not transitive.

[2] Any two rows of a matrix are said to be disjoint, if they do not share a common column index with non-zero entry, and non-disjoint otherwise.

We say that two rows $A_{i\bullet}$ and $A_{j\bullet}$ are *connected*, denoted $A_{i\bullet}|\lambda_i \rightsquigarrow A_{j\bullet}|\lambda_j$, if and only if they are connected in $G(A|\boldsymbol{\lambda})$. $\rightsquigarrow$ is an equivalence relation on $\text{row}(A|\boldsymbol{\lambda})$, effectively partitioning $\text{row}(A|\boldsymbol{\lambda}) = A_1|\boldsymbol{\lambda}_1 \sqcup \cdots \sqcup A_c|\boldsymbol{\lambda}_c$. For each component $A_i|\boldsymbol{\lambda}_i$ of $A|\boldsymbol{\lambda}$, let $\overline{A}_i$ denote the *column-reduced form* of A_i, which is obtained by simply dropping all the zero columns from A_i. Then, it is easy to see that the induced subgraph $G[A_i|\boldsymbol{\lambda}_i, \text{col}(\overline{A}_i)]$ is a component $G(A|\boldsymbol{\lambda})$, and a system-graph in its own right. As a consequence, with a slight abuse of notations, we also write $A_i|\boldsymbol{\lambda}_i$ to denote the $q_i \times (r+1)$ submatrix (also referred as a *component*) of $A|\boldsymbol{\lambda}$ corresponding to the equivalence class $A_i|\boldsymbol{\lambda}_i = \{A_{j_1\bullet}|\lambda_{j_1}, \ldots, A_{j_{q_i}\bullet}|\lambda_{j_{q_i}}\}$, i.e.

$$A_i|\boldsymbol{\lambda}_i = \begin{pmatrix} A_{j_1\bullet}|\lambda_{j_1} \\ \vdots \\ A_{j_{q_i}\bullet}|\lambda_{j_{q_i}} \end{pmatrix},$$

where $\sum_i q_i = q$. Let $r_i := |\text{col}(\overline{A}_i)|$ and $\sum_i r_i = r$. For any $i \in [c]$, we say that $A_i|\boldsymbol{\lambda}_i$ is *isolated* if $q_i = 1$. By extension, $A|\boldsymbol{\lambda}$ is said to be isolated if $A_i|\boldsymbol{\lambda}_i$ is isolated for all $i \in [c]$.

Note that, both $\sim$ and $\rightsquigarrow$ are independent of $\boldsymbol{\lambda}$. Accordingly, we often view them as relations on $\text{row}(A)$.

Definition 2 (Canonical Component Form). *Let $A_1|\boldsymbol{\lambda}_1 \sqcup \ldots \sqcup A_c|\boldsymbol{\lambda}_c$ be the partitioning of* $\text{row}(A|\boldsymbol{\lambda})$ *with respect to $\rightsquigarrow$. The component form (CF) of $A|\boldsymbol{\lambda}$ with respect to an arbitrary ordering $(A_{i_1}|\boldsymbol{\lambda}_{i_1}, \ldots, A_{i_c}|\boldsymbol{\lambda}_{i_c})$ is defined as the block matrix*

$$\text{CF}(A|\boldsymbol{\lambda}) := \begin{pmatrix} \overline{A}_{i_1} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\lambda}_{i_1} \\ \mathbf{0} & \overline{A}_{i_2} & \cdots & \mathbf{0} & \boldsymbol{\lambda}_{i_2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \overline{A}_{i_c} & \boldsymbol{\lambda}_{i_c} \end{pmatrix}$$

$A|\boldsymbol{\lambda}$ can have several component forms. Unless stated otherwise, we always assume that the system $A|\boldsymbol{\lambda}$ is in some component form, for if not, it can be placed in CF by a swapping of rows and columns.

Definition 3 (Acyclic System). *Any system $A|\boldsymbol{\lambda}$ is said to be cyclic if and only if the corresponding system-graph $G(A|\boldsymbol{\lambda})$ is cyclic, and acyclic otherwise.*

The following proposition is a trivial consequence of the acyclic nature of the system-graph.

Proposition 2. *Any acyclic system has full row-rank.*

See Example 1 for a short explanation on the notations and definitions introduced thus far.

Example 1. Consider the following system of 6 equations in 15 variables over $\mathbb{F}_N$:

$$A|\boldsymbol{\lambda} = \left(\begin{array}{ccccccccccccccc|c} \alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 \\ 0 & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \\ \alpha_7 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_8 & \alpha_9 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \\ 0 & 0 & 0 & 0 & \alpha_{10} & 0 & \alpha_{11} & 0 & 0 & 0 & 0 & \alpha_{12} & 0 & 0 & 0 & \lambda_4 \\ 0 & 0 & \alpha_{13} & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{14} & \alpha_{15} & 0 & 0 & 0 & 0 & \lambda_5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{16} & \alpha_{17} & \alpha_{18} & \lambda_6 \end{array}\right)$$

for non-zero $\alpha_1, \ldots, \alpha_{18} \in \mathbb{F}_N$. The corresponding system-graph is illustrated in Figure 1.

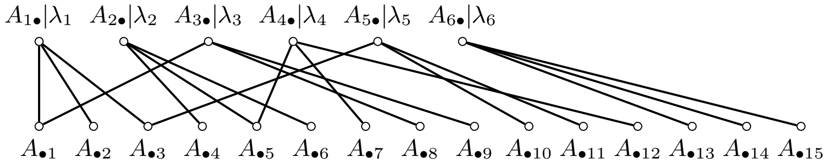


Fig. 1. The system-graph corresponding to the system in Example 1. The edge labels are omitted for readability.

Here,

- $A_{3\bullet}|\lambda_3 \sim A_{1\bullet}|\lambda_1 \sim A_{5\bullet}|\lambda_5$ giving $A_1|\boldsymbol{\lambda}_1 = \{A_{1\bullet}|\lambda_1, A_{3\bullet}|\lambda_3, A_{5\bullet}|\lambda_5\}$,
- $A_{2\bullet}|\lambda_2 \sim A_{4\bullet}|\lambda_4$ giving $A_2|\boldsymbol{\lambda}_2 = \{A_{2\bullet}|\lambda_2, A_{4\bullet}|\lambda_4\}$, and
- $A_{6\bullet}|\lambda_6 \sim A_{6\bullet}|\lambda_6$ giving $A_3|\boldsymbol{\lambda}_3 = \{A_{6\bullet}|\lambda_6\}$,

resulting in the following component form:

$$\left(\begin{array}{ccc|c} \overline{A}_1 & \mathbf{0} & \mathbf{0} & \boldsymbol{\lambda}_1 \\ \mathbf{0} & \overline{A}_2 & \mathbf{0} & \boldsymbol{\lambda}_2 \\ \mathbf{0} & \mathbf{0} & \overline{A}_3 & \boldsymbol{\lambda}_3 \end{array}\right) = \left(\begin{array}{ccccccccccccccc|c} \alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_1 \\ \alpha_7 & 0 & 0 & \alpha_8 & \alpha_9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_3 \\ 0 & 0 & \alpha_{13} & 0 & 0 & \alpha_{14} & \alpha_{15} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_5 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_4 & \alpha_5 & \alpha_6 & 0 & 0 & 0 & 0 & 0 & \lambda_2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{10} & 0 & \alpha_{11} & \alpha_{12} & 0 & 0 & 0 & \lambda_4 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \alpha_{16} & \alpha_{17} & \alpha_{18} & \lambda_6 \end{array}\right)$$

The resulting system $\mathrm{CF}(A|\boldsymbol{\lambda})$ is acyclic and same as $A|\boldsymbol{\lambda}$ up to a relabeling of variables and constants. Furthermore, one of the components $A_3|\boldsymbol{\lambda}_3$ is isolated, although the overall system itself is non-isolated.

Solutions to a System of Equations: Let $\eta(A|\boldsymbol{\lambda})$ denote the number of solutions to the system $A|\boldsymbol{\lambda}$. Throughout we assume that the system is consistent, i.e., $\mathrm{rank}(A|\boldsymbol{\lambda}) = \mathrm{rank}(A)$, otherwise $\eta(A|\boldsymbol{\lambda}) = 0$.

The component form of a system gives a very simple product rule for the number of solutions:

$$\eta(A|\boldsymbol{\lambda}) = \prod_{i=1}^{c} \eta(\overline{A}_i|\boldsymbol{\lambda}_i), \tag{7}$$

which stems from the simple observation that any two components are completely disjoint, i.e., involve distinct variables.

Definition 4 (Constrained System). *For any positive integers* q, r, t *such that* $q, t < r$*, a* (q, r, t)*-constrained system* $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$ *over* $\mathbb{F}_N$ *is the system* $A|\boldsymbol{\lambda}$ *of* q *equations in* r *variables, over* $\mathbb{F}_N$*, endowed with an equivalence relation* P *on* $\mathrm{col}(A)$ *resulting in the partition* $\mathrm{col}(A) = \mathsf{P}_1 \sqcup \ldots \sqcup \mathsf{P}_t$*.*

The dimension and rank of $\mathbb{S}$*, denoted* $\dim(\mathbb{S})$ *and* $\mathrm{rank}(\mathbb{S})$*, are simply the dimension and rank of* A*, respectively.*

For what follows, we fix a (q, r, t)-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$, where $A|\boldsymbol{\lambda}$ is in a component form. Whenever convenient, we drop P from the notation.

Since $\mathbb{S}$ is effectively a system of equations, all the notations and notions are analogously extended unless stated otherwise, except for a minor change in the definition of the system-graph $G(\mathbb{S})$ associated with $\mathbb{S}$ which is now endowed with an implicit coloring of the vertices $\mathrm{col}(A)$ that has a one to one correspondence with P. More precisely, for any $i \in [t]$, any two columns $A_{\bullet j}, A_{\bullet j'} \in \mathsf{P}_i$ share the same implicit color.

The ordered sequence $(\mathbb{S}_1 \prec \cdots \prec \mathbb{S}_c)$ denotes the component form of $\mathbb{S}$, denoted $\mathrm{CF}(\mathbb{S})$, where each $\mathbb{S}_i$ is the (q_i, r_i, t_i)-constrained system $(\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P}^{(i)})$, with $\mathsf{P}^{(i)} \subseteq \mathsf{P}$ being the equivalence relation on the set $\mathrm{col}(\overline{A}_i) \subseteq [r]$, that partitions $\mathrm{col}(\overline{A}_i)$ into t_i subsets $\mathsf{P}_1^{(i)}, \ldots, \mathsf{P}_{t_i}^{(i)}$.

$\mathbb{S}$ is said to be:

- a *clique* iff for all $j, j' \in \mathrm{col}(A)$, $(j, j') \in \mathsf{P}$.
- a *partite* iff for all $A_{i\bullet} \in \mathrm{row}(A)$, and for all $j, j' \in \mathrm{col}(\overline{A}_{i\bullet})$, $(j, j') \notin \mathsf{P}$.

Since $\mathsf{P}^{(i)} \subseteq \mathsf{P}$, for brevity we continue to use P instead of $\mathsf{P}^{(i)}$ for all i. Wlog we also assume that $\mathbb{S}$ is in component form or simply CF.

See Example 2 for an explanation on the notations and definitions related to constrained systems.

Example 2. Recall Example 1, and endow the system $A|\boldsymbol{\lambda}$ with an implicit equivalence relation P (as evident from the updated system-graph illustrated in Figure 2), resulting in the partition $\mathrm{col}(A) = \mathsf{P}_1 \sqcup \mathsf{P}_2 \sqcup \mathsf{P}_3$, where $\mathsf{P}_i = \{j \in [15] : j \equiv i \pmod 3\}$ for all $i \in [3]$.

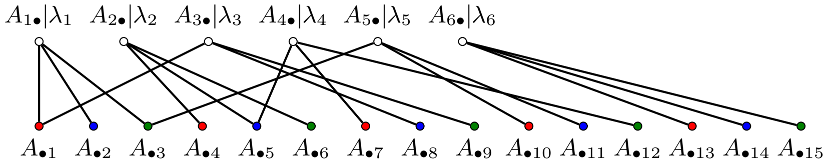


Fig. 2. The system-graph corresponding to the constrained system in Example 2. Yet again the edge labels are omitted for readability.

For the $(6, 15, 3)$-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$, we have

- $\dim(\mathbb{S}) = 6 \times 15$, $\mathrm{rank}(\mathbb{S}) = 6$,
- $\mathrm{CF}(\mathbb{S}) = (\mathbb{S}_1 \prec \mathbb{S}_2 \prec \mathbb{S}_3)$, where $\mathbb{S}_i = (\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P})$,
- $\mathbb{S}_3$ is isolated, but $\mathbb{S}$ is not, and
- $\mathbb{S}$ is acyclic and partite.

4 Solutions to a Constrained System

Definition 5 (Solution to a Constrained System). *For a family of sets $\mathcal{R} = \{\mathcal{R}_i \subseteq \mathbb{F}_N\}_{i\in[t]}$, any $\boldsymbol{y} = (y_1, \ldots, y_r) \in \mathbb{F}_N^r$ is said to be an $\overline{\mathcal{R}}$-solution for $\mathbb{S}$ if and only if the following conditions are satisfied:*

1. *$\boldsymbol{y}$ satisfies the system $A|\boldsymbol{\lambda}$,*
2. *for any $i \in [t]$, and any $j \in \mathsf{P}_i$, $y_j \notin \mathcal{R}_i$,*
3. *for any $i \in [t]$, and any $j \neq j' \in \mathsf{P}_i$, $y_j \neq y_{j'}$.*

In words, all elements in $\mathcal{R}_1, \ldots, \mathcal{R}_t$ are forbidden. In this context, $\mathcal{R}_i$ are referred as *forbidden sets.* Furthermore any two distinct P-related variables[3] must have distinct values.

Let $(\mathbb{S}\,|\,\mathcal{R})$ denote the $\overline{\mathcal{R}}$-solution space of $\mathbb{S}$ and $\eta(\mathbb{S}\,|\,\mathcal{R}) := |(\mathbb{S}\,|\,\mathcal{R})|$, the number of $\overline{\mathcal{R}}$-solutions of $\mathbb{S}$. The central problem that we study in this work is to find a good lower bound on $\eta(\mathbb{S}\,|\,\mathcal{R})$ under some assumptions on A, $\boldsymbol{\lambda}$ and $\mathcal{R}$.

Fix a (q, r, t)-constrained system $\mathbb{S} = (A|\boldsymbol{\lambda}\,;\,\mathsf{P})$ and a family of sets $\mathcal{R} = \{\mathcal{R}_i\}_{i\in[t]}$. Fix a component form $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$ for $\mathbb{S}$. For any $(i, j) \in [c] \times [t]$, let $r_i^{(j)} := \left|\mathrm{col}(\overline{A}_i) \cap \mathsf{P}_j\right|$, and define $r^{(j)} = \sum_{i=1}^{c} r_i^{(j)}$.

Without loss of generality, we assume $|\mathcal{R}_i| = s_i \leq s$ for some $s < N$, or else, $(\mathbb{S}\,|\,\mathcal{R}) = \varnothing$. Then, under the assumption that $\boldsymbol{\lambda}$ is uniform at random, one would expect that the number of $\overline{\mathcal{R}}$-solutions for $\mathbb{S}$ is approximately

$$\mathbb{E}\left(\mathbb{S}\,|\,\mathcal{R}\right) := \frac{\prod_{j=1}^{t}(N - s_i)_{r^{(j)}}}{N^q} \tag{8}$$

Of course, the assumption and the expression are both quite speculative at a first glance. However, as we show later, $\eta(\mathbb{S}\,|\,\mathcal{R})$ is very close to $\mathbb{E}\left(\mathbb{S}\,|\,\mathcal{R}\right)$ for a large class of constrained systems. Indeed, for certain binary matrices A and $\mathcal{R} = \varnothing$ case, Cogliati et al. prove [15] exactly this result. We aim to prove it in a more general setting where $\mathcal{R}$ may not be empty.

While tackling the problem in its full generality is an interesting and technically challenging endeavor, it might not captivate the general cryptography community. Instead, we focus on a specific class of constrained systems that includes, among other things, known instances in symmetric cryptography, particularly those discussed in this paper.

[3] The equivalence relation P on $\mathrm{col}(A)$ can be equivalently defined over the set of variables $\{x_1, \ldots, x_r\}$.

Definition 6 (Weight). *The weight of any* $A \in \mathbb{F}_N^{q\times r}$ *is defined as*

$$\mathtt{H}(A) := \min\{\mathtt{H}(\boldsymbol{v}) : \boldsymbol{v} \in \mathrm{rowsp}^+(A)\},$$

where $\mathrm{rowsp}^+(A) := \{a_1 A_{1\bullet} \oplus \cdots \oplus a_q A_{q\bullet} : \forall (a_1, \ldots, a_q) \neq \mathbf{0}\}$ *and* $\mathtt{H}(\boldsymbol{v})$ *denotes the number of non-zero coordinates in* v.

We have the following fact that relates the weight of a matrix (and its components) with its row rank.

Proposition 3. *Suppose* $A \in \mathbb{F}_N^{q\times r}$ *has* $\mathtt{H}(A) = k > 0$. *Then,*
(1) *A has full row rank.*
(2) *for every* $r' \geq r - k + 1$ *and* $1 \leq i_1 < \cdots < i_{r'} \leq r$, *the matrix* $A' = (A_{\bullet i_1}|\cdots|A_{\bullet i_{r'}})$ *has full row rank, where* $A_{\bullet i}$ *denotes the* i*-th column of* A *viewed as a* q*-dimensional vector.*
(3) $r - k + 1 \geq q$.

Proof. (1) follows from the definition. For (2), suppose to the contrary that A' does not have full rank. Then, we must have $\mathbf{0} \in \mathrm{rowsp}^+(A')$. Specifically, one can find $(a_1, \ldots, a_q) \neq \mathbf{0} \in \mathbb{F}_2^q$, such that $a_1 A'_{1\bullet} \oplus \cdots \oplus a_q A'_{q\bullet} = \mathbf{0}$. Then, $\boldsymbol{v} = a_1 A_{1\bullet} \oplus \cdots \oplus a_q A_{q\bullet} \in \mathrm{rowsp}^+(A)$, and $\mathtt{H}(\boldsymbol{v}) \leq r - r' \leq k - 1$. Thus, $\mathtt{H}(\mathbb{S}) < k$, which is a contradiction. Finally, (3) follows from (2). □

Looking ahead momentarily the higher the weight of a system, the closer our bound to $\mathbb{E}(\mathbb{S}|\mathcal{R})$, and point (2) and (3) of Proposition 3 play a crucial role towards establishing this fact. The following definition and subsequent results provide an easy-to-check condition for determining the weight of a matrix.

Definition 7 (Regularity). *Any* $A \in \mathbb{F}_N^{q\times r}$ *is said to be* k*-regular if and only if* $\mathtt{H}(A_{i\bullet}) = k$, *for all* $i \in [q]$.

Note that, the above definition can be equivalently formulated as $\mathrm{row}(A|\boldsymbol{\lambda})$ is regular[4] in $G(A|\boldsymbol{\lambda})$. The following propositions show that acyclic and highly regular systems have high weight.

Proposition 4. *For any* $k \geq 2$, *any* k*-regular and acyclic* $A \in \mathbb{F}_N^{q\times r}$ *has* $\mathtt{H}(A) = k$.

Proof. The result is trivial for $q = 1$. Assume for contradiction that $\mathtt{H}(A) < k$ for some $q \geq 2$. Then, for some $2 \leq l \leq q$, there exists a sequence of rows $A_{i_1\bullet}, \ldots, A_{i_l\bullet}$ and a sequence of non-zero field elements $a_1, \ldots, a_l$, such that $\boldsymbol{v} = a_1 A_{i_1\bullet} \oplus \ldots \oplus a_l A_{i_l\bullet}$ has $H(\boldsymbol{v}) < k$. Since, A is acyclic, one can always find two distinct rows $A_{i_a\bullet}$ and $A_{i_b\bullet}$ such that there exists at most one $A_{i_c\bullet} \in \{A_{i_1\bullet}, \ldots, A_{i_l\bullet}\} \setminus A_{i_a\bullet}$ and one $A_{i_d\bullet} \in \{A_{i_1\bullet}, \ldots, A_{i_l\bullet}\} \setminus A_{i_b\bullet}$ such that $A_{i_a\bullet} \sim A_{i_c\bullet}$ and $A_{i_b\bullet} \sim A_{i_d\bullet}$, respectively. For if not, then due to the finiteness of l, the matrix

$$\begin{pmatrix} A_{i_1\bullet} \\ \vdots \\ A_{i_l\bullet} \end{pmatrix}$$

[4] A vertex set is said to be regular if all the constituent vertices have the same degree.

is cyclic which contradicts the acyclic nature of A. Then, using the k-regularity of A, at least $k-1 \geq 1$ non-zero columns in each of $A_{i_a \bullet}$ and $A_{i_b \bullet}$ have a single non-zero entry. Therefore, these columns contribute non-zero coordinates to $\boldsymbol{v}$. Thus, $H(\boldsymbol{v}) \geq 2k-2$ which is at least k for $k \geq 2$. □

Proposition 5. *For any $q \geq 2$ and any $k \geq 3$, let $A \in \mathbb{F}_N^{q \times r}$ be acyclic and k-regular. Then, for any $1 \leq i_1 < \ldots < i_k \leq r$, the matrix $A' = A \setminus \{A_{\bullet i_1}, \ldots, A_{\bullet i_k}\}$ has:*

$$\operatorname{rank}(A') = \begin{cases} q-1 & \textit{if } \{i_1, \ldots, i_k\} = \operatorname{col}(\overline{A}_{j\bullet}) \textit{ for some } j \in [q], \\ q & \textit{otherwise.} \end{cases}$$

Proof. First consider the case: $\{i_1, \ldots, i_k\} = \operatorname{col}(\overline{A}_{j\bullet})$ for some $j \in [q]$, i.e., all the non-zero columns of $A_{j\bullet}$ are deleted, and hence $A_{j\bullet}$ can be dropped without affecting the rank of A'. Thus, $\operatorname{rank}(A') \leq q-1$. Furthermore, since the system is acyclic and A is k-regular, A' must be acyclic and at least $(k-1)$-regular. Then, using Proposition 4, we have $\mathtt{H}(A') \geq k-1 \geq 2$, and thus using Proposition 3, $\operatorname{rank}(A') = q-1$.

Now suppose $\{i_1, \ldots, i_k\} \neq \operatorname{col}(\overline{A}_{j\bullet})$ for all $j \in [q]$. Thus, A' has q non-zero rows. Assume towards a contradiction that $\operatorname{rank}(A') < q$. Then one can find a sequence of distinct rows $A'_{j_1\bullet}, A'_{j_2\bullet}, \ldots, A'_{j_l\bullet} \in \operatorname{row}(A')$ and a sequence of non-zero coefficients $a_1, a_2, \ldots, a_l$ such that $\boldsymbol{v} = a_1 A'_{j_1\bullet} \oplus \ldots \oplus a_l A'_{j_l\bullet} = \mathbf{0}$. Let

$$A'' = \begin{pmatrix} A_{j_1\bullet} \\ A_{j_2\bullet} \\ \vdots \\ A_{j_l\bullet} \end{pmatrix}$$

We claim that the number of columns in A'' with a single non-zero entry in each of these columns is at least $2k-2$. Indeed, in the worst case, all the rows are connected to each other. So after a relabeling of rows one can find a sequence $A_{j'_1\bullet} \sim A_{j'_2\bullet} \sim \ldots \sim A_{j'_{l'}\bullet}$ for some $l' \leq l$. Since A'' is acyclic and k-regular, $A_{j'_1\bullet}$ and $A_{j'_{l'}\bullet}$ contribute at least $k-1$ columns each with a single non-zero entry. Now, even if one deletes k columns from A'', there are still at least $k-2 \geq 1$ columns that contribute non-zero entries in any linear combination, including $\boldsymbol{v} = a_1 A'_{j_1\bullet} \oplus \ldots \oplus a_l A'_{j_l\bullet}$. Therefore, $v \neq \mathbf{0}$, contradicting $\operatorname{rank}(A') < q$. □

Column-Uniform System: $\mathbb{S} = (A|\boldsymbol{\lambda}; \mathsf{P})$ is said to be a *column-uniform matrix* if for each column j of A, there exists a non-zero scalar α_j such that all non-zero entries in column j are equal to α_j. Formally, for each column j, there exists a non-zero scalar α_j, such that for all row i of A the following condition holds:

$$A_{ij} = \begin{cases} \alpha_j & \text{if } A_{ij} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

In this paper, *we focus on lower bounding $\eta(\mathbb{S}|\mathcal{R})$ for column-uniform, acyclic and k-regular* (or k-CAR) *system* $\mathbb{S} = (A|\boldsymbol{\lambda}; \mathsf{P})$.

Additional Notations and Conventions: Without loss of generality assume a component form $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$, such that all the isolated components appear before the non-isolated ones. Let $\mathtt{NI}(\mathbb{S})$ denote the set of indices of all the non-isolated components, $\xi_{\mathbb{S}} \coloneqq \max\{r_i : i \in [c]\}$, $\Delta_{\mathbb{S}} \coloneqq \max_d |\{i \in [q] : \lambda_i = d\}|$, and for any $i \in [c]$, let:

- $\mathbb{S}_{\leq i}$ denote the system $(\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_i)$,
- $\boldsymbol{y}_{\leq i}$ denote the solution of the sub-system $\mathbb{S}_{\leq i}$,
- $\mathcal{P}$ and $\mathcal{F}$ define families of set indexed by $j \in [t]$ such that

$$\mathcal{P}_j(\boldsymbol{y}_{\leq i}) \coloneqq \{y_k \in \boldsymbol{y}_{\leq i} : k \in \mathsf{P}_j\} \quad \text{and} \quad \mathcal{F}_j(\boldsymbol{y}_{\leq i}) \coloneqq \mathcal{R}_j \sqcup \mathcal{P}_j(\boldsymbol{y}_{\leq i}).$$

Let $|\mathcal{P}_j(\boldsymbol{y}_{\leq i})| \coloneqq r^{(j)}_{\leq i}$ and $|\mathcal{F}_j(\boldsymbol{y}_{\leq i})| = f^{(j)}_{\leq i} \coloneqq s_j + r^{(j)}_{\leq i}$.
Extending the notation for $i = 0$, let $\boldsymbol{y}_{\leq 0}$ denote any empty sequence, and thus, $\mathcal{P}_j(\boldsymbol{y}_{\leq 0}) = \varnothing$ and $\mathcal{F}_j(\boldsymbol{y}_{\leq 0}) = \mathcal{R}_j$. In addition, for the sake of convenience we also assume that $0^n \in \mathcal{R}_j$ for all $j \in [t]$. Note that, $r^{(j)}_{\leq i}$ and hence $f^{(j)}_{\leq i}$ are independent of the actual elements in $\mathcal{P}_j(\boldsymbol{y}_{\leq i})$ and $\mathcal{F}_j(\boldsymbol{y}_{\leq i})$, respectively. In particular, we have $r^{(j)}_{\leq i} \leq q$, as each equation can have at most one variable in P_j, and thus, $f^{(j)}_{\leq i} \leq s_j + q \leq s + q$.

4.1 The Case of CAR Partite System

For any t-CAR and partite (t-CARP) (q, r, t)-system $\mathbb{S}$, there exists a fixed coefficient vector $\boldsymbol{\alpha}_{\mathbb{S}} = (\alpha_1, \ldots, \alpha_t) \in \mathbb{F}_N^t$ common across all equations. Further, we have the obvious bijective map $\alpha_j \mapsto \mathsf{P}_j$. With this in mind, we define three families of sets $\hat{\mathcal{R}}$, $\hat{\mathcal{P}}$ and $\hat{\mathcal{F}}$ indexed by $j \in [t]$ such that

$$\begin{aligned}\hat{\mathcal{R}}_j &\coloneqq \alpha_j \cdot \mathcal{R}_j \\ \hat{\mathcal{P}}_j(\boldsymbol{y}_{\leq i}) &\coloneqq \{\alpha_j \cdot y_k \in \boldsymbol{y}_{\leq i} : k \in \mathsf{P}_j\} \\ \hat{\mathcal{F}}_j(\boldsymbol{y}_{\leq i}) &\coloneqq \hat{\mathcal{R}}_j \sqcup \hat{\mathcal{P}}_j(\boldsymbol{y}_{\leq i}).\end{aligned}$$

It is obvious that $|\hat{\mathcal{R}}_j| = s_j$, $|\hat{\mathcal{P}}_j(\boldsymbol{y}_{\leq i})| = r^{(j)}_{\leq i}$ and $|\hat{\mathcal{F}}_j(\boldsymbol{y}_{\leq i})| = f^{(j)}_{\leq i}$.

Theorem 1 (Partite Bound). *Let $t \geq 2$, and $\mathcal{R}$ be a family of sets. For any (q, r, t)-constrained system $\mathbb{S}$ which is t-CARP and satisfies $\xi_{\mathbb{S}}(s + q) \leq N/2$, we have $\eta(\mathbb{S} \,|\, \mathcal{R}) \geq (1 - \varepsilon)\, \mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R})$, where*

$$\varepsilon \leq \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}, \mathcal{R})}{N^{t-1}} + \frac{2q\Delta_{\mathbb{S}}}{N^{t-1}} + \frac{6q(s+q)^t}{N^t} + \sum_{i \in \mathtt{NI}(\mathbb{S})} \left(\frac{2r_i^t(s+q)^t}{N^t} + \frac{q_i(s+q)^{t-1}}{N^{t-1}} \right).$$

A proof of this result is derived in two stages. First, in Lemma 4, we derive an initial bound that would be useful when the local[5] error terms can be shown to be sufficiently small in expectation for a random constrained system. We then go on to derive a bound on the global error term which completes the proof of the aforementioned theorem.

[5] The adjective "local" here corresponds to individual components.

Consider the i-th component $\mathbb{S}_i = (\overline{A}_i|\boldsymbol{\lambda}_i\,;\,\mathsf{P})$. Since $\mathbb{S}$ is in CF, $\mathrm{col}(\overline{A}_i) = \{r_{\leq(i-1)}+1,\ldots,r_{\leq(i-1)}+t\}$, where $r_{\leq(i-1)} = r_1+\ldots+r_{i-1}$. For brevity, we ignore the $r_{\leq(i-1)}$ shift in indexing.

Now, towards a proof of Theorem 1, observe that

$$\eta(\mathbb{S}_{\leq i}\,|\,\mathcal{R}) = \sum_{\boldsymbol{y}_{\leq(i-1)}} \eta(\mathbb{S}_i\,|\,\mathcal{F}(\boldsymbol{y}_{\leq(i-1)})), \tag{9}$$

For a fixed $\boldsymbol{y}_{\leq(i-1)}$, the set of $\overline{\mathcal{R}}$-solutions to $\mathbb{S}_i$ is given by

$$(\mathbb{S}_i\,|\,\mathcal{F}) := \{y = (y_1,\ldots,y_{r_i}) \in \overline{\mathcal{F}}_{(1)} \times \ldots \times \overline{\mathcal{F}}_{(r_i)} : \overline{A}_i y = \boldsymbol{\lambda}_i\},$$

where, for all $j \in [r_i]$, $\mathcal{F}_{(j)} := \mathcal{F}_k(\boldsymbol{y}^{\leq i-1})$ for a unique $k \in [t]$. Let $f_{(j)} = |\mathcal{F}_{(j)}|$, and thus $f_{(j)} = f^{(k)}_{\leq(i-1)}$ for a unique $k \in [t]$. Let $\mathcal{A}_{\varnothing} := \{y \in \mathbb{F}_N^{r_i} \;:\; \overline{A}_i y = \boldsymbol{\lambda}_i\}$. Moreover, for each $j \in [r_i]$, we define

$$\mathcal{A}_{\{j\}} := \mathcal{A}_{\varnothing} \bigcap (\mathbb{F}_N^{j-1} \times \mathcal{F}_{(j)} \times \mathbb{F}_N^{t-j}).$$

Then, we have

$$(\mathbb{S}_i\,|\,\mathcal{F}) = \mathcal{A}_{\varnothing} \smallsetminus \left(\bigcup_{j\in\mathrm{col}(\overline{A}_i)} \mathcal{A}_{\{j\}} \right).$$

For any non-empty $\mathcal{J} \subseteq \mathrm{col}(\overline{A}_i)$, let $\mathcal{A}_{\mathcal{J}} := \cap_{j\in\mathcal{J}} \mathcal{A}_{\{j\}}$. Using the principal of inclusion-exclusion, we have

$$\begin{aligned} \eta(\mathbb{S}_i\,|\,\mathcal{F}) &= |\mathcal{A}_{\varnothing}| - \left| \left(\bigcup_{j\in[r_i]} \mathcal{A}_{\{j\}} \right) \right| \\ &= \sum_{\mathcal{J}\subseteq[r_i]} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \end{aligned} \tag{10}$$

Now, $|\mathcal{A}_{\varnothing}| = N^{r_i - q_i}$ follows from elementary linear algebra; In fact, by virtue of $\mathbb{S}$ being an acyclic and t-regular system, Proposition 4 and 3 allows for an analogous argument to prevail for any $\mathcal{A}_{\mathcal{J}}$ with $|\mathcal{J}| \leq t-1$. In particular, for any $\mathcal{J} = \{l_1,\ldots,l_{|\mathcal{J}|}\}$, and any $y_{\mathcal{J}} = (y_{l_1},\ldots,y_{l_{|\mathcal{J}|}}) \in \mathcal{F}_{(l_1)} \times \ldots \times \mathcal{F}_{(l_{|\mathcal{J}|})}$, we obtain an equation in exactly $r_i - |\mathcal{J}| \geq r_i - t + 1 \geq q_i$ variables, which has exactly $N^{r_i - |\mathcal{J}| - q_i}$ solutions. There are exactly $f_{(\mathcal{J})} = f_{(l_1)} \cdots f_{(l_{|\mathcal{J}|})}$ such $y_{\mathcal{J}}$. Thus, we have $|\mathcal{A}_{\mathcal{J}}| = f_{(\mathcal{J})} \cdot N^{t-|\mathcal{J}|-q_i}$ for all $\mathcal{J} \subset [t]$.

Crude Bound: Digressing a little, from (10) and the above discussion, we have

$$N^{r_i-q_i} - r_i(s+q)N^{r_i-q_i-1} \leq \eta(\mathbb{S}_i\,|\,\mathcal{F}) \leq N^{r_i-q_i}$$

for any acyclic system $\mathbb{S}$ where we use the fact $f_{(j)} \leq (s+q)$ for all $j \in [r_i]$. This along with (9) gives the following crude bound

$$N^{r_i-q_i} - r_i(s+q)N^{r_i-q_i-1} \leq \frac{\eta(\mathbb{S}_{\leq i}\,|\,\mathcal{R})}{\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})} \leq N^{r_i-q_i} \tag{11}$$

Now coming back to (10) for a proof of Theorem 1, we study the right hand side separately for isolated and non-isolated components, starting with an isolated component.

Lemma 3. *Suppose $\mathbb{S}_i$ is isolated. Then, for any $y_{\le(i-1)} \in (\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R})$, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N} \left(1 - \frac{2}{N^{t-1}} \left| \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f^{([t])}_{\le(i-1)}}{N} \right| \right),$$

where $f^{([t])}_{\le(i-1)} = f^{(1)}_{\le(i-1)} \cdot \dots \cdot f^{(t)}_{\le(i-1)}$.

Proof. Since $\mathbb{S}_i$ is t-regular, partite and isolated, $r_i = t$ and $q_i = 1$. Then, recall from (10) and the subsequent discussion

$$\begin{aligned}
\eta(\mathbb{S}_i \,|\, \mathcal{F}) &= \sum_{\mathcal{J} \subseteq [t]} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \\
&= \sum_{\mathcal{J} \subset [t]} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{t-|\mathcal{J}|-1} + (-1)^t \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) \\
&= \frac{1}{N} \left(\sum_{\mathcal{J} \subset [t]} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{t-|\mathcal{J}|} + f_{([t])} - f_{([t])} + (-1)^t N \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) \right) \\
&= \frac{1}{N} \left(\prod_{j=1}^{t} (N - f_{(j)}) + (-1)^t N \left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f_{([t])}}{N} \right) \right) \\
&= \frac{1}{N} \left(\prod_{j=1}^{t} (N - f^{(j)}_{\le(i-1)}) + (-1)^t N \left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f^{([t])}_{\le(i-1)}}{N} \right) \right) \\
&\ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N} \left(1 - \frac{N}{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})} \left| \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f^{([t])}_{\le(i-1)}}{N} \right| \right) \\
&\ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N} \left(1 - \frac{2}{N^{t-1}} \left| \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{F}) - \frac{f^{([t])}_{\le(i-1)}}{N} \right| \right),
\end{aligned} \tag{12}$$

where the second equality is due to (2), the fifth equality is from a simple relabeling, and the last inequality follows from the fact that $f^{(j)}_{\le(i-1)} \le (s+q)$ and $t(s+q) \le \xi_{\mathbb{S}}(s+q) \le N/2$. □

Now, on to a lower bound on $\eta(\mathbb{S}_{\le i} \,|\, \mathcal{R})$ for isolated $\mathbb{S}_i$.

Lemma 4. *Suppose $\mathbb{S}_i$ is isolated. Then, we have*

$$\eta(\mathbb{S}_{\le i} \,|\, \mathcal{R}) \ge \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\le(i-1)})}{N} \left(1 - \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i, \mathcal{R})}{N^{t-1}} - \frac{2\Delta_{\mathbb{S}}}{N^{t-1}} - \frac{6(s+q)^t}{N^t} \right) \eta(\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R}).$$

Proof. From (9) and Lemma 3, we have

$$\begin{aligned}
\eta(\mathbb{S}_{\leq i}\,|\,\mathcal{R}) &= \sum_{\boldsymbol{y}_{\leq(i-1)}} \eta(\mathbb{S}_i\,|\,\mathcal{F}(\boldsymbol{y}_{\leq(i-1)})) \\
&\geq \sum_{\boldsymbol{y}_{\leq(i-1)}} \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(1 - \frac{2}{N^{t-1}}\left|\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^{([t])}_{\leq(i-1)}}{N}\right|\right) \\
&\geq \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{2f^{([t])}_{\leq(i-1)}}{N^t}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{2}{N^{t-1}}\sum_{\boldsymbol{y}_{\leq(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F})\right) \\
&\geq \frac{\prod_{j=1}^{t}(N - f^{(j)}_{\leq(i-1)})}{N}\left(\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{2f^{([t])}_{\leq(i-1)}}{N^t}\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}) - \frac{2}{N^{t-1}}\sum_{\boldsymbol{y}_{\leq(i-1)}}\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F})\right)
\end{aligned} \tag{13}$$

Claim. We claim

$$\sum_{\boldsymbol{y}_{\leq(i-1)}} \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) \leq \left(\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R}) + \mathit{\Delta}_{\mathbb{S}} + \frac{2(s+q)^t}{N}\right)\eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$$

Proof. We have

$$\begin{aligned}
\sum_{\boldsymbol{y}_{\leq(i-1)}} \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) &= \sum_{\boldsymbol{y}_{\leq(i-1)}}\sum_{\mathcal{I}\subseteq[t]} \mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_{\mathcal{I}},\hat{\mathcal{R}}_{[t]\smallsetminus\mathcal{I}}) \\
&= \sum_{\mathcal{I}\subseteq[t]}\sum_{\boldsymbol{y}_{\leq(i-1)}} \mu(\boldsymbol{\lambda}_i,\hat{\mathcal{P}}_{\mathcal{I}},\hat{\mathcal{R}}_{[t]\smallsetminus\mathcal{I}})
\end{aligned}$$

where $\hat{\mathcal{P}}_{\mathcal{I}} = \hat{\mathcal{P}}_{j_1} \times \ldots \times \hat{\mathcal{P}}_{j_m}$ and $\hat{\mathcal{R}}_{[t]\smallsetminus\mathcal{I}} = \hat{\mathcal{R}}_{k_1} \times \ldots \times \hat{\mathcal{R}}_{k_{m'}}$, for every $\mathcal{I} = \{j_1,\ldots,j_m\}$ and $[t] \smallsetminus \mathcal{I} = \{k_1,\ldots,k_{m'}\}$. For brevity we simply write $\mathcal{I} = [m]$. Consider the following two cases:

- Case A: $\mathcal{I} = \varnothing$. In this case the definition straightaway gives

  $$\sum_{\boldsymbol{y}_{\leq(i-1)}} \mu(\boldsymbol{\lambda}_i,\hat{\mathcal{R}}_{[t]}) = \mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R}) \times \eta(\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R}).$$

  We remark that for $i = 1$ this is the only possible case.

- Case B: $\mathcal{I} \neq \varnothing \subseteq [t]$. Fix some $(a_{t-m+1},\ldots,a_t) \in \hat{\mathcal{R}}_{[t]\smallsetminus\mathcal{I}}$ and define $a_{\oplus} := a_{t-m+1} \oplus \ldots \oplus a_t$, with $a_{\oplus} = 0$ whenever $\mathcal{I} = [t]$. Fix some $(y_{l_1},\ldots,y_{l_m}) \in \hat{\mathcal{P}}_1 \times \ldots \times \hat{\mathcal{P}}_m$. Then, we have

  $$\sum_{\boldsymbol{y}_{\leq(i-1)}} \mu(\boldsymbol{\lambda}_i, y_{l_1},\ldots,y_{l_m},a_{t-m+1},\ldots,a_t) = \sum_{\boldsymbol{y}_{\leq(i-1)}} \mu(\boldsymbol{\lambda}_i \oplus a_{\oplus}, y_{l_1},\ldots,y_{l_m}) \tag{14}$$

  Thus, we want to count the number of solutions for $\mathbb{S}_{\leq(i-1)}$ that additionally satisfies the equation $\alpha_{l_1} \cdot \boldsymbol{x}_{l_1} \oplus \ldots \oplus \alpha_{l_m} \cdot \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \oplus a_{\oplus}$.

Let $\mathbb{S}'_{\le(i-1)} = \mathbb{S}_{\le(i-1)} \cup \{\alpha_{l_1} \cdot \boldsymbol{x}_{l_1} \oplus \ldots \oplus \alpha_{l_m} \cdot \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \oplus a_\oplus\}$ be the constrained system $\mathbb{S}_{\le(i-1)}$ extended with the additional equation $\boldsymbol{x}_{l_1} \oplus \ldots \boldsymbol{x}_{l_m} = \boldsymbol{\lambda}_i \oplus a_\oplus$. Then, by definition, we have

$$\sum_{\boldsymbol{y}_{\le(i-1)}} \mu(\boldsymbol{\lambda}_i \oplus a_\oplus, y_{l_1}, \ldots, y_{l_m}) = \eta(\mathbb{S}'_{\le(i-1)} \,|\, \mathcal{R}).$$

Let $A'_{\le(i-1)}$ denote the corresponding coefficient matrix. We can have two cases based on the rank of $A'_{\le(i-1)}$:

- Case B1: $A'_{\le(i-1)}$ has full row rank. Suppose $l_m \in \mathrm{col}_{\overline{A}_j}$ for some $j \le (i-1)$ and let $\mathbb{S}_{\le(i-1)\setminus j}$ denote the constrained system that excludes $\mathbb{S}_j$. Then, using the fact that $A'_{\le(i-1)}$ is full rank, we have

  $$\eta(\mathbb{S}'_{\le(i-1)} \,|\, \mathcal{R}) \le N^{t-2} \times \eta(\mathbb{S}_{\le(i-1)\setminus j} \,|\, \mathcal{R}),$$

  and further, using the crude bound (11), we have $\eta(\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R}) \ge (N^{t-1} - t(s+q)N^{t-2}) \times \eta(\mathbb{S}_{\le(i-1)\setminus j} \,|\, \mathcal{R})$ holds as $\mathbb{S}_{\le(i-1)}$ is acyclic and t-regular. Thus,

  $$\eta(\mathbb{S}'_{\le(i-1)} \,|\, \mathcal{R}) \le \frac{2}{N} \eta(\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R}),$$

  where we use the fact that $t(s+q) \le N/2$. There are at most $\binom{t}{m}$ choices for $\mathcal{I}$ and for each such choice there are at most $q^m s^{t-m}$ choices for $(l_1, \ldots, l_m, a_{t-m+1}, \ldots, a_t)$, which finally gives

  $$\sum_{\mathcal{I} \subseteq [t]} \sum_{\boldsymbol{y}_{\le(i-1)}} \mu(\boldsymbol{\lambda}_i, \mathcal{P}_{\mathcal{I}}, \mathcal{R}_{[t]\setminus\mathcal{I}}) \le \frac{2(s+q)^t}{N} \eta(\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R}).$$

- Case B2: $A'_{\le(i-1)}$ does not have full row rank. This case is only possible if the additional equation is defined by the equations in $\mathbb{S}_{\le(i-1)}$. Since $\mathbb{S}_{\le(i-1)}$ is isolated, this case is only possible if the additional equation is redundant, i.e., $\mathcal{I} = [t]$, $\{l_1, \ldots, l_t\} = \mathrm{col}(\overline{A}_j)$ for some $j \le (i-1)$, and $\lambda_j = \lambda_i$. Since there is only one choice for $\mathcal{I}$, and at most $\Delta_{\mathbb{S}}$ choices for j, the number of solutions in this case is bounded by $\Delta_{\mathbb{S}} \eta(\mathbb{S}_{\le(i-1)} \,|\, \mathcal{R})$.

The claim then follows by combining the bounds in all cases, and the lemma follows by substituting the claimed bound in (13). □

Now on to non-isolated components.

Lemma 5. *Suppose $\mathbb{S}_i$ is non-isolated. Then, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \ge \frac{\prod_{j=1}^{t} (N - f^{(j)}_{\le(i-1)})^{r^{(j)}_i}}{N^{q_i}} \left(1 - \frac{2 r^t_i (s+q)^t}{N^t} - \varepsilon_{\mathrm{odd}}(q, r, s, t)\right),$$

where

$$\varepsilon_{\mathrm{odd}}(q, r, s, t) = \begin{cases} \frac{2 q_i (s+q)^{t-1}}{N^{t-1}} & \textit{for odd } t, \\ 0 & \textit{for even } t. \end{cases}$$

Proof. Recall from (10) that

$$\eta(\mathbb{S}_i \mid \mathcal{F}) = \sum_{\mathcal{J} \subseteq [r_i]} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}|.$$

First consider the even t case, where using Bonferroni's inequality, we have

$$\begin{aligned}
\eta(\Psi_i \mid \mathcal{F}) &\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t-1}} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t-1}} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{r_i - |\mathcal{J}| - q_i} \\
&\geq \frac{1}{N^{q_i}} \left(\sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t}} (-1)^{|\mathcal{J}|} f_{(\mathcal{J})} N^{r_i - |\mathcal{J}|} - \sum_{\substack{\mathcal{J}' \subseteq [r_i] \\ |\mathcal{J}'| = t}} f_{(\mathcal{J}')} N^{r_i - t} \right) \\
&\geq \frac{1}{N^{q_i}} \left(\prod_{j=1}^{r_i} (N - f_{(j)}) - r_i^t (s+q)^t N^{r_i - t} \right) \\
&\geq \frac{\prod_{j=1}^{r_i} (N - f_{(j)})}{N^{q_i}} \left(1 - \frac{2 r_i^t (s+q)^t}{N^t} \right),
\end{aligned} \tag{15}$$

where the last inequality follows from the fact that $f_{(j)} \leq (s+q)$ for any j and $r_i(s+q) \leq \xi_{\mathbb{S}}(s+q) \leq N/2$.

As for the odd t case, using Bonferroni's inequality, we have

$$\begin{aligned}
\eta(\Psi_i \mid \mathcal{F}) &\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| \leq t}} (-1)^{|\mathcal{J}|} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| < t}} (-1)^{|\mathcal{J}|} f_{\mathcal{J}} N^{r_i - |\mathcal{J}| - q_i} - \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \\
&\geq \frac{1}{N^{q_i}} \left(\sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| < t}} (-1)^{|\mathcal{J}|} f_{\mathcal{J}} N^{r_i - |\mathcal{J}|} - N^{q_i} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right) \\
&\geq \frac{1}{N^{q_i}} \left(\prod_{j=1}^{r_i} (N - f_{(j)}) - N^{q_i} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right) \\
&\geq \frac{\prod_{j=1}^{r_i} (N - f_{(j)})}{N^{q_i}} \left(1 - \frac{2}{N^{r_i - q_i}} \sum_{\substack{\mathcal{J} \subseteq [r_i] \\ |\mathcal{J}| = t}} |\mathcal{A}_{\mathcal{J}}| \right)
\end{aligned} \tag{16}$$

Claim. We claim

$$\sum_{\substack{\mathcal{J}\subseteq[r_i]\\|\mathcal{J}|=t}} |\mathcal{A}_{\mathcal{J}}| \le q_i(s+q)^{t-1}N^{r_i-t-q_i+1} + r_i^t(s+q)^t N^{r_i-t-q_i}.$$

Proof. Let $\mathcal{J}=\{l_1,\ldots,l_t\}$ and suppose $\mathbb{S}_i'$ denote the updated system after the removal of these t columns from $\mathbb{S}_i$. Using Proposition 5, we have two cases:

- Case A: $\mathcal{J}=\mathrm{col}(\overline{A}_{j\bullet})$ for some $A_{j\bullet}\in\mathrm{row}(A_i)$. From Proposition 5 we know that $\mathrm{rank}(\mathbb{S}_i')=q_i-1$. Thus, we have
$$\sum_{\substack{\mathcal{J}=\mathrm{col}(\overline{A}_{j\bullet})\\A_{j\bullet}\in\mathrm{row}(A_i)}} |\mathcal{A}_{\mathcal{J}}| \le q_i(s+q)^{t-1}N^{r_i-t-q_i+1}.$$
- Case B: $\mathcal{J}\ne\mathrm{col}(\overline{A}_{j\bullet})$ for all $A_{j\bullet}\in\mathrm{row}(A_i)$. From Proposition 5 we know that $\mathrm{rank}(\mathbb{S}_i')=q_i$. Thus, we have
$$\sum_{\mathcal{J}\ne\mathrm{col}(\overline{A}_{j\bullet})} |\mathcal{A}_{\mathcal{J}}| \le r_i^t(s+q)^t N^{r_i-t-q_i}.$$

This proves the claim. □

The result follows by substituting the claimed bound in (16) by realizing that

$$\prod_{j=1}^{r_i}(N-f_{(j)}) = \prod_{k=1}^{t}(N-f_{\le(i-1)}^k)^{r_i^{(k)}}$$ □

Since the bound in Lemma 5 is independent of $\boldsymbol{y}_{\le(i-1)}$, we have the following corollary.

Corollary 2. *Suppose $\mathbb{S}_i$ is non-isolated. Then, we have*

$$\eta(\mathbb{S}_{\le i}\,|\,\mathcal{R}) \ge \frac{\prod_{j=1}^t(N-f_{\le(i-1)}^{(j)})^{r_i^{(j)}}}{N^{q_i}}\left(1-\frac{2r_i^t(s+q)^t}{N^t}-\varepsilon_{\mathrm{odd}}(q,r,s,t)\right)\eta(\mathbb{S}_{\le(i-1)}\,|\,\mathcal{R}),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s,t) = \begin{cases}\frac{2q_i(s+q)^{t-1}}{N^{t-1}} & \text{for odd } t,\\ 0 & \text{for even } t.\end{cases}$$

Theorem 1 now follows from the appropriate recursive application of Lemma 4 and Corollary 2 for all i from c down to 1, carefully accumulating the bound for non-isolated components.

4.2 The Case of CAR Clique System

Towards a variation of Theorem 1, suppose $\mathbb{S}$ is column-uniform, acyclic, k-regular (k-CARC) for some $k\ge 2$, and clique. Thus, $t=1$ in this case.

A system $\mathbb{S}$ is said to be *trivial* if and only if there exists $\boldsymbol{v} \in \mathrm{rowsp}^{+}(A)$ such that

$$H(\boldsymbol{v}) = 2 \quad \text{and} \quad (\boldsymbol{v}|0^n) \in \mathrm{rowsp}^{+}(A|\boldsymbol{\lambda}),$$

and *non-trivial* otherwise. For all trivial systems, $\eta(\mathbb{S}\,|\,\mathcal{R}) = 0$. Accordingly, we assume that the system is non-trivial. Beyond this obvious limitation, the case of clique systems is quite similar to the partite case.

Indeed we reuse the same notations and arguments to a large extent. First, we redefine

$$\mathbb{E}(\mathbb{S}\,|\,\mathcal{R}) \coloneqq \frac{(N-s)_r}{N^q}$$

Next, suppose $\overline{\mathbb{S}}$ denote an arbitrary partite version of $\mathbb{S}$. Set $\mathcal{R}_1 = \cdots = \mathcal{R}_k$, $s_1 = \ldots = s_k$, and reuse the definitions of $\mathcal{A}_{\varnothing}$ and $\mathcal{A}_{\{j\}}$ for any $j \in \mathrm{col}(\overline{A}_i)$. Furthermore, for each $j_1 \neq j_2 \in \mathrm{col}(\overline{A}_i)$, let

$$\mathsf{EQ}_{j_1,j_2} \coloneqq \{y = (y_1,\ldots,y_{r_i}) \in \mathbb{F}_N^{r_i} \;:\; \overline{A}_i y = \boldsymbol{\lambda}_i \wedge y_{j_1} = y_{j_2}\}.$$

Then, for any $i \in [q]$, we have

$$(\mathbb{S}_i\,|\,\mathcal{F}) = \mathcal{A}_{\varnothing} \smallsetminus \left(\left(\bigcup_{j=1}^{r_i} \mathcal{A}_{\{j\}}\right) \cup \left(\bigcup_{j_1<j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2}\right)\right),$$

More importantly,

$$\begin{aligned}
\eta(\mathbb{S}_i\,|\,\mathcal{F}) &= |\mathcal{A}_{\varnothing}| - \left|\bigcup_{j=1}^{r_i} \mathcal{A}_{\{j\}}\right| - \left|\bigcup_{j_1<j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2}\right| \\
&= \eta(\overline{\mathbb{S}}_i\,|\,\mathcal{F}) - \left|\bigcup_{j_1<j_2 \in \mathrm{col}(\overline{A}_i)} \mathsf{EQ}_{j_1,j_2}\right| \\
&\geq \eta(\overline{\mathbb{S}}_i\,|\,\mathcal{F}) - \binom{r_i}{2} N^{r_i-1-q_i}
\end{aligned}$$

where the inequality follows from the fact that $|\mathsf{EQ}_{j_1,j_2}| \leq N^{r_i-1-q_i}$ as $\mathtt{H}(A) \geq k \geq 2$. This gives the following clique counterparts for the results derived in the partite case.

Lemma 6. *Suppose $\mathbb{S}_i$ is isolated and non-trivial. Then, for any $y_{\leq(i-1)} \in (\mathbb{S}_{\leq(i-1)}\,|\,\mathcal{R})$, we have*

$$\eta(\mathbb{S}_i\,|\,\mathcal{F}) \geq \frac{(N-f_{\leq(i-1)})^k}{N}\left(1 - \frac{2}{N^{k-1}}\left|\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{F}) - \frac{f^k_{\leq(i-1)}}{N}\right| - \frac{k^2}{N}\right).$$

Lemma 7. *Suppose $\mathbb{S}_i$ is isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_{\leq i}\,|\,\mathcal{R}) \geq \frac{(N-f_{\leq(i-1)})^k}{N}\left(1 - \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}_i,\mathcal{R})}{N^{k-1}} - \frac{2\Delta_{\mathbb{S}}}{N^{k-1}} - \frac{6(s+kq)^k}{N^k} - \frac{k^2}{N}\right).$$

Lemma 8. *Suppose $\mathbb{S}_i$ is non-isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_i \,|\, \mathcal{F}) \geq \frac{(N - f_{\leq(i-1)})^{r_i}}{N^{q_i}} \left(1 - \frac{2r_i^k(s+kq)^k}{N^k} - \varepsilon_{\mathrm{odd}}(q,r,s) - \frac{r_i^2}{N}\right),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s) = \begin{cases} \frac{2q_i(s+kq)^{k-1}}{N^{k-1}} & \textit{for odd } k, \\ 0 & \textit{for even } k. \end{cases}$$

Corollary 3. *Suppose $\mathbb{S}_i$ is non-isolated and non-trivial. Then, we have*

$$\eta(\mathbb{S}_{\leq i} \,|\, \mathcal{R}) \geq \frac{(N - f_{\leq(i-1)})^{r_i}}{N^{q_i}} \left(1 - \frac{2r_i^k(s+kq)^k}{N^k} - \varepsilon_{\mathrm{odd}}(q,r,s) - \frac{r_i^2}{N}\right) \eta(\mathbb{S}_{\leq(i-1)} \,|\, \mathcal{R}),$$

where

$$\varepsilon_{\mathrm{odd}}(q,r,s) = \begin{cases} \frac{2q_i(s+kq)^{k-1}}{N^{k-1}} & \textit{for odd } k, \\ 0 & \textit{for even } k. \end{cases}$$

Theorem 2 (Clique Bound). *Let $k \geq 2$ and $\mathcal{R}$ be a family of sets. For any $(q,r,1)$-constrained system $\mathbb{S}$ which is non-trivial, k-CARC and which satisfies $\xi_{\mathbb{S}}(q+s) \leq N/2$, we have $\eta(\mathbb{S} \,|\, \mathcal{R}) \geq (1-\varepsilon)\,\mathbb{E}\,(\mathbb{S} \,|\, \mathcal{R})$, where*

$$\varepsilon \leq \frac{2\mu_{\boldsymbol{\alpha}_{\mathbb{S}}}(\boldsymbol{\lambda}, \mathcal{R})}{N^{k-1}} + \frac{2q\mathit{\Delta}_{\mathbb{S}}}{N^{k-1}} + \frac{6q(s+kq)^k}{N^k} + \frac{2qk^2}{N} + \sum_{i \in \mathtt{NI}(\mathbb{S})} \left(\frac{2r_i^k(s+kq)^k}{N^k} + \frac{q_i(s+kq)^{k-1}}{N^{k-1}} + \frac{r_i^2}{N} \right).$$

5 Single-keyed Double-block Hash-then-Sum

Let π be a permutation of $\{0,1\}^n$. We define three injective functions $\pi_0, \pi_1, \pi_2 : \{0,1\}^{n-2} \to \{0,1\}^n$ as follows:

$$\pi_0(\cdot) \coloneqq \pi(00\|\cdot) \qquad \pi_1(\cdot) \coloneqq \pi(01\|\cdot) \qquad \pi_2(\cdot) \coloneqq \pi(10\|\cdot)$$

For $0 \leq j \leq 2$, we define $\mathcal{I}_j\,(n) \coloneqq \{\pi_j : \pi \in \mathcal{P}\,(n)\}$.

Definition 8 (Single-keyed Permutation-based `DBHtS`). *For some permutation π of $\{0,1\}^n$ and a permutation-based rate-c^{-1} diblock hash function $H : \mathcal{I}_0\,(n) \times \{0,1\}^* \to \{0,1\}^{n-2} \times \{0,1\}^{n-2}$, we define the single-keyed* `DBHtS`*, denoted* $\mathtt{1k\text{-}DBHtS}_{\pi,H}$ *construction by the mapping:*

$$m \mapsto \pi_1(H_{\pi_0}(m)) \oplus \pi_2(H_{\pi_0}(m)). \tag{17}$$

The construction is illustrated in Fig. 3.

We drop the parameters π and H whenever they are clear from the context. We reemphasize here that the π_0,π_1,π_2 are all domain-separated versions of the same permutation π.

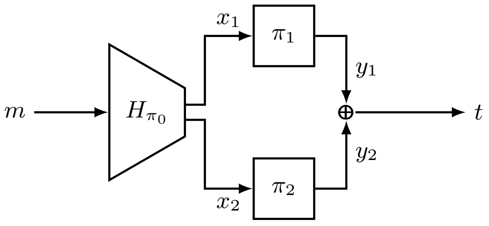


Fig. 3. The $\texttt{1k-DBHtS}_{\pi,H}$ construction.

Theorem 3. *Let $c, q, \ell, \sigma \geq 0$ satisfying $q\ell < \sigma$ and $\overline{\sigma} = c\sigma + 2q \leq 2^{n-3}$. Suppose $H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^{2n-4}$ is a ratec^{-1} $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CFH. Then, for $\rho = (q, \ell, \sigma)$ and $\rho' = (2, \ell, 2\ell)$, the PRF advantage of any ρ-distinguisher $\mathcal{A}$ against $\texttt{1k-DBHtS}_{\Pi,H}$ satisfies*

$$\mathbf{Advt}^{\mathrm{prf}}_{\texttt{1k-DBHtS}_{\Pi,H}}(\mathcal{A}) \leq \epsilon_1 + \epsilon_2,$$

where

$$\epsilon_1 := 2\epsilon_2(\rho, 4) + \delta(\rho) + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3\left(\rho, 2^n/4\overline{\sigma}\right).$$

$$\epsilon_2 := \frac{16q^2\overline{\sigma}^2\epsilon_1(\rho')}{2^{2n}} + \frac{8q^2\epsilon_1(\rho')}{2^n} + \frac{3q\overline{\sigma}}{2^{3n/2}} + \frac{40q\overline{\sigma}^{5/2}}{2^{5n/2}} + \frac{4q\overline{\sigma}^2 + 16q^2\overline{\sigma}^2 + 16q^3\overline{\sigma}}{2^{3n}}.$$

Proof. Without loss of generality assume that $\mathcal{A}$ is deterministic. Let

- $\mathsf{M}^i := (\mathsf{M}^i_1, \ldots, \mathsf{M}^i_{\ell_i})$, denote the i-th query of the distinguisher, containing $\ell_i \leq \ell$ blocks.
- T^i, denote the i-th response of the oracle.

In addition, the oracle releases additional information to the distinguisher, once the distinguisher is done querying the oracle, but before it outputs its decision bit.

In the real world, the oracle releases:

- $\mathsf{X}^i := (\mathsf{X}^i_1, \mathsf{X}^i_2) = H_{\Pi_0}(\mathsf{M}^i)$, the $(2n-2)$-bit internal hash output, or *finalization input* corresponding to the i-th query.
- $\mathsf{Y}^i := (\mathsf{Y}^i_1, \mathsf{Y}^i_2) = (\Pi_1(\mathsf{X}^i_1), \Pi_2(\mathsf{X}^i_2))$, the $2n$-bit *finalization output* corresponding to the i-th query.
- $\mathcal{R}$, the set of all image points sampled during the computation of $H_{\Pi_0}(\mathsf{M}^i)$ for all $i \in [q]$. Since H is a rate-c^{-1} hash function, $|\mathcal{R}| = c\sigma$ for M.

Thus, the full real world transcript can be described as

$$\Theta_{\mathrm{re}} := ((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i, \mathsf{Y}^i : i \in [q]), \mathcal{R}).$$

In the ideal world, the oracle first samples a dummy random permutation Π', and then computes $\mathsf{X}^i := H_{\Pi'_0}(\mathsf{M}^i)$ for all $i \in [q]$. In other words, X^i is generated faithfully for all $i \in [q]$. Note that, $\mathcal{R}$ can be derived here as well, as the ideal oracle is faithfully generating the hash outputs.

Sampling Y in the ideal world: The sampling mechanism for Y^i is on the other hand a bit more sophisticated. The goal is to sample Y^i's in such a way that

$$(\mathsf{X}_1^i = \mathsf{X}_1^j \iff \mathsf{Y}_1^i = \mathsf{Y}_1^j), \qquad\qquad (\mathsf{X}_2^i = \mathsf{X}_2^j \iff \mathsf{Y}_2^i = \mathsf{Y}_2^j),$$

is satisfied for all $i \neq j \in [q]$. We refer to this predicate as the *permutation compatibility* or PC *condition.*

For any $i \in [q]$, let $(i)_1 := \min\{j < i \ : \ \mathsf{X}_1^i = \mathsf{X}_1^j\}$ and $(i)_2 := \min\{j < i \ : \ \mathsf{X}_2^i = \mathsf{X}_2^j\}$. Let $r = |\{(i)_1, (i)_2 \ : \ i \in [q]\}|$. Consider the 2-regular and binary, $(q, r, 1)$-constrained system $\mathbb{S} := \{\mathsf{Y}_1^{(i)_1} \oplus \mathsf{Y}_2^{(i)_2} = \mathsf{T}^i : i \in [q]\}$.

Any $\overline{\mathcal{R}}$-solution for $\mathbb{S}$ satisfies the PC condition, apart from fully defining Y. As long as the system is acyclic and non-trivial, we can use the results developed in the previous section. Keeping this in mind, we now define some bad predicates on the partial transcript $((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i : i \in [q]), \mathcal{R})$:

$$\begin{array}{ll}
\mathsf{A}_1 : \ \exists^* i, j, k, l \in [q], & \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^j = \mathsf{X}_2^k \wedge \mathsf{X}_1^k = \mathsf{X}_1^l. \\
\mathsf{A}_2 : \ \exists^* i, j \in [q], & \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{T}^i \oplus \mathsf{T}^j = 0^n. \\
\mathsf{A}_3 : \ \exists^* k \geq 2^{n-2}/(c\sigma + 2q), i_1, \ldots, i_k \in [q], & \mathsf{X}_1^{i_1} = \mathsf{X}_1^{i_2} = \ldots = \mathsf{X}_1^{i_k}. \\
\mathsf{B}_1 : \ \exists^* i, j, k, l \in [q], & \mathsf{X}_2^i = \mathsf{X}_2^j \wedge \mathsf{X}_1^j = \mathsf{X}_1^k \wedge \mathsf{X}_2^k = \mathsf{X}_2^l. \\
\mathsf{B}_2 : \ \exists^* i, j \in [q], & \mathsf{X}_2^i = \mathsf{X}_2^j \wedge \mathsf{T}^i \oplus \mathsf{T}^j = 0^n. \\
\mathsf{B}_3 : \ \exists^* k \geq 2^{n-2}/(c\sigma + 2q), i_1, \ldots, i_k \in [q], & \mathsf{X}_2^{i_1} = \mathsf{X}_2^{i_2} = \ldots = \mathsf{X}_2^{i_k}. \\
\mathsf{C} : \ \exists^* i \in [q], & \mathsf{T}^i = 0^n. \\
\mathsf{D} : \ \exists^* i, j \in [q], & \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^i = \mathsf{X}_2^j. \\
\mathsf{E} : \ \exists^* i, j, k \in [q], & \mathsf{X}_1^i = \mathsf{X}_1^j \wedge \mathsf{X}_2^j = \mathsf{X}_2^k \wedge \mathsf{T}^i \oplus \mathsf{T}^j \oplus \mathsf{T}^k = 0^n.
\end{array}$$

Define $\mathsf{Cyclic} := \mathsf{A}_1 \vee \mathsf{B}_1 \vee \mathsf{D}$, $\mathsf{Trivial} := \mathsf{A}_2 \vee \mathsf{B}_2 \vee \mathsf{C} \wedge \mathsf{E}$, and $\mathsf{Giant} := \mathsf{A}_3 \vee \mathsf{B}_3$. It is not difficult to see that as long as Cyclic, $\mathsf{Trivial}$, and Giant are false, $\mathbb{S}$ is acyclic and non-trivial, and satisfies $\chi_{\mathbb{S}}(c\sigma + 2q) \leq 2^{n-1}$ for $(c\sigma + 2q) < 2^{3n/4}$. For notational convenience, let $s = c\sigma$.

Onwards we describe the sampling of Y conditioned on the fact that $\neg(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant})$ holds. Let $\mathrm{CF}(\mathbb{S}) = (\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_c)$ such that all the isolated components appear before the non-isolated ones. Let $\mathsf{Y}_0 = \mathcal{R}$, and Y_i denote the $\overline{\mathsf{Y}_{\leq(i-1)}}$-solution for $\mathbb{S}_i$, where $\mathsf{Y}_{\leq(i-1)}$ denotes a $\overline{\mathsf{Y}_0}$-solution for $\mathbb{S}_{\leq(i-1)} = (\mathbb{S}_1 \prec \ldots \prec \mathbb{S}_{i-1})$. Let $\mathcal{F}(\mathsf{Y}_{\leq(i-1)}) := \mathsf{Y}_0 \cup \mathsf{Y}_{\leq(i-1)}$ and and $f_{\leq(i-1)} := |\mathcal{F}(\mathsf{Y}_{\leq(i-1)})|$.

Sampling Y_i *in isolated case:* For the i-th isolated component, using Lemma 6, the number of solutions conditioned on the forbidden set Y_0 and a compatible solution $\mathsf{Y}_{\leq(i-1)}$ of $\mathbb{S}_{\leq(i-1)}$ is given by

$$\eta(\mathbb{S} \,|\, \mathcal{F}(\mathsf{Y}_{\leq(i-1)})) \geq \frac{(2^n - f_{\leq(i-1)})^2}{2^n} \left(1 - \frac{2}{2^n} \left| \mu(\mathsf{T}^{(i)}, \mathcal{F}) - \frac{f_{\leq(i-1)}^2}{2^n} \right| - \frac{4}{2^n}\right), \tag{18}$$

where $\mathsf{T}^{(i)} = \mathsf{T}^j$ for some $j \in [q]$ and $f_{\leq(i-1)} = s + 2(i-1)$.

Sampling Y^i *in non-isolated case:* For the i-th non-isolated component, using Lemma 8, the number of solutions conditioned on the forbidden set Y_0 and a compatible solution $\mathsf{Y}_{\le(i-1)}$ of $\mathbb{S}_{\le(i-1)}$ is given by

$$\eta(\mathbb{S}\,|\,\mathcal{F}(\mathsf{Y}_{\le(i-1)})) \ge \frac{(2^n - f_{\le(i-1)})^{r_i}}{2^{nq_i}}\left(1 - \frac{2r_i^2(s+2q)^2}{2^{2n}} - \frac{r_i^2}{2^n}\right). \tag{19}$$

Now, for all $i \in [c]$, we sample $\mathsf{Y}_i \twoheadleftarrow (\mathbb{S}\,|\,\mathcal{F}(\mathsf{Y}_{\le(i-1)}))$. This concludes the sampling in the ideal world, and finally the ideal world transcript is given by

$$\Theta_{\text{id}} := ((\mathsf{M}^i, \mathsf{T}^i, \mathsf{X}^i, \mathsf{Y}^i : i \in [q]), \mathcal{R}).$$

where the PC condition is satisfied as long as $\neg(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant})$ holds; otherwise the transcript is defined arbitrarily.

(Bad) Transcript Definition and Analysis: The set of transcripts Ω is the set of all tuples $\omega = ((m^i, t^i, x^i, y^i : i \in [q]), \mathcal{R})$, where $m^i \in \{0,1\}^*$, $t^i \in \{0,1\}^n$, $x^i \in \{0,1\}^{2n-2}$, $y^i \in \{0,1\}^{2n}$ and $\mathcal{R} \subseteq (\{0,1\}^n)^{c\sigma}$, where $\sigma = \sum_{i=1}^{q} \lceil |m^i|/n \rceil$.

A transcript ω is said to be *bad*, i.e., $\omega \in \Omega_{\text{bad}}$ if and only if it satisfies $\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant}$, and *good* otherwise.

Lemma 9. *Suppose H is an $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-coverfree hash function. Then*

$$\Pr(\Theta_{\text{id}} \in \Omega_{\text{bad}}) \le 2\epsilon_2(\rho, 4) + \delta(\rho) + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3\left(\rho, \frac{2^{n-2}}{c\sigma + 2q}\right).$$

Proof. Let $s' = 2^{n-2}/(c\sigma + q)$. We have

$$\begin{aligned}
\Pr(\Theta_{\text{id}} \in \Omega_{\text{bad}}) &= \Pr(\mathsf{Cyclic} \vee \mathsf{Trivial} \vee \mathsf{Giant}) \\
&\le \Pr(\mathsf{Cyclic}) + \Pr(\mathsf{Trivial}) + \Pr(\mathsf{Giant}) \\
&\le \Pr(\mathsf{A}_1) + \Pr(\mathsf{B}_1) + \Pr(\mathsf{D}) + \Pr(\mathsf{A}_2) + \Pr(\mathsf{B}_2) + \Pr(\mathsf{C}) + \Pr(\mathsf{E}) + \Pr(\mathsf{A}_3) + \Pr(\mathsf{B}_3) \\
&\le \Pr\left(\mathsf{AP1}_H^4(\mathsf{M})\right) + \Pr\left(\mathsf{AP2}_H^4(\mathsf{M})\right) + \Pr(\mathsf{COLL}_H(\mathsf{M})) + \frac{\Pr(\mathsf{COLL1}_H(\mathsf{M}))}{2^n} \\
&\quad + \frac{\Pr(\mathsf{COLL2}_H(\mathsf{M}))}{2^n} + \frac{q}{2^n} + \frac{\Pr\left(\mathsf{AP1}_H^3(\mathsf{M})\right)}{2^n} + \Pr\left(\mathsf{MC1}_H^{s'}(\mathsf{M})\right) + \Pr\left(\mathsf{MC2}_H^{s'}(\mathsf{M})\right) \\
&\le 2\epsilon_2(\rho, 4) + \delta + \frac{q + 2\epsilon_1(\rho) + \epsilon_2(\rho, 3)}{2^n} + 2\epsilon_3(\rho, s'),
\end{aligned}$$

where the the first three (in)equalities follow from the definition and a trivial application of union bound, the fourth inequality just maps the bad predicates to corresponding coverfree hash events, and finally the fifth inequality follows from the coverfree bound of H. □

Good Transcript Analysis: Fix a good transcript $\omega \in \Omega_{\text{good}}$. We will recycle notations from the sampling phase.

In the real world, $\mathsf{\Pi}$ is sampled exactly $s+r$ times ($|\mathcal{R}| = s$ and $|\{(i)_1, (i)_2 : i \in [q]\}| = r$). Thus, we have

$$\Pr\left(\Theta_{\mathrm{re}} = \omega\right) = \frac{1}{(2^n)_{s+r}} \tag{20}$$

In the ideal world, first T is sampled uniformly from a set of size 2^{nq}, followed by $\mathcal{R}$ which is sampled faithfully via $\mathsf{\Pi}$. Finally, Y is sampled. Let $\mathrm{CF}(\mathbb{S}) = (\mathbb{S}_1, \ldots, \mathbb{S}_c)$ where the first $c' \le c$ components are isolated and the remaining components are non-isolated. Then, we have

$$\begin{aligned}\Pr\left(\Theta_{\mathrm{id}} = \omega\right) &= \frac{1}{2^{nq}} \times \frac{1}{(2^n)_s} \times \prod_{i=1}^{c'} \frac{1}{\eta(\mathbb{S}_i \,|\, \mathcal{F}(\mathsf{Y}_{\le(i-1)}))} \times \prod_{i'=c'+1}^{c} \frac{1}{\eta(\mathbb{S}_{i'} \,|\, \mathcal{F}(\mathsf{Y}_{\le(i'-1)}))} \\ &\le \frac{1}{2^{nq}} \times \frac{1}{(2^n)_s} \times \prod_{i=1}^{c'} \frac{2^n}{(1-\mu_i)(2^n - f_{\le(i-1)})^2} \times \prod_{i'=c'+1}^{c} \frac{2^{nq_{i'}}}{(1-\nu_{i'})(2^n - f_{\le(i'-1)})^{r_{i'}}}\end{aligned}$$

where

$$\mu_i = \frac{2}{2^n}\left|\mu(\mathsf{T}^{(i)}, \mathcal{F}) - \frac{f^2_{\le(i-1)}}{2^n}\right| + \frac{4}{2^n}, \tag{21}$$

$$\nu_{i'} = \frac{2r^2_{i'}(s+2q)^2}{2^{2n}} + \frac{r^2_{i'}}{2^n}. \tag{22}$$

Continuing on we have

$$\begin{aligned}\Pr\left(\Theta_{\mathrm{id}} = \omega\right) &\le \frac{1}{(2^n)_s} \times \prod_{i=1}^{c'} \frac{1}{(1-\mu_i)(2^n - f_{\le(i-1)})^2} \times \prod_{i'=c'+1}^{c} \frac{1}{(1-\nu_{i'})(2^n - f_{\le(i'-1)})^{r_{i'}}} \\ &\le \frac{1}{\left(1 - \sum_{i=1}^{c'} \mu_i\right)} \times \frac{1}{\left(1 - \sum_{i'=c'+1}^{c} \nu_{i'}\right)} \times \prod_{i=1}^{c} \frac{1}{(2^n - f_{\le(i-1)})^{r_i}}\end{aligned} \tag{23}$$

On dividing (20) by (23), we have

$$\begin{aligned}\frac{\Pr\left(\Theta_{\mathrm{re}} = \omega\right)}{\Pr\left(\Theta_{\mathrm{id}} = \omega\right)} &\ge \left(1 - \sum_{i=1}^{c'} \mu_i - \sum_{i'=c'+1}^{c} \nu_{i'}\right) \times \frac{\prod_{i=1}^{c}(2^n - f_{\le(i-1)})^{r_i}}{(2^n)_{s+r}} \\ &\ge \left(1 - \sum_{i=1}^{c'} \mu_i - \sum_{i'=c'+1}^{c} \nu_{i'}\right).\end{aligned} \tag{24}$$

In anticipation of applying the Expectation Method Corollary 1, we have to compute

$$\mathbb{E}\left(\sum_{i=1}^{c'} \mu_i\right) \qquad \mathbb{E}\left(\sum_{i'=c'+1}^{c} \nu_{i'}\right)$$

First, let $\sim_1$ (res. $\sim_2$) be equivalence relations on $[q]$, such that $i \sim_1 j$ (res. $i \sim_2 j$) if and only if $\mathsf{X}^i_1 = \mathsf{X}^j_1$ (res. $\mathsf{X}^i_2 = \mathsf{X}^j_2$). Let $\mathcal{C}^1_1, \ldots, \mathcal{C}^1_{t_1}$ and $\mathcal{C}^2_1, \ldots, \mathcal{C}^2_{t_2}$ denote the non-singleton equivalence classes of $[q]$ with respect to $\sim_1$ and $\sim_2$, respectively. For $i \in [t_1]$ and $j \in [t_2]$, let $\mathtt{mc}_i^{(1)} = |\mathcal{C}_i^1|$ and $\mathtt{mc}_j^{(2)} = |\mathcal{C}_j^2|$.

$$\begin{aligned}
\mathbb{E}\left(\sum_{i'=c'+1}^{c} \nu_{i'}\right) &= \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right)\mathbb{E}\left(\sum_{i'=c'+1}^{c} r_{i'}^2\right) \\
&\leq \left(\frac{2(s+2q)^2}{2^{2n}} + \frac{1}{2^n}\right) \times 2\left(\sum_{j=1}^{t_1}\mathbb{E}\left(\mathtt{mc}_j^{(1)}\right) + \sum_{j'=1}^{t_2}\mathbb{E}\left(\mathtt{mc}_{j'}^{(2)}\right)\right) \\
&\leq \frac{16q^2(s+2q)^2\epsilon_1(2,\ell,2\ell)}{2^{2n}} + \frac{8q^2\epsilon_1(2,\ell,2\ell)}{2^n}. \qquad (25)
\end{aligned}$$

Second, using Proposition 1, we have

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{c'}\mu_i\right) &= \mathbb{E}\left(\frac{2}{2^n}\sum_{i=1}^{c'}\left|\mu(\mathsf{T}^{(i)},\mathcal{F}) - \frac{f_{\leq(i-1)}^2}{2^n}\right| + \sum_{i=1}^{c'}\frac{4}{2^n}\right) \\
&= \frac{2}{2^n}\sum_{i=1}^{c'}\mathbb{E}\left(\left|\mu(\mathsf{T}^{(i)},\mathcal{F}) - \frac{f_{\leq(i-1)}^2}{2^n}\right|\right) + \frac{4q}{2^n} \\
&\leq \frac{2}{2^n}\sum_{i=1}^{c'}\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right)} + \frac{2}{2^n}\sum_{i=1}^{c'}\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) - \frac{f_{\leq(i-1)}^2}{2^n}\right| + \frac{4q}{2^n} \qquad (26)
\end{aligned}$$

We claim:

$$\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) - \frac{f_{\leq(i-1)}^2}{2^n}\right| \leq \frac{2s^2 + 8q(s+2q)^2 + 8q^2(s+2q)}{2^{2n}} \qquad (27)$$

$$\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right)} \leq \frac{\sqrt{2}(s+2q)}{2^{n/2}} + \frac{20(s+2q)^{5/2}}{2^{3n/2}} \qquad (28)$$

A proof of this claim is given in Appendix A. Theorem 3 then follows from Lemma 9 and (25)-(28). □

6 Instantiations of Cover-free Hash functions

For a diblock hash function $H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$ we can construct the truncated diblock hash $\mathtt{T}H : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^{n-2} \times \{0,1\}^{n-2}$ as $\mathtt{T}H(x) := (\mathtt{Trunc}(H_1(x)), \mathtt{Trunc}(H_2(x)))$, where $\mathtt{Trunc} : \{0,1\}^n \to \{0,1\}^{n-2}$ truncates the first two bits of its n-bit input.

Now let us define the functions $\mathtt{PHash} : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$ and $\mathtt{LightHash} : \mathcal{I}_0(n) \times \{0,1\}^* \to \{0,1\}^n \times \{0,1\}^n$, as follows:

Two instances of CfHs will be the truncated versions of the above hash functions, $\mathtt{TPHash}$ and $\mathtt{TLightHash}$, respectively. In fact, we have that $\mathtt{1k\text{-}PMAC+} = \mathtt{1k\text{-}DBHtS}_{\mathtt{TPHash}}$ and $\mathtt{1k\text{-}LightMAC+} = \mathtt{1k\text{-}DBHtS}_{\mathtt{TLightHash}}$.

```
PHash_{Π_0}
Input: $m = m[1]\|\cdots\|m[k] \in (\{0,1\}^{n-2})^k$
$\Delta_0 \leftarrow \texttt{Trunc}(\Pi_0(0))$
$\Delta_1 \leftarrow \texttt{Trunc}(\Pi_0(1))$
for $i \in [k]$,
    $W[i] \leftarrow m[i] \oplus 2^i \cdot \Delta_0 \oplus 2^{2i} \cdot \Delta_1$
    $Z[i] \leftarrow \Pi_0(W[i])$
$x[1] \leftarrow Z[1] \oplus Z[2] \cdots \oplus Z[k]$
$x[2] \leftarrow Z[1] \oplus 2 \cdot Z[2] \cdots \oplus 2^{k-1} \cdot Z[k]$
return $x := (x[1]\|x[2])$
```

```
LightHash_{Π_0}
Input: $m = m[1]\|\cdots\|m[k] \in (\{0,1\}^{n-s})^k$
for $i \in [k]$,
    $Z[i] \leftarrow \Pi_0(\langle i \rangle_{s-2}\|m[i])$
$x[1] \leftarrow Z[1] \oplus Z[2] \oplus \cdots \oplus Z[k]$
$x[2] \leftarrow 2^{k-1} \cdot Z[1] \oplus 2^{k-2} \cdot Z[2] \cdots \oplus Z[k]$
return $x := (x[1]\|x[2])$
```

6.1 Affine bad events.

For a diblock hash function H, any $\boldsymbol{x} = (x_1, \dots, x_q) \in (\mathcal{X})_q$, and $c, c_1, c_2, c_3 \in \{0,1\}^2$, we define:

$$\mathsf{COLL}_H^{c_1,c_2}(\boldsymbol{x}): \ \exists^* i, j \in [q] \text{ such that } H_K(x_i) \oplus H_K(x_j) = (c_1\|0^{n-2}, c_2\|0^{n-2})$$

$$\mathsf{COLL1}_H^{c}(\boldsymbol{x}): \ \exists^* i, j \in [q] \text{ such that } H_K^1(x_i) \oplus H_K^1(x_j) = c\|0^{n-2}.$$

$$\mathsf{COLL2}_H^{c}(\boldsymbol{x}): \ \exists^* i, j \in [q] \text{ such that } H_K^2(x_i) \oplus H_K^2(x_j) = c\|0^{n-2}.$$

$$\mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x}): \ \exists^* i, j, k, l \in [q] \text{ such that}$$

$$H_K^1(x_i) \oplus H_K^1(x_j) = c_1\|0^{n-2} \wedge H_K^2(x_j) \oplus H_K^2(x_k) = c_2\|0^{n-2}$$
$$\wedge H_K^1(x_k) \oplus H_K^1(x_l) = c_3\|0^{n-2}.$$

$$\mathsf{AP2}_H^{c_1,c_2,c_3}(\boldsymbol{x}): \ \exists^* i, j, k, l \in [q] \text{ such that}$$

$$H_K^2(x_i) \oplus H_K^2(x_j) = c_1\|0^{n-2} \wedge H_K^1(x_j) \oplus H_K^1(x_k) = c_2\|0^{n-2}$$
$$\wedge H_K^2(x_k) \oplus H_K^2(x_l) = c_3\|0^{n-2}.$$

$$\mathsf{AP2}_H^{c_1,c_2}(\boldsymbol{x}): \ \exists^* i, j, k \in [q] \text{ such that}$$

$$H_K^2(x_i) \oplus H_K^2(x_j) = c_1\|0^{n-2} \wedge H_K^1(x_j) \oplus H_K^1(x_k) = c_2\|0^{n-2}$$

$$\mathsf{MC1}_H^{c_1,\dots,c_s}(\boldsymbol{x}): \ \exists^* i, j, k, l \in [q] \text{ such that}$$

$$H_K^1(x_i) \oplus H_K^1(x_j) = c_1\|0^{n-2} \wedge \cdots \wedge H_K^1(x_{i_{s-1}}) \oplus H_K^1(x_{i_s}) = c_s\|0^{n-2}$$

$$\mathsf{MC2}_H^{c_1,\dots,c_s}(\boldsymbol{x}): \ \exists^* i, j, k, l \in [q] \text{ such that}$$

$$H_K^2(x_i) \oplus H_K^2(x_j) = c_1\|0^{n-2} \wedge \cdots \wedge H_K^2(x_{i_{s-1}}) \oplus H_K^2(x_{i_s}) = c_s\|0^{n-2}$$

One can readily check that

$$\mathsf{COLL1}_{\texttt{TH}}(\boldsymbol{x}) = \bigvee_{c \in \{0,1\}^2} \mathsf{COLL1}_H^c(\boldsymbol{x}) \qquad \mathsf{COLL2}_{\texttt{TH}}(\boldsymbol{x}) = \bigvee_{c \in \{0,1\}^2} \mathsf{COLL2}_H^c(\boldsymbol{x})$$

$$\begin{aligned}
\mathsf{AP1}^4_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2,c_3)\\ \in(\{0,1\}^2)^3}} \mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x}) & \mathsf{AP2}^4_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2,c_3)\\ \in(\{0,1\}^2)^3}} \mathsf{AP2}_H^{c_1,c_2,c_3}(\boldsymbol{x})\\
\mathsf{COLL}_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2)\\ \in(\{0,1\}^2)^2}} \mathsf{COLL}_H^{c_1,c_2}(\boldsymbol{x}) & \mathsf{AP1}^3_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{(c_1,c_2)\\ \in(\{0,1\}^2)^3}} \mathsf{AP1}_H^{c_1,c_2,c_3}(\boldsymbol{x})\\
\mathsf{MC1}^s_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{c^s\\ \in(\{0,1\}^2)^s}} \mathsf{MC1}_H^{c_1,\cdots,c_s}(\boldsymbol{x}) & \mathsf{MC2}^s_{\mathtt{T}H}(\boldsymbol{x}) &= \bigvee_{\substack{c^s\\ \in(\{0,1\}^2)^s}} \mathsf{MC2}_H^{c_1,\cdots,c_s}(\boldsymbol{x})
\end{aligned} \tag{29}$$

6.2 TPHash.

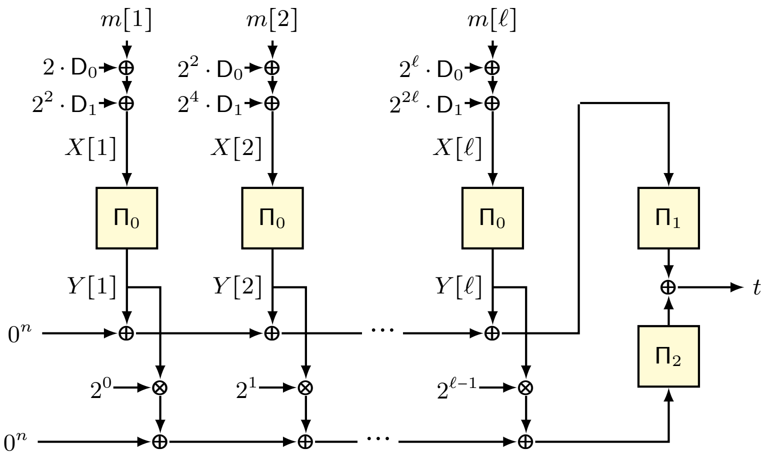


Fig. 4. 1k-PMAC+

Our bad event analysis heavily depends on the one presented in [27]. We tailor their bounds according to our needs while highlighting the main aspects of similarity and departure between their results and ours.

Similar to the PMAC+ analysis in [27] we define analogous auxiliary events as follows: Let the i-th message be $m^i = m^i[1]\|\cdots\|m^i[\ell_i] \in (\{0,1\}^{n-2})^{\ell_i}$, $i \in [q]$. For $i \neq j \in [q]$, let $\ell = \min\{\ell_i, \ell_j\}$ and $\ell' = \max\{\ell_i, \ell_j\}$, then we can define the index set for which $m^i[k] \neq m^j[k]$ as

$$I_{ij} := \{k \in [\ell] : m^i[k] \neq m^j[k]\} \sqcup [\ell + 1..\ell']$$

We define the following random variables: $\mathsf{D}_0 := \mathtt{Trunc}(\Pi_0(0))$, $\mathsf{D}_1 := \mathtt{Trunc}(\Pi_0(1))$, and $\mathsf{W}^i = \mathsf{W}^i[1]\|\cdots\|\mathsf{W}^i[\ell_i]$, where $\mathsf{W}^i[k] = m^i[k] \oplus 2^k \cdot \mathsf{D}_0 \oplus 2^{2k} \cdot \mathsf{D}_1$. We further define the random index sets where W^i and W^j collide as follows:

$$\mathtt{I}_{\mathsf{col}} = \{(i,j) \in ([q])_2 : \exists^* k, k' \text{ such that } \mathsf{W}^i[k] = \mathsf{W}^j[k']\}$$

$$\mathsf{J_{col}} = \{(i,j) \in ([q])_2 : \min\{I_{ij}\} \leq \ell_i \text{ and } \exists k \text{ such that } \mathsf{W}^i[\min\{I_{ij}\}] = \mathsf{W}^j[k]\}$$

Then the auxiliary events are:
Aux_1 **:** $\mathsf{D}_0 = 0 \vee \mathsf{D}_1 = 0$
Aux_2 **:** $\exists i \in [q], \exists^* k, k'$ such that $\mathsf{W}^i[k] = \mathsf{W}^i[k']$.
Aux_3 **:** $\exists i \in [q], k \in [\ell_i]$ such that $\mathsf{W}^i[k] \in \{0, 1, \Pi_0^{-1}(0)\}$.
Aux_4 **:** $|\mathsf{I_{col}}| > s$, where $s = 2^n/4\overline{\sigma}$.
Aux_5 **:** $|\mathsf{J_{col}}| > s'$ where $s' = \ell q$
and let $\mathsf{Aux} = \bigvee_{i \in [5]} \mathsf{Aux}_i$.

Lemma 10. *For* $\boldsymbol{m} = (m^i : i \in [q])$ *and* $c, c_1, c_2, c_3 \in \{0,1\}^2$,

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) \leq \frac{4\ell q^2}{2^{2n}}$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) \leq \frac{2s'^2}{2^{2n}} + \frac{4s}{2^n} + \frac{2}{2^n} + \frac{2\sqrt{2}q^2}{2^{3n/2}} + \frac{8sq^2}{2^{2n}} + \frac{96q^2}{2^{2n}} + \frac{8q^4}{2^{3n}}$$

Proof Sketch: First we note that, the following pairs of events, the left defined in [27] and the right defined in this paper, are equivalent in the single-key scenario:

$$\mathsf{Bad}_1 \equiv \mathsf{COLL}^{0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}), \qquad \mathsf{Bad}_2 \equiv \mathsf{AP1}^{0,0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$$

Analogous to Eq. (10) and (11) of [27], we can write, for any $c \in \{0,1\}^2$,

$$\mathtt{PHash}^1_{\Pi_0}(m^i) \oplus \mathtt{PHash}^1_{\Pi_0}(m^j) = c\|0^{n-2} \iff A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = c\|0^{n-2}$$
$$\mathtt{PHash}^2_{\Pi_0}(m^i) \oplus \mathtt{PHash}^2_{\Pi_0}(m^j) = c\|0^{n-2} \iff B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = c\|0^{n-2}$$

where, for $(i,j) \in ([q])_2$, $\{\mathsf{W}[1], \ldots, \mathsf{W}[t]\} := \{\mathsf{W}^i[1], \ldots, \mathsf{W}^i[\ell_i]\} \cup \{\mathsf{W}^j[1], \ldots, \mathsf{W}^j[\ell_j]\}$, and for $k \in [t]$, $\mathsf{Z}[k] := \Pi_0(\mathsf{W}[k])$.

Thus, borrowing from the analysis of [27], each of the events in the statement of this lemma can be written as an event that a system of equations $\boldsymbol{AZ} = \boldsymbol{c}$ holds, where $\boldsymbol{Z}$ is a vector with k-th component $\mathsf{Z}[k]$, and $\boldsymbol{c}$ depends on the indices c, c_1, c_2, c_3 of the corresponding event. If $\boldsymbol{c} \notin \mathcal{C}(\boldsymbol{A})$, then this system of equations will hold with 0 probability. If $\boldsymbol{c} \in \mathcal{C}(\boldsymbol{A})$ then the probability that this system of equations holds, depends on the rank of $\boldsymbol{A}$ and not on $\boldsymbol{c}$. So we have that

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) \leq \Pr\left(\mathsf{COLL}^{0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) = \Pr\left(\mathsf{Bad}_1 \wedge \neg\mathsf{Aux}\right)$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) \leq \Pr\left(\mathsf{AP1}^{0,0,0}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m}) \wedge \neg\mathsf{Aux}\right) = \Pr\left(\mathsf{Bad}_2 \wedge \neg\mathsf{Aux}\right)$$

Thus we can use the bounds on the corresponding bad events from [27] to get our result. □

The probability analysis of the events $\mathsf{AP2}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{AP1}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ are similar to the analysis of the events $\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{COLL}^{c_1,c_2}_{\mathtt{PHash}_{\Pi_0}}(\boldsymbol{m})$, respectively.

Lemma 11. *For $\ell \le 2^{n-2}$, $m \ne m' \in (\{0,1\}^{n-2})^{\le \ell}$, and $c \in \{0,1\}^2$, we have*

$$\Pr\left(\mathit{PHash}^1_{\Pi_0}(m) \oplus \mathit{PHash}^1_{\Pi_0}(m') = c \| 0^{n-2}\right) \le \frac{26\ell}{2^n}$$

$$\Pr\left(\mathit{PHash}^2_{\Pi_0}(m) \oplus \mathit{PHash}^2_{\Pi_0}(m') = c \| 0^{n-2}\right) \le \frac{26\ell}{2^n}$$

Proof. Let $m \in (\{0,1\}^{n-2})^{\ell}$ and $m' \in (\{0,1\}^{n-2})^{\ell'}$. Note that the claim is trivial $\ell = 1$ and we ignore this case.

Let i be the maximum block-index where m and m' are distinct, precisely,

$$i = \begin{cases} \ell, & \text{if } \ell > \ell' \\ \max\{j \le \ell : m[j] \ne m'[j]\}, & \text{if } \ell = \ell' \end{cases}$$

Consider the random variables:

$$\begin{aligned} &\mathsf{D}_0 = \mathsf{trunc}(\Pi(0)), && \mathsf{D}_1 = \mathsf{trunc}(\Pi(1)), && \\ &\mathsf{W}[i] = m[i] \oplus 2^i \cdot \mathsf{D}_0 \oplus 2^{2i} \cdot \mathsf{D}_1, && \mathsf{Z}[i] = \Pi_0(\mathsf{W}[i]), && i \in [\ell] \\ &\mathsf{W}'[i] = m'[i] \oplus 2^i \cdot \mathsf{D}_0 \oplus 2^{2i} \cdot \mathsf{D}_1, && \mathsf{Z}'[i] := \Pi_0(\mathsf{W}'[i]), && i \in [\ell'] \end{aligned}$$

Let us define the following events:

$$\begin{aligned} &\mathsf{E}_1 : \mathsf{D}_0 = 0 \\ &\mathsf{E}_2 : \bigvee_{j \in [\ell]} (\mathsf{W}[j] = 0 \vee \mathsf{W}[j] = 1) \vee \bigvee_{j \in [\ell']} (\mathsf{W}'[j] = 0 \vee \mathsf{W}'[j] = 1) \\ &\mathsf{E}_3 : \bigvee_{\substack{j \in [\ell] \\ j \ne i}} (\mathsf{W}[i] = \mathsf{W}[j]) \vee \bigvee_{j \in [\ell']} (\mathsf{W}[i] = \mathsf{W}'[j]) \end{aligned}$$

Note that $\Pr(c \cdot \mathtt{Trunc}(\Pi(a)) = b) = 4/2^n$ for any $a \in \{0,1\}^n$ and $b, c \in \{0,1\}^{n-2}$ with $c \ne 0$. Hence, for any $a_1, \ldots, a_r \in \{0,1\}^n$ and $b, c_1, \ldots, c_r \in \{0,1\}^{n-2}$ with $c_r \ne 0$, we have

$$\begin{aligned} &\Pr(c_1 \cdot \mathtt{Trunc}(\Pi(a_1)) \oplus \cdots \oplus c_r \cdot \mathtt{Trunc}(\Pi(a_r)) = b) \\ &= \sum_{\substack{b'_1, \ldots, b'_{r_1} \\ \in \{0,1\}^n \\ \text{all distinct}}} \Pr(\mathtt{Trunc}(\Pi(a_r)) = b') \Pr(\Pi(a_i) = b'_i \ \forall i \in [r-1]) \\ &\le \frac{4}{2^n - r + 1} \end{aligned}$$

where $b_i = \mathsf{trunc}(b'_i)$ and $b' = c_r^{-1} \cdot (b \oplus c_1 \cdot b_1 \oplus \cdots \oplus c_{r-1} \cdot b_{r-1})$. Similarly for any $a_1, \ldots, a_r \in \{0,1\}^n$ and $b, c_1, \ldots, c_r \in \{0,1\}^{n-2}$ with at least one $c_i \ne 0$, we have

$$\Pr(c_1 \cdot \Pi(a_1) \oplus \cdots \oplus c_r \cdot \Pi(a_r) = b) \le \frac{1}{2^n - r + 1}. \tag{30}$$

This implies $\Pr(\mathsf{E}_1) = \Pr(\mathsf{trunc}(\Pi(0)) = 0) = 4/2^n$, $\Pr(\mathsf{E}_2 \mid \mathsf{E}_1^c) \le 4\ell \cdot 4/2^n$, and $\Pr(\mathsf{E}_3 \mid \mathsf{E}_1^c \wedge \mathsf{E}_2^c) \le (2\ell - 1) \cdot 4/2^n$.

Now the event $\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2}$, is equivalent to $\mathsf{Z}[1] \oplus \cdots \oplus \mathsf{Z}[\ell] \oplus \mathsf{Z}'[1] \oplus \cdots \oplus \mathsf{Z}'[\ell'] = c\|0^{n-2}$. Of course, if any two Z-random variables are identically equal then they cancel out. However, conditional on $\mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c$ we have $\mathsf{Z}[i] \neq \mathsf{Z}[j], \mathsf{Z}'[j']$ for any $j \in [m] \smallsetminus \{i\}, j' \in [m']$ and $\mathsf{Z}[i] \neq 0, \Pi(0), \Pi(1)$. Hence from Eq. (30), we have

$$\Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2} \,\middle|\, \mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c\right) \leq \frac{1}{2^n - (m-1) - m' - 2} \leq \frac{1}{2^n - 2\ell} \leq 2/2^n$$

assuming $\ell \leq 2^{n-2}$.

Since for any two events A and B, we have $\Pr(\mathsf{A}) = \Pr(\mathsf{A} \wedge \mathsf{B}) + \Pr(\mathsf{A} \wedge \mathsf{B}^c)$ and $\Pr(\mathsf{A} \wedge \mathsf{B}) \leq \Pr(\mathsf{A})$ and $\Pr(\mathsf{A} \wedge \mathsf{B}) \leq \Pr(\mathsf{A} \mid \mathsf{B})$, we have

$$\begin{aligned}
&\Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2}\right) \\
&\leq \Pr(\mathsf{E}_1) + \Pr(\mathsf{E}_2 \mid \mathsf{E}_1^c) + \Pr(\mathsf{E}_3 \mid \mathsf{E}_1^c \wedge \mathsf{E}_2^c) \\
&\quad + \Pr\left(\mathtt{PHash}^1_{\Pi_0}(m) \oplus \mathtt{PHash}^1_{\Pi_0}(m') = c\|0^{n-2} \,\middle|\, \mathsf{E}_1^c \wedge \mathsf{E}_2^c \wedge \mathsf{E}_3^c\right) \\
&\leq \frac{4}{2^n} + \frac{16\ell}{2^n} + \frac{8\ell - 4}{2^n} + \frac{2}{2^n} \leq \frac{26\ell}{2^n}
\end{aligned}$$

Same argument shows that $\Pr\left(\mathtt{PHash}^2_{\Pi_0}(m) \oplus \mathtt{PHash}^2_{\Pi_0}(m') = c\|0^{n-2}\right) \leq 26\ell/2^n$.

Corollary 4.

$$\Pr\left(\mathsf{COLL1}^c_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{13\ell q^2}{2^n} \qquad \Pr\left(\mathsf{COLL2}^c_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{13\ell q^2}{2^n}$$

$$\Pr\left(\mathsf{MC1}^{c_1,\ldots,c_s}_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{13\ell q^2}{s \cdot 2^n} \qquad \Pr\left(\mathsf{MC2}^{c_1,\ldots,c_s}_{\mathit{PHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{13\ell q^2}{s \cdot 2^n}$$

The Corollary 4 follows from Lemma 11 by simple application of the Markov's inequality.y

Finally, we bound the auxilliary events

Lemma 12. *We have*

$$\Pr(\mathsf{Aux}_1 \vee \mathsf{Aux}_3) \leq \frac{3\ell q}{2^n - 2} + \frac{2}{2^n} \qquad \Pr(\mathsf{Aux}_2) \leq \frac{\ell^2 q}{2^{n+1}}$$

$$\Pr(\mathsf{Aux}_4) \leq \frac{\ell^2 q^2}{s \cdot 2^n} \qquad \Pr(\mathsf{Aux}_5) \leq \frac{\ell q^2}{s' \cdot 2^n}$$

Combining these bounds we have

$$\Pr(\mathsf{Aux}) \leq \frac{(\ell^2 + 8\ell)q}{2^{n+1}} + \frac{\ell^2 q^2}{s \cdot 2^n} + \frac{\ell q^2}{s' \cdot 2^n}$$

Combining Eq. (29), Lemma 10, Corollary 4 and Lemma 12 we have the following result:

Lemma 13. *$\mathtt{TPHash}_{\Pi_0}$ is a $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CfH where*

$$\epsilon_1(\rho) = \frac{26\ell q^2}{2^n}, \qquad \epsilon_2(\rho, 3) = \frac{16\ell q^2}{2^{2n}}, \qquad \epsilon_3(\rho, s) = \frac{2^s \cdot 13\ell q^2}{s \cdot 2^n}, \qquad \delta(\rho) = \frac{16\ell q^2}{2^{2n}}$$

$$\epsilon_2(\rho, 4) = 8 \cdot \left(\frac{2s'^2}{2^{2n}} + \frac{4s}{2^n} + \frac{2}{2^n} + \frac{2\sqrt{2}q^2}{2^{3n/2}} + \frac{8sq^2}{2^{2n}} + \frac{96q^2}{2^{2n}} + \frac{8q^4}{2^{3n}} \right)$$

6.3 `TLightHash`.

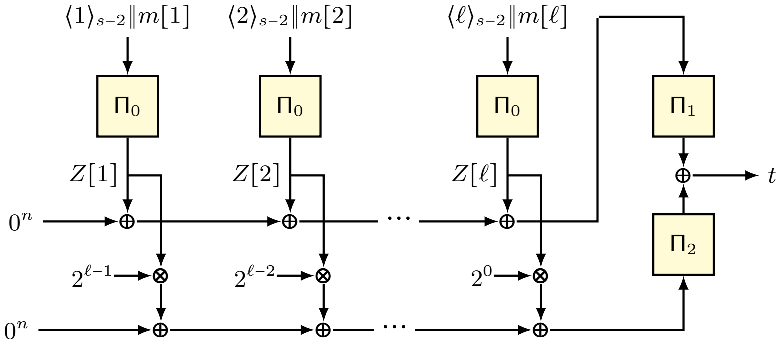


Fig. 5. `1k-LightMAC+`

As before, we let the i-th message be $m^i = m^i[1] \| \cdots \| m^i[\ell_i] \in (\{0,1\}^{n-s})^{\ell_i}$, $i \in [q]$. Note that, $m^i[k] \neq m^j[k] \iff \mathsf{Z}^i[k] \neq \mathsf{Z}^j[k]$ for any $k \in [\max\{\ell_i, \ell_j\}]$, where $\mathsf{Z}^i[k] := \Pi_0(\langle k \rangle_{s-2} \| m^i[k])$. Moreover, $\mathsf{Z}^i[k] \neq \mathsf{Z}^j[k']$ for any $k \neq k'$, $i, j \in [q]$. Let us fix $(i,j) \in ([q])_2$, define $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\} := \{\mathsf{Z}^i[k] : k \in [\ell_i]\} \cup \{\mathsf{Z}^j[k] : k \in [\ell_j]\}$ and partition $[t] := \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{i\bar{j}}$, where

$$\begin{aligned}
\mathtt{I}_{\bar{i}j} &:= \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^i[k'] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j\}]\} \\
\mathtt{I}_{\bar{i}\bar{j}} &:= \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^i[k'] = \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j\}]\} \\
\mathtt{I}_{i\bar{j}} &:= \{k \in [t] : \mathsf{Z}[k] = \mathsf{Z}^j[k'] \neq \mathsf{Z}^i[k'], k' \in [\max\{\ell_i, \ell_j\}]\}
\end{aligned}$$

Then we have

$$\begin{aligned}
&\mathtt{LightHash}^1_{\Pi_0}(m^i) \oplus \mathtt{LightHash}^1_{\Pi_0}(m^j) = c \| 0^{n-2} \\
&\qquad \iff A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = c \| 0^{n-2} \\
&\mathtt{LightHash}^2_{\Pi_0}(m^i) \oplus \mathtt{LightHash}^2_{\Pi_0}(m^j) = c \| 0^{n-2} \\
&\qquad \iff B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = c \| 0^{n-2}
\end{aligned}$$

where

- $A_k = 1$ for $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, $A_k = 0$, otherwise.

- $B_k = 2^\beta$ for some β, if $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, otherwise $B_k = 2^\beta \oplus 2^\gamma$ for some β, γ.

Due to this similarity with $\mathtt{PHash}$, the argument given in Lemma 10 also holds here, giving us

$$\Pr\left(\mathsf{COLL}^{c_1,c_2}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{COLL}^{0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$
$$\Pr\left(\mathsf{AP1}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{AP1}^{0,0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$
$$\Pr\left(\mathsf{AP2}^{c_1,c_2,c_3}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \Pr\left(\mathsf{AP2}^{0,0,0}_{\mathtt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right)$$

Lemma 14. *Assume $\ell \leq 2^n/4$. Then in the ideal world,*

$$\Pr\left(\mathsf{COLL}^{0,0}_{LightHash_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{2q^2}{2^{2n}}$$

Proof. We fix $(i,j) \in ([q])_2$ as above, thus fixing $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\}$ and partitioning $[t] = \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{i\bar{j}}$. We can make the following observations about the index sets:

- $\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} \neq \varnothing$ since otherwise m^i and m^j will not be distinct.
- $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \geq 2$ because otherwise $\mathtt{LightHash}^1_{\Pi_0}(m^i) \neq \mathtt{LightHash}^1_{\Pi_0}(m^j)$.

If we consider the system of linear equations representing the events $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)$ and $\mathtt{LightHash}^2_{\Pi_0}(m^i) = \mathtt{LightHash}^2_{\Pi_0}(m^j)$, respectively:

$$A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = 0^n$$
$$B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = 0^n$$

then the above observations about the index sets imply that there are two distinct indices $k, k' \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$ such that $A_k = A_{k'} = 1$ and $B_k = 2^\beta$, $B_{k'} = 2^\gamma$ for distinct β and γ. This implies that the above system of linear equations has rank 2, and hence will be satisfied with probability $(2^n)_{t-2}/(2^n)_t = 1/(2^n - t + 2)(2^n - t + 1) \leq (2^n - 2\ell + 2)(2^n - 2\ell + 1) \leq 4/2^{2n}$ for $\ell \leq 2^n/4$. Since there are $q(q-1)/2$ tuples $(i,j) \in ([q])_2$, we have our result.

Lemma 15. *Assume that $\ell \leq 2^n/8$. Then in the ideal world, one has,*

$$\Pr\left(\mathsf{AP1}^{0,0,0}_{LightHash_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{q^4}{3 \cdot 2^{3n}} + \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n} + \frac{96q^2}{2^{2n}}$$

Proof. Let us fix $(i,j,r,s) \in ([q])_4$. We want to find the probability of the event

$$\begin{aligned}\mathsf{E}(i,j,r,s) : \left(\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)\right) \\ \wedge \left(\mathtt{LightHash}^2_{\Pi_0}(m^j) = \mathtt{LightHash}^2_{\Pi_0}(m^r)\right) \\ \wedge \left(\mathtt{LightHash}^1_{\Pi_0}(m^r) = \mathtt{LightHash}^1_{\Pi_0}(m^s)\right)\end{aligned}$$

Let $\{\mathsf{Z}[1], \ldots, \mathsf{Z}[t]\} = \{\mathsf{Z}^i[k] : k \in [\ell_i]\} \cup \{\mathsf{Z}^j[k] : k \in [\ell_j]\} \cup \{\mathsf{Z}^r[k] : k \in [\ell_r]\} \cup \{\mathsf{Z}^s[k] : k \in [\ell_s]\}$. Also let us partition $[t]$ in three ways as $[t] = \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} \sqcup \mathtt{I}_{\bar{i}\bar{j}} \sqcup \mathtt{I}_{ij} = \mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}} \sqcup \mathtt{I}_{\bar{j}\bar{r}} \sqcup \mathtt{I}_{jr} = \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}} \sqcup \mathtt{I}_{\bar{r}\bar{s}} \sqcup \mathtt{I}_{rs}$ where

$$\mathtt{I}_{\bar{i}j} := \{k : \mathsf{Z}[k] = \mathsf{Z}^i[k'] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$

$$\mathtt{I}_{i\bar{j}} := \{k : \mathsf{Z}[k] = \mathsf{Z}^j[k'] \neq \mathsf{Z}^i[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$
$$\mathtt{I}_{\bar{i}\bar{j}} := \{k : \mathsf{Z}[k] = \mathsf{Z}^i[k'] = \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$
$$\mathtt{I}_{ij} := \{k : \mathsf{Z}[k] \neq \mathsf{Z}^i[k'], \mathsf{Z}[k] \neq \mathsf{Z}^j[k'], k' \in [\max\{\ell_i, \ell_j, \ell_r, \ell_s\}]\}$$

and the rest of the index sets are defined analogously.

Then the above event can be represented by the following system of equations

$$A_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus A_t \cdot \mathsf{Z}[t] = 0^n$$
$$B_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus B_t \cdot \mathsf{Z}[t] = 0^n$$
$$C_1 \cdot \mathsf{Z}[1] \oplus \cdots \oplus C_t \cdot \mathsf{Z}[t] = 0^n$$

where

- $A_k = 1$ if $k \in \mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}$, and $A_k = 0$ otherwise.
- $B_k = 2^\beta$ for some β if $k \in \mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}}$, $B_k = 2^\beta \oplus 2^\gamma$ for some β, γ if $k \in \mathtt{I}_{\bar{j}\bar{r}}$, and $B_k = 0$ otherwise.
- $C_k = 1$ if $k \in \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}$, and $C_k = 0$ otherwise.

As observed in the proof of Lemma 14, $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \geq 2$ and $|\mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}| \geq 2$. Let us call the coefficient matrix of the above system of equations $M^{(i,j,r,s)}$, its first row as $A^{(i,j,r,s)}$, second row as $B^{(i,j,r,s)}$ and third row as $C^{(i,j,r,s)}$. Let us write $([q])_4$ as union of three index sets, $([q])_4 = \mathtt{J}_1 \sqcup \mathtt{J}_2 \sqcup \mathtt{J}_3$, where $\mathtt{J}_i$ are defined as follows:

$$\mathtt{J}_1 := \{(i,j,r,s) : \mathrm{rank}(M^{(i,j,r,s)}) = 3\}$$
$$\mathtt{J}_2 := \{(i,j,r,s) : A^{(i,j,r,s)} = C^{(i,j,r,s)}\}\}$$
$$\mathtt{J}_2 := \{(i,j,r,s) : B^{(i,j,r,s)} = c_1 A^{(i,j,r,s)} \oplus c_2 C^{(i,j,r,s)} \text{ for } c_1, c_2 \neq 0\}$$

For $(i,j,r,s) \in \mathtt{J}_1$, the probability of the Z-variables satisfying the system of equations is $(2^n)_{t-3}/(2^n)_t \leq 8/2^{3n}$ for $\ell \leq 2^n/8$, since $t \leq 4\ell$. Thus we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_1} \mathsf{E}(i,j,r,s)\right] \leq \frac{q^4}{3 \cdot 2^{3n}} \tag{31}$$

Now let us define the equivalence relation over $([q])_2$ as $(i,j) \sim (r,s)$ if $\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}} = \mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}$. If $(i,j) \sim (r,s)$, then $A^{(i,j,r,s)} = C^{(i,j,r,s)}$, which implies $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j) \iff \mathtt{LightHash}^1_{\Pi_0}(m^r) = \mathtt{LightHash}^1_{\Pi_0}(m^s)$. Suppose the above relations partitions $([q])_2$ into c equivalence classes $([q])_2 = C_1 \sqcup \cdots \sqcup C_c$. For $a = 1, \ldots, c$, consider the events E_a that $\mathtt{LightHash}^1_{\Pi_0}(m^i) = \mathtt{LightHash}^1_{\Pi_0}(m^j)$ for every $(i,j) \in C_a$. Thus from Eq. (30) we have that

$$\Pr[\mathsf{E}_a] \leq \frac{1}{2^n - 2\ell + 1}$$

since $|\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}| \leq 2\ell$ for all $(i,j) \in C_a$. Now we have

$$\Pr\left[\bigvee_{(i,j,r,s)\in\mathtt{J}_2} \mathsf{E}(i,j,r,s)\right] = \Pr\left[\bigvee_{a\in[c]} \bigvee_{(i,j),(r,s)\in C_a} \mathsf{E}(i,j,r,s)\right]$$

$$
\begin{aligned}
&\leq \sum_{a=1}^{c} \Pr\left[\bigvee_{(i,j),(r,s)\in C_a} \mathsf{E}(i,j,r,s) \right] \\
&= \sum_{a=1}^{c} \Pr[\mathsf{E}_a] \cdot \Pr\left(\bigvee_{(i,j),(r,s)\in C_a} \mathtt{LightHash}^2_{\Pi_0}(m^j) = \mathtt{LightHash}^2_{\Pi_0}(m^r) \,\middle|\, \mathsf{E}_a \right) \\
&\leq \sum_{a=1}^{c} \frac{1}{2^n - 2\ell + 1} \cdot \min\left\{ \frac{|C_a|^2}{2(2^n - 2\ell + 1)}, 1 \right\}
\end{aligned}
$$

where the last inequality follows from Eq. (30) and the facts that $A^{(i,j,r,s)}$ and $B^{(i,j,r,s)}$ are linearly independent, and that $|\mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}} \sqcup \mathtt{I}_{\bar{j}\bar{r}}| \leq 2\ell$ for all $(j,r) \in C_a$. Note that $1/(2^n - 2\ell + 1) \leq 2/2^n$ for $\ell \leq 2^n/8$. Subject to the condition that $\sum_{a=1}^{c} |C_a| = \binom{q}{2}$, the sum $\sum_{a=1}^{c} \min\{|C_a|^2/(2(2^n - 2\ell + 1)), 1\}$ is maximized when $c = \left\lfloor \binom{q}{2}/2^{n/2} \right\rfloor + 1$, $|C_a| = 2^{n/2}$ for $a \in [c-1]$ and $|C_c| = \binom{q}{2} - (c-1)2^{n/2}$, in which case we have

$$
\sum_{c=1}^{a} \frac{2}{2^n} \cdot \min\left\{ \frac{|C_a|^2}{2^n}, 1 \right\} \leq \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n}.
$$

Thus we have

$$
\Pr\left[\bigvee_{(i,j,r,s)\in \mathsf{J}_1} \mathsf{E}(i,j,r,s) \right] \leq \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n} \tag{32}
$$

Finally we consider $(i,j,r,s) \in \mathsf{J}_3$. In this case $B^{(i,j,r,s)} = c_1 A^{(i,j,r,s)} + c_2 C^{(i,j,r,s)}$. This linear dependence implies the following:

- $c_1 = 2^\beta$ and $c_2 = 2^\gamma$ for some β, γ.
- $(\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}) \triangle (\mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}) = \mathtt{I}_{\bar{j}r} \sqcup \mathtt{I}_{j\bar{r}}$.[6] Also $B_k, k \in \mathtt{I}_{\bar{j}r}$ are all distinct, and similarly, $B_k, k \in \mathtt{I}_{j\bar{r}}$ are all distinct
- $(\mathtt{I}_{\bar{i}j} \sqcup \mathtt{I}_{i\bar{j}}) \cap (\mathtt{I}_{\bar{r}s} \sqcup \mathtt{I}_{r\bar{s}}) = \mathtt{I}_{\bar{j}\bar{r}}$. From the definition of the index sets, this reduces to $\mathtt{I}_{i\bar{j}} \cap \mathtt{I}_{\bar{r}s} = \mathtt{I}_{\bar{j}\bar{r}}$. If for $k \in \mathtt{I}_{\bar{j}\bar{r}}$, $\mathsf{Z}[k] = \mathsf{Z}^j[k'] = \mathsf{Z}^r[k']$, then $B_k = 2^{\ell_j - k'} + 2^{\ell_r - k'}$. Since $2^a + 2^b = 2^c + 2^d$ implies either $(a,b) = (c,d)$ or $(a,b) = (d,c)$, and since in this case for every $k \in \mathtt{I}_{\bar{j}\bar{r}}$, $B_k = 2^\beta + 2^\gamma$, we have $|\mathtt{I}_{\bar{j}\bar{r}}| = 1$.

Thus the following assumptions made in proof of Lemma 4 of [27] holds true:

- $B^{(i,j,r,s)}$ does not contain the same entry more than twice.
- $B^{(i,j,r,s)}$ contains at least two different non-zero entries.
- Each of $A^{(i,j,r,s)}$ and $C^{(i,j,r,s)}$ contains at least three ones.

The rest of the analysis is exactly the one presented in the proof of Lemma 4 of [27], except the ignorable fact that the coefficient of $\mathsf{Z}^j[k']$ is $2^{\ell_j - k'}$ (instead of $2^{k'}$ as in the [27]), which however makes no changes in the argument presented. Thus following the proof of Lemma 4 of [27], we have

$$
\Pr\left[\bigvee_{(i,j,r,s)\in \mathsf{J}_3} (i,j,r,s) \right] \leq \frac{24q^2}{(2^n - 4\ell + 1)(2^n - 4\ell + 2)} \leq \frac{96q^2}{2^{2n}} \tag{33}
$$

for $\ell \leq 2^n/8$.

Combining Eqs. (31), (32) and (33) we have our result.

[6] For two sets A, B, we denote their symmetric difference as $A \triangle B := (A \setminus B) \cup (B \setminus A)$

The probability analysis of the events $\mathsf{AP2}^{c_1,c_2,c_3}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{AP1}^{c_1,c_2}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ are similar to the analysis of the events $\mathsf{AP1}^{c_1,c_2,c_3}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})$ and $\mathsf{COLL}^{c_1,c_2}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})$, respectively, and we get the same probability bounds.

The exact same arguments given to prove Lemma 11 can be used to prove the following statement, keeping in mind that we do not need to consider the events E_1 and E_2, described in the proof of Lemma 11, for `LightHash`:

Lemma 16. *For $\ell \leq 2^{n-2}$, $m \neq m' \in (\{0,1\}^{n-2})^{\leq \ell}$, and $c \in \{0,1\}^2$, we have*

$$\Pr\left(\texttt{LightHash}^1_{\Pi_0}(m) \oplus \texttt{LightHash}^1_{\Pi_0}(m') = c\|0^{n-2}\right) \leq \frac{8\ell}{2^n}$$
$$\Pr\left(\texttt{LightHash}^2_{\Pi_0}(m) \oplus \texttt{LightHash}^2_{\Pi_0}(m') = c\|0^{n-2}\right) \leq \frac{8\ell}{2^n}$$

Corollary 5.

$$\Pr\left(\mathsf{COLL1}^{c}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{4\ell q^2}{2^n} \qquad \Pr\left(\mathsf{COLL2}^{c}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{4\ell q^2}{2^n}$$
$$\Pr\left(\mathsf{MC1}^{c_1,\ldots,c_s}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{4\ell q^2}{s \cdot 2^n} \qquad \Pr\left(\mathsf{MC2}^{c_1,\ldots,c_s}_{\texttt{LightHash}_{\Pi_0}}(\boldsymbol{m})\right) \leq \frac{4\ell q^2}{s \cdot 2^n}$$

Thus we get our desired result:

Lemma 17. *$\texttt{TLightHash}_{\Pi_0}$ is a $(\epsilon_1, \epsilon_2, \epsilon_3, \delta)$-CfH, where*

$$\epsilon_1(\rho) = \frac{8\ell q^2}{2^n}, \qquad \epsilon_2(\rho,3) = \frac{8q^2}{2^{2n}}, \qquad \epsilon_3(\rho,s) = \frac{2^s \cdot 4\ell q^2}{s \cdot 3^n}, \qquad \delta(\rho) = \frac{8q^2}{2^{2n}}$$
$$\epsilon_2(\rho,4) = 8 \cdot \left(\frac{q^4}{3 \cdot 2^{3n}} + \frac{q^2}{2 \cdot 2^{3n/2}} + \frac{2}{2^n} + \frac{96q^2}{2^{2n}}\right)$$

7 PRF Security of Sum of k Even-Mansour

For any $r \geq 2$, let $(\pi_1, \ldots, \pi_r) \twoheadleftarrow \mathcal{P}(n)^r$ be a tuple of r permutations of $\{0,1\}^n$ and let $(K_1, \ldots, K_r) \in (\{0,1\}^n)^r$ be a r-tuple of n-bit strings.

One-round Even-Mansour construction is a keyed permutation of $\{0,1\}^n$ defined by the mapping

$$x \longmapsto \pi_1(x \oplus K_1) \oplus K_1,$$

where K_1 denotes the key.

The r-sum of Even-Mansour construction, π-$\texttt{SOEM}^r$ is a length-preserving keyed function of $\{0,1\}^n$ defined by the mapping

$$m \longmapsto \bigoplus_{i=1}^{r} \pi_i(m \oplus K_i),$$

where $K = (K_1, \ldots, K_r)$ denotes the key. See Figure 6 for a pictorial illustration. Notice that we skipped the post-permutation key masking. This is motivated by a similar simplification [35] by Sibleyras and Todo who studied the $r = 2$ case. Thus, we study the same problem for any arbitrary $r \geq 2$.

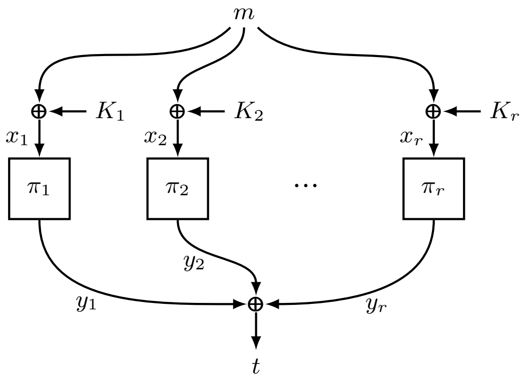


Fig. 6. The π-$\mathtt{SOEM}^r$ construction instantiated with key $K = (K_1, \ldots, K_r)$.

Theorem 4. *Fix some $r \geq 2$, $q + p \leq 2^{\frac{r}{r+1}n - \log_2(n)}$, and $\Pi = (\Pi_1, \ldots, \Pi_r) \twoheadleftarrow \mathcal{P}(n)^r$. For any (q,p)-distinguisher $\mathcal{A}$ we have*

$$\mathbf{Advt}^{\mathrm{prf}}_{\Pi\text{-}\mathtt{SOEM}^r}(\mathcal{A}) \leq \frac{1}{2^n} + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{20\sqrt{nq}(2p+2q)^{r-1}}{2^{n(r-1)}} + \frac{10q(2p+2q)^r}{2^{nr}}.$$

Proof. For the purpose of this proof let $\mathsf{F}_{\mathsf{K}}(\cdot) = \Pi\text{-}\mathtt{SOEM}^r_{\mathsf{K}}(\cdot)$, and let $\Gamma \twoheadleftarrow \{0,1\}^n$. $\mathcal{A}$'s goal is to distinguish between the *real* oracle $(\mathsf{F}_{\mathsf{K}}, \Pi^{\pm})$ and the *ideal* oracle $(\Gamma, \Pi^{\pm})$, where F_{K} and Γ are referred as the construction oracle and $\Pi^{\pm}$ is referred as the primitive oracle.

Fix a (q,p)-distinguisher $\mathcal{A}$. Let

- $(\mathsf{M}^i, \mathsf{T}^i)$ denote the i-th query-response tuple corresponding to the construction oracle. Let $\mathsf{M} := \{\mathsf{M}^i : i \in [q]\}$ and $\mathsf{T} := \{\mathsf{T}^i : i \in [q]\}$.
- $(\mathsf{U}^i_j, \mathsf{V}^i_j)$ denote the i-th query-response tuple corresponding to the permutation Π_j. Unless stated otherwise, we assume that all these queries are in the forward direction. Let $\mathsf{U}_j := \{\mathsf{U}^i_j : i \in [p]\}$, $\mathsf{V}_j := \{\mathsf{V}^i_j : i \in [p]\}$, $\mathsf{U} := (\mathsf{U}_1, \ldots, \mathsf{U}_r)$, and $\mathsf{V} := (\mathsf{V}_1, \ldots, \mathsf{V}_r)$.
- $(\mathsf{X}^i_j, \mathsf{Y}^i_j)$ denote the input-output tuple to the j-th permutation, for all $j \in [r]$, within the i-th construction query in the real world, i.e., $\mathsf{X}^i_j = \mathsf{M}^i \oplus \mathsf{K}_j$. Let $\mathsf{X}^i := (\mathsf{X}^i_j : j \in [r])$ and $\mathsf{Y}^i := (\mathsf{Y}^i_j : j \in [r])$. Let $\mathsf{X} := \{\mathsf{X}^i : i \in [q]\}$ and $\mathsf{Y} := \{\mathsf{Y}^i : i \in [q]\}$.

We study a modified game where the real oracle releases (X, Y) to $\mathcal{A}$ once the query-response phase is over, but before $\mathcal{A}$ outputs. This obviously does not decrease $\mathcal{A}$'s advantage.

Ideal World Transcript Extension: Naturally, in the ideal world, the sampling is extended to generate this additional information. We have

$$\mathcal{SC}(\mathsf{T}, \mathsf{V}) = \left\{ (\mathsf{T}^i, \mathsf{V}^{j_1}_1, \mathsf{V}^{j_2}_2, \ldots, \mathsf{V}^{j_r}_r) \in \mathsf{T} \times \mathsf{V} : \bigoplus_{k=1}^{r} \mathsf{V}^{j_k}_k = \mathsf{T}^i \right\}$$
$$\mu(\mathsf{T}, \mathsf{V}) = |\mathcal{SC}(\mathsf{T}, \mathsf{V})|$$

Further due to the increasing nature of $\mu^r(\mathsf{T},\cdot)$, $\mu(\mathsf{T},\mathsf{V}) \le \mu^r(\mathsf{T},p+q)$. We define the predicate

$$\mathsf{LSC}(\mathsf{T},p+q): \left(\mu^r(\mathsf{T},p+q) > \frac{q(p+q)^r}{2^n} + 4(p+q)^{r-1}\sqrt{nq}\right)$$

The subsequent two-step sampling mechanism for (X,Y) in the ideal world is defined under the condition that $\neg\mathsf{LSC}(\mathsf{T},p+q)$ holds:

1. In the first step, a dummy key tuple is sampled uniformly at random, i.e., $\mathsf{K} \twoheadleftarrow (\{0,1\}^n)^r$, which determines $\mathsf{X}^i_j := \mathsf{M}^i \oplus \mathsf{K}_j$. Consider the following predicates:

$$\mathsf{KG}(\mathsf{M},\mathsf{U},\mathsf{K}): \exists\, i \in [q],\, j_1,\ldots,j_r \in [p] \text{ such that } \left(\forall k \in [r],\, \mathsf{X}^i_k = \mathsf{U}^{j_k}_k\right)$$

$$\mathsf{SC}(\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K}): \exists\,(i,j_1,j_2,\ldots,j_r) \in \mathcal{SC}(\mathsf{T},\mathsf{V}),\, k \in [r], \text{ such that } (\mathsf{X}^i_k \ne \mathsf{U}^{j_k}_k) \text{ and } \left(\forall k' \ne k,\, \mathsf{X}^i_{k'} = \mathsf{U}^{j_{k'}}_{k'}\right)$$

Going forward we assume that $\neg(\mathsf{KG}(\mathsf{M},\mathsf{U},\mathsf{K}) \vee \mathsf{SC}(\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K}))$ holds. For each $i \in [q]$:

(a) if there exists $j \in [p]$ and $k \in [r]$, such that $\mathsf{X}^i_k = \mathsf{U}^j_k$, then define $\mathsf{Y}^i_k := \mathsf{V}^j_k$;

(b) let $\mathcal{I}_i = \{j \in [r] : \mathsf{X}^i_j \notin \mathsf{U}_j\}$ to be the set of permutation indices with fresh input for the i-th construction query.

(c) let $\sim$ be a relation on $[q]$ defined as: $i_1 \sim i_2 \iff \mathcal{I}_{i_1} = \mathcal{I}_{i_2}$. Clearly, $\sim$ is an equivalence relation. Let $\mathcal{Q}^{(1)}_{(0)} \sqcup \ldots \mathcal{Q}^{(r)}_{(0)} \sqcup \mathcal{Q}_{(1)} \sqcup \ldots \sqcup \mathcal{Q}_{(c)}$ denote the corresponding partitioning of $[q]$, where $\mathcal{Q}^{(j)}_{(0)} = \{i \in [q] : \mathcal{I}_i = \{j\}\}$. Let $|\mathcal{Q}^{(j)}_{(0)}| = q^{(j)}_0$, $q_0 := \sum_{j\in[r]} q^{(j)}_0$ and $|\mathcal{Q}_{(i)}| = q_i$. Then $q_0 + \sum_{i\in[c]} q_i = q$. Also, note that, $c \le \sum_{j=2}^{r-1}\binom{r}{j} \le 2^r$.

(d) for all $j \in [r]$ and $i \in \mathcal{Q}^{(j)}_{(0)}$, define $\mathsf{Y}^i_j := \oplus_{l\in[r]\setminus j}\mathsf{Y}^i_l \oplus \mathsf{T}^i$ and

$$\mathsf{Y}^{(0)} = \{\mathsf{Y}^i_j \oplus_{l\in[r]\setminus j} \mathsf{Y}^i_l \oplus \mathsf{T}^i \;:\; j \in [r], i \in \mathcal{Q}^{(j)}_{(0)}\}.$$

This concludes the first step. We encourage the readers to verify that at the end of this step Y^i_j is undefined for exactly the indices in $\mathcal{I}_i$ and $|\mathcal{I}_i| \ge 2$. Furthermore, due to $\neg(\mathsf{KG}(\mathsf{MU},\mathsf{K}) \vee \mathsf{SC}(\mathsf{M},\mathsf{T},\mathsf{U},\mathsf{V},\mathsf{K}))$, the partially defined (X,Y) is permutation-consistent.

Constrained system formulation: For each $i \in [c]$, let $\mathcal{J}_{(i)} = \{j_1,\ldots,j_{t_i}\}$ denote the set of permutation indices with fresh input for the i-th equivalence class $\mathcal{Q}_{(i)}$. Let $r_i = q_i t_i$.

Then, for each $i \in [c]$, we obtain a (q_i, r_i, t_i)-constrained system $\mathbb{S}^{(i)}$:

$$\mathbb{S}^{(i)} = \left\{ \bigoplus_{k\in\mathcal{J}_{(i)}} \mathsf{Y}^j_k = \mathsf{T}^j \bigoplus_{k'\in[r]\setminus\mathcal{J}_{(i)}} \mathsf{Y}^j_{k'} \right\}_{j\in\mathcal{Q}_{(i)}}$$

which is binary, acyclic, partite, isolate and t_i-regular.

2. In the second step, we sample a solution for each of the c constrained systems. First fix any arbitrary ordering of $\mathbb{S}^{(1)}, \ldots, \mathbb{S}^{(c)}$. Now, for the i-th system:
   - let $\mathcal{R}^{(j)}_{\le(i-1)} = \mathsf{V}_j \cup \{\mathsf{Y}^k_j \ : \ k \in \mathcal{Q}^{(j)}_{(0)}\} \cup \{\mathsf{Y}^k_j \ : \ k \in \mathcal{Q}_{(1)} \sqcup \ldots \sqcup \mathcal{Q}_{(i-1)}\}$, for all $j \in [r]$, and let $|\mathcal{R}^{(j)}_{\le(i-1)}| = r^{(j)}_{\le(i-1)} \le (p+q)$,
   - let $\mathcal{R}_{\le(i-1)} = (\mathcal{R}^{(j)}_{\le(i-1)} \ : \ j \in [r])$ and $\widehat{\mathcal{R}}_{\le(i-1)} = (\mathcal{R}^{(j)}_{\le(i-1)} \ : \ j \in \mathcal{J}_{(i)})$,
   - let $\mathsf{T}^{(i)} = (\mathsf{T}^k \ : \ k \in \mathcal{Q}_{(i)})$ and $\widehat{\mathsf{T}}^{(i)} = (\mathsf{T}^k \bigoplus_{j \in [r] \setminus \mathcal{J}_{(i)}} \mathsf{Y}^k_j \ : \ k \in \mathcal{Q}_{(i)})$. Then, $|\mathsf{T}^{(i)}|, |\widehat{\mathsf{T}}^{(i)}| \le q_i$.
   - let $\mathsf{Y}^{(i)} = \{\mathsf{Y}^k_j \ : \ k \in \mathcal{Q}_{(i)}, j \in \mathcal{J}_{(i)}\}$. Then, $|\mathsf{Y}^{(i)}| = r_i$.

   We sample $\mathsf{Y}^{(i)} \twoheadleftarrow (\mathbb{S}^{(i)} \,|\, \widehat{\mathcal{R}}_{\le(i-1)})$, where using Theorem 1, we have

$$\eta(\mathbb{S}^{(i)} \,|\, \widehat{\mathcal{R}}_{\le(i-1)}) \ge \frac{\prod_{j \in \mathcal{J}_{(i)}} (2^n - r^{(j)}_{\le(i-1)})_{q_i}}{2^{nq_i}} \left(1 - \varepsilon^{(i)}\right) \tag{34}$$

$$\varepsilon^{(i)} \le \frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\le(i-1)})}{2^{n(t_i-1)}} + \frac{2q_i \varDelta_{\mathbb{S}^{(i)}}}{2^{n(t_i-1)}} + \frac{6q_i(p+q)^{t_i}}{2^{nt_i}} \tag{35}$$

   Since the solution for each system is sampled in a consistent manner given a consistent solution for the previous system, the cumulative sampling is also permutation-compatible. This completes the second step.

At this stage the full transcript in the ideal world, i.e., $\Theta_{\mathrm{id}} = (\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}, \mathsf{Y})$ is fully determined.

Some Notations on Transcripts: For any wo $\in$ {re, id}, and $\Theta_{\mathrm{wo}} = (\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}, \mathsf{Y})$, let:

- $\Theta^{\mathsf{key}}_{\mathrm{wo}}$ denote the restriction of Θ_{wo} to the key K,
- $\Theta^{\mathsf{con}}_{\mathrm{wo}}$ denote the restriction of Θ_{wo} to the construction query-response tuple (M, T),
- $\Theta^{\mathsf{prim}}_{\mathrm{wo}}$ denote the restriction of Θ_{wo} to the key (U, V),
- $\Theta^{\mathsf{int}}_{\mathrm{wo}}$ denote the restriction of Θ_{wo} to the construction-specific primitive query-response (X, Y).

Bad Transcript Definition and Analysis: A transcript $\omega = (M, T, U, V, K, Y) \in \varOmega$ is said to be *bad* if and only if $\mathsf{LSC}(T, p{+}q) \vee \mathsf{KG}(M, U, K) \vee \mathsf{SC}(M, T, U, V, K)$ holds.

Lemma 18.

$$\Pr\left(\Theta_{\mathrm{id}} \in \varOmega_{\mathrm{bad}}\right) \le \frac{1}{2^n} + \frac{4\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}} + \frac{2q(p+q)^r}{2^{nr}}$$

Proof. We have

$$\begin{aligned}\Pr\left(\Theta_{\mathrm{id}} \in \varOmega_{\mathrm{bad}}\right) &= \Pr\left(\mathsf{LSC}(\mathsf{T}, p+q) \vee \mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}) \vee \mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K})\right) \\ &\le \Pr\left(\mathsf{LSC}(\mathsf{T}, p+q)\right) + \Pr\left(\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K})\right) + \Pr\left(\mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}) \mid \neg\mathsf{LSC}(\mathsf{T}, p+q)\right) \\ &\le \frac{1}{2^n} + \frac{qp^r}{2^{nr}} + \frac{q(p+q)^r}{2^{nr}} + \frac{4(p+q)^{r-1}\sqrt{nq}}{2^{n(r-1)}},\end{aligned}$$

where the first term on the right hand side corresponds to $\Pr(\mathsf{LSC}(\mathsf{T}, p+q))$ and follows from Lemma 1, the second term corresponds to $\Pr(\mathsf{KG}(\mathsf{M}, \mathsf{U}, \mathsf{K}))$ and follows from the uniformity of K. The last two terms correspond to $\Pr(\mathsf{SC}(\mathsf{M}, \mathsf{T}, \mathsf{U}, \mathsf{V}, \mathsf{K}) \mid \neg\mathsf{LSC}(\mathsf{T}, p+q))$. To argue this, first notice that given $\neg\mathsf{LSC}(\mathsf{T}, p+q)$, we have

$$\mu(\mathsf{T}, \mathsf{V}) \leq \frac{q(p+q)^r}{2^n} + 4(p+q)^{r-1}\sqrt{nq}.$$

For each choice of $k \in [r]$, the predicate $\forall k' \neq k$, $\mathsf{X}^i_{k'} = \mathsf{U}^{j_{k'}}_{k'}$ is satisfied with at most $2^{-n(r-1)}$ probability. Now, we get the desired terms using union bound. □

Good Transcript Analysis: Let $\omega = (M, T, U, V, K, Y)$ be a good transcript. Since the transcript is good, $\neg(\mathsf{LSC}(T, p+q) \vee \mathsf{KG}(M, U, K) \vee \mathsf{SC}(M, T, U, V, K))$ holds.

Before moving forward, recall the notations introduced while discussing the sampling in the ideal world. We assume analogous notations for any arbitrary transcript.

We also ignore the probability computation of obvious events, such as: the message tuple being realized.

Real World: In the real world, we have

$$\begin{aligned}\Pr(\Theta_{\text{re}} = \omega) &= \Pr\left(\Theta^{\text{key}}_{\text{re}} = K, \Theta^{\text{prim}}_{\text{re}} = (U, V), \Theta^{\text{int}}_{\text{re}} = (X, Y), \Theta^{\text{con}}_{\text{re}} = (M, T)\right) \\ &= \Pr\left(\Theta^{\text{key}}_{\text{re}} = K\right) \times \Pr\left(\Theta^{\text{prim}}_{\text{re}} = (U, V)\right) \times \Pr\left(\Theta^{\text{int}}_{\text{re}} = (X, Y) \,\middle|\, \Theta^{\text{key}}_{\text{re}}, \Theta^{\text{prim}}_{\text{re}}\right) \\ &= \frac{1}{2^{nr}} \times \frac{1}{(2^n)^r_p} \times \Pr\left(\Theta^{\text{int}}_{\text{re}} = (X, Y) \,\middle|\, \Theta^{\text{key}}_{\text{re}}, \Theta^{\text{prim}}_{\text{re}}\right),\end{aligned}$$

where the first term on the right hand side follows from the uniformity of K, the second term follows from the uniformity of $\Pi = (\Pi_1, \ldots, \Pi_r)$.

As for the last term, consider the partition imposed by $\sim$ in an arbitrary order, and also the associated notations introduced earlier. Then, conditioned on $(\Theta^{\text{key}}_{\text{re}}, \Theta^{\text{prim}}_{\text{re}})$, we have

$$\Pr\left(\Theta^{\text{int}}_{\text{re}} = (X, Y) \,\middle|\, \Theta^{\text{key}}_{\text{re}}, \Theta^{\text{prim}}_{\text{re}}\right) = \prod_{j=1}^{r} \frac{1}{(2^n - p)_{q_0^{(j)}}} \times \prod_{\substack{i \in [c] \\ j' \in \mathcal{J}_{(i)}}} \frac{1}{(2^n - r^{(j')}_{\leq(i-1)})_{q_i}}.$$

Indeed, the first product term corresponds to the query indices with exactly one fresh primitive input, i.e. the ones in $\mathcal{Q}^{(j)}_{(0)}$ for some $j \in [r]$, and the second product correspond to the query indices with at least two fresh primitive inputs, computed using a simple application of chain rule over the partitions $\mathcal{Q}_{(1)}, \ldots, \mathcal{Q}_{(c)}$. By combining everything, we have

$$\Pr(\Theta_{\text{re}} = \omega) = \frac{1}{2^{nr}} \times \frac{1}{(2^n)^r_p} \times \prod_{j=1}^{r} \frac{1}{(2^n - p)_{q_0^{(j)}}} \times \prod_{\substack{i \in [c] \\ j' \in \mathcal{J}_{(i)}}} \frac{1}{(2^n - r^{(j')}_{\leq(i-1)})_{q_i}}, \tag{36}$$

Ideal World: In the ideal world, we have

$$\begin{aligned}\Pr\left(\Theta_{\mathrm{id}} = \omega\right) &= \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{key}} = K, \Theta_{\mathrm{id}}^{\mathrm{prim}} = (U,V), \Theta_{\mathrm{id}}^{\mathrm{int}} = (X,Y), \Theta_{\mathrm{id}}^{\mathrm{con}} = (M,T)\right)\\ &= \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{key}} = K\right) \times \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{con}} = (M,T)\right) \times \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{prim}} = (U,V)\right)\\ &\qquad \times \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{int}} = (X,Y) \,\middle|\, \Theta_{\mathrm{id}}^{\mathrm{key}}, \Theta_{\mathrm{id}}^{\mathrm{prim}}, \Theta_{\mathrm{id}}^{\mathrm{con}}\right)\\ &= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \Pr\left(\Theta_{\mathrm{id}}^{\mathrm{int}} = (X,Y) \,\middle|\, \Theta_{\mathrm{id}}^{\mathrm{key}}, \Theta_{\mathrm{id}}^{\mathrm{prim}}, \Theta_{\mathrm{id}}^{\mathrm{con}}\right)\\ &= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{i\in[c]} \Pr\left(\mathsf{Y}^{(i)} = Y^{(i)} \,\middle|\, \widehat{\mathcal{R}}_{\le(i-1)}\right)\\ &= \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{i\in[c]} \frac{1}{\eta(\mathbb{S}^{(i)} \,|\, \widehat{\mathcal{R}}_{\le(i-1)})}\end{aligned}$$

where the first three terms are obvious. The fourth term corresponds to the indices in $\mathcal{Q}_{(i)}$ for all $i \in [c]$. Further, using (34), we have

$$\begin{aligned}\Pr\left(\Theta_{\mathrm{id}} = \omega\right) &\ge \frac{1}{2^{nr}} \times \frac{1}{2^{nq}} \times \frac{1}{(2^n)_p^r} \times \prod_{\substack{i\in[c]\\ j'\in[r]}} \frac{2^{nq_i}}{\left(1-\varepsilon^{(i)}\right)(2^n - r_{\le(i-1)}^{(j')})_{q_i}}\\ &= \frac{1}{2^{nr}} \times \frac{1}{2^{nq_0}} \times \frac{1}{(2^n)_p^r} \times \prod_{\substack{i\in[c]\\ j'\in[r]}} \frac{1}{\left(1-\varepsilon^{(i)}\right)(2^n - r_{\le(i-1)}^{(j')})_{q_i}}, \qquad (37)\end{aligned}$$

where the equality follows from the fact that $q = q_0 \sum_{i\in[c]} q_i$.

The Ratio: On dividing (36) by (37), we have

$$\begin{aligned}\frac{\Pr\left(\Theta_{\mathrm{re}} = \omega\right)}{\Pr\left(\Theta_{\mathrm{id}} = \omega\right)} &\ge \prod_{i\in[c]} \left(1-\varepsilon^{(i)}\right) \qquad (38)\\ &\ge 1 - \sum_{i\in[c]} \varepsilon^{(i)}\\ &\ge 1 - \underbrace{\sum_{i\in[c]} \left(\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\le(i-1)})}{2^{n(t_i-1)}} + \frac{2q_i \varDelta_{\mathbb{S}^{(i)}}}{2^{n(t_i-1)}} + \frac{6q_i(p+q)^{t_i}}{2^{nt_i}}\right)}_{\varepsilon_{\mathrm{ratio}}(\omega)}. \qquad (39)\end{aligned}$$

Now, we have

$$\begin{aligned}\mathbb{E}\left(1_{\mathrm{good}}\varepsilon_{\mathrm{ratio}}\right) &= \sum_{i\in[c]} \mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}})\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\le(i-1)})}{2^{n(t_i-1)}}\right) + \sum_{i\in[c]} \frac{2\mathbb{E}\left(q_i\right)\mathbb{E}\left(\varDelta_{\mathbb{S}^{(i)}}\right)}{2^{n(t_i-1)}} + \sum_{i\in[c]} \frac{6\mathbb{E}\left(q_i\right)(p+q)^{t_i}}{2^{nt_i}}\\ &\qquad (40)\\ &\le \sum_{i\in[c]} \mathbb{E}\left(1_{\mathrm{good}}(\Theta_{\mathrm{id}})\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\le(i-1)})}{2^{n(t_i-1)}}\right) + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{6q(2(p+q))^r}{2^{nr}}\\ &\qquad (41)\end{aligned}$$

where the first equality follows from linearity of expectation and the fact that $\mathbb{E}\left(\chi \mathtt{R}\right) \leq \mathbb{E}\left(\mathtt{R}\right)$ for any non-negative random variable $\mathtt{R}$ and indicator random variable χ. The second/third term in the second inequality follows from $\mathbb{E}\left(q_i\right) \leq qp^{r-t_i}/2^{n(r-t_i)} \leq q(p+q)^{r-t_i}/2^{n(r-t_i)}$, $t_i \geq 2$, $c \leq 2^r$. Additionally, due to the uniformity of T and $q < 2^n$, $\mathbb{E}\left(\Delta_{\mathbb{S}(i)}\right) \leq 4n$. Now, for the first term, when $t_i = r$, we have

$$\begin{aligned}
\mathbb{E}\left(1_{\text{good}}(\Theta_{\text{id}})\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(r-1)}}\right) &\leq \frac{2\mu(\mathsf{T}, \mathsf{V})}{2^{n(r-1)}} \\
&\leq \frac{2\mu^r(\mathsf{T}, p+q)}{2^{n(r-1)}} \\
&\leq \frac{2q(p+q)^r}{2^{nr}} + \frac{8\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}}, \qquad (42)
\end{aligned}$$

where the last inequality follows from $1_{\text{good}}(\Theta_{\text{id}}) = 1$. For, $t_i < r$, let $\mathcal{J}_{(i)} = \{j_1, \ldots, j_{t_i}\}$, $[r] \setminus \mathcal{J}_{(i)} = \{j'_1, \ldots, j'_{r-t_i}\}$, and

$$\mathcal{KSC}_{(i)} := \left\{(\mathsf{T}^{i'}, \mathsf{V}^{k_1}_{j'_1}, \ldots, \mathsf{V}^{k_{r-t_i}}_{j'_{r-t_i}}, \mathsf{Z}_{\mathcal{J}_{(i)}}) \in \mathcal{SC}(\mathsf{T}, \mathsf{V}_{[r]\setminus\mathcal{J}_{(i)}}, \mathcal{R}^{(\mathcal{J}_{(i)})}_{\leq(i-1)}) \;:\; \mathsf{X}^{i'}_{j'_l} = \mathsf{U}^{k_l}_{j'_l}\right\}$$

Then, $|\mathcal{KSC}_{(i)}| = \mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})$, and thus

$$\begin{aligned}
\mathbb{E}\left(1_{\text{good}}(\Theta_{\text{id}})\frac{2\mu(\widehat{\mathsf{T}}^{(i)}, \widehat{\mathcal{R}}_{\leq(i-1)})}{2^{n(t_i-1)}}\right) &\leq \frac{2}{2^{n(t_i-1)}}\mathbb{E}\left(1_{\text{good}}(\Theta_{\text{id}})|\mathcal{KSC}_{(i)}|\right) \\
&\leq \frac{2}{2^{n(t_i-1)}} \times \frac{\mu^r(\mathsf{T}, p+q)}{2^{n(r-t_i)}} \\
&\leq \frac{2q(p+q)^r}{2^{nr}} + \frac{8\sqrt{nq}(p+q)^{r-1}}{2^{n(r-1)}} \qquad (43)
\end{aligned}$$

where the second inequality follows from the uniformity of K, and the last inequality follows from $1_{\text{good}}(\Theta_{\text{id}}) = 1$. Using (42) and (43) in (41), we have

$$\mathbb{E}\left(1_{\text{good}}\varepsilon_{\text{ratio}}\right) \leq \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{16\sqrt{nq}(2(p+q))^{r-1}}{2^{n(r-1)}} + \frac{8q(2(p+q))^r}{2^{nr}} \qquad (44)$$

Finally, using the fine-grained variant of the Expectation method (see Lemma 2) along with Lemma 18 and (44), we have

$$\mathbf{Advt}^{\text{prf}}_{\Pi\text{-}\mathtt{SOEM}^r}(\mathcal{A}) \leq \frac{1}{2^n} + \frac{16nq(2p)^{r-2}}{2^{n(r-1)}} + \frac{20\sqrt{nq}(2p+2q)^{r-1}}{2^{n(r-1)}} + \frac{10q(2p+2q)^r}{2^{nr}},$$

which completes the proof. □

Remark 1. We remark that a similar bound is also possible via the usual Expectation method with an additional $qp^r/2^{rn}$ term.

References

1. Andreeva, E., Daemen, J., Mennink, B., Assche, G.V.: Security of keyed sponge constructions using a modular proof approach. In: Leander, G. (ed.) Fast Software Encryption - FSE 2015, Revised Selected Papers. Lecture Notes in Computer Science, vol. 9054, pp. 364–384. Springer (2015). https://doi.org/10.1007/978-3-662-48116-5_18
2. Banik, S., Pandey, S.K., Peyrin, T., Sasaki, Y., Sim, S.M., Todo, Y.: GIFT: A small present - towards reaching the limit of lightweight encryption. In: Fischer, W., Homma, N. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2017, Proceedings. Lecture Notes in Computer Science, vol. 10529, pp. 321–345. Springer (2017). https://doi.org/10.1007/978-3-319-66787-4_16
3. Bellare, M., Canetti, R., Krawczyk, H.: Keying hash functions for message authentication. In: Koblitz, N. (ed.) Advances in Cryptology - CRYPTO 1996, Proceedings. Lecture Notes in Computer Science, vol. 1109, pp. 1–15. Springer (1996). https://doi.org/10.1007/3-540-68697-5_1
4. Bellare, M., Impagliazzo, R.: A tool for obtaining tighter security analyses of pseudorandom function based constructions, with applications to PRP to PRF conversion. IACR Cryptol. ePrint Arch. p. 24 (1999)
5. Bellare, M., Kilian, J., Rogaway, P.: The security of the cipher block chaining message authentication code. J. Comput. Syst. Sci. **61**(3), 362–399 (2000). https://doi.org/10.1006/JCSS.1999.1694
6. Bernstein, D.J., Kölbl, S., Lucks, S., Massolino, P.M.C., Mendel, F., Nawaz, K., Schneider, T., Schwabe, P., Standaert, F., Todo, Y., Viguier, B.: Gimli : A cross-platform permutation. In: Fischer, W., Homma, N. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2017, Proceedings. Lecture Notes in Computer Science, vol. 10529, pp. 299–320. Springer (2017). https://doi.org/10.1007/978-3-319-66787-4_15
7. Bertoni, G., Daemen, J., Hoffert, S., Peeters, M., Assche, G.V., Keer, R.V.: Farfalle: parallel permutation-based cryptography. IACR Trans. Symmetric Cryptol. **2017**(4), 1–38 (2017)
8. Bertoni, G., Daemen, J., Peeters, M., Van Assche, G.: Keccak. In: Johansson, T., Nguyen, P.Q. (eds.) Advances in Cryptology – EUROCRYPT 2013, Proceedings. pp. 313–314 (2013)
9. Bierbrauer, J., Johansson, T., Kabatianskii, G., Smeets, B.J.M.: On families of hash functions via geometric codes and concatenation. In: Stinson, D.R. (ed.) Advances in Cryptology - CRYPTO 1993, Proceedings. Lecture Notes in Computer Science, vol. 773, pp. 331–342. Springer (1993). https://doi.org/10.1007/3-540-48329-2_28
10. Black, J., Rogaway, P.: A block-cipher mode of operation for parallelizable message authentication. In: Knudsen, L.R. (ed.) Advances in Cryptology - EUROCRYPT 2002, Proceedings. Lecture Notes in Computer Science, vol. 2332, pp. 384–397. Springer (2002). https://doi.org/10.1007/3-540-46035-7_25
11. den Boer, B.: A simple and key-economical unconditional authentication scheme. J. Comput. Secur. **2**, 65–72 (1993)
12. Bogdanov, A., Knezevic, M., Leander, G., Toz, D., Varici, K., Verbauwhede, I.: spongent: A lightweight hash function. In: Preneel, B., Takagi, T. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2011 - 13th International Workshop, Nara, Japan, September 28 - October 1, 2011. Proceedings. Lecture Notes in Computer Science, vol. 6917, pp. 312–325. Springer (2011). https://doi.org/10.1007/978-3-642-23951-9_21

13. Bogdanov, A., Knudsen, L.R., Leander, G., Paar, C., Poschmann, A., Robshaw, M.J.B., Seurin, Y., Vikkelsoe, C.: PRESENT: an ultra-lightweight block cipher. In: Paillier, P., Verbauwhede, I. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2007, Proceedings. Lecture Notes in Computer Science, vol. 4727, pp. 450–466. Springer (2007). https://doi.org/10.1007/978-3-540-74735-2_31
14. Chen, Y.L., Lambooij, E., Mennink, B.: How to build pseudorandom functions from public random permutations. In: Boldyreva, A., Micciancio, D. (eds.) Advances in Cryptology - CRYPTO 2019, Proceedings, Part I. Lecture Notes in Computer Science, vol. 11692, pp. 266–293. Springer (2019). https://doi.org/10.1007/978-3-030-26948-7_10
15. Cogliati, B., Dutta, A., Nandi, M., Patarin, J., Saha, A.: Proof of mirror theory for a wide range of $\xi _{\max }$. In: Hazay, C., Stam, M. (eds.) Advances in Cryptology - EUROCRYPT 2023, Proceedings, Part IV. Lecture Notes in Computer Science, vol. 14007, pp. 470–501. Springer (2023). https://doi.org/10.1007/978-3-031-30634-1_16
16. Cogliati, B., Patarin, J.: Mirror theory: A simple proof of the pi+pj theorem with xi_max=2. IACR Cryptol. ePrint Arch. p. 734 (2020), https://eprint.iacr.org/2020/734
17. Dai, W., Hoang, V.T., Tessaro, S.: Information-theoretic indistinguishability via the chi-squared method. In: Katz, J., Shacham, H. (eds.) Advances in Cryptology - CRYPTO 2017, Proceedings, Part III. Lecture Notes in Computer Science, vol. 10403, pp. 497–523. Springer (2017). https://doi.org/10.1007/978-3-319-63697-9_17
18. Datta, N., Dutta, A., Nandi, M., Paul, G., Zhang, L.: Single key variant of pmac_plus. IACR Trans. Symmetric Cryptol. **2017**(4), 268–305 (2017). https://doi.org/10.13154/TOSC.V2017.I4.268-305
19. Dinur, I.: Tight indistinguishability bounds for the XOR of independent random permutations by fourier analysis. In: Joye, M., Leander, G. (eds.) Advances in Cryptology - EUROCRYPT 2024, Proceedings, Part I. Lecture Notes in Computer Science, vol. 14651, pp. 33–62. Springer (2024). https://doi.org/10.1007/978-3-031-58716-0_2
20. Dutta, A., Nandi, M., Saha, A.: Proof of mirror theory for $\xi_{\max} = 2$. IEEE Trans. Inf. Theory **68**(9), 6218–6232 (2022). https://doi.org/10.1109/TIT.2022.3171178
21. Eberhard, S.: More on additive triples of bijections. CoRR **abs/1704.02407** (2017), http://arxiv.org/abs/1704.02407
22. Guo, J., Peyrin, T., Poschmann, A., Robshaw, M.J.B.: The LED block cipher. In: Preneel, B., Takagi, T. (eds.) Cryptographic Hardware and Embedded Systems - CHES 2011. Proceedings. Lecture Notes in Computer Science, vol. 6917, pp. 326–341. Springer (2011). https://doi.org/10.1007/978-3-642-23951-9_22
23. Hoang, V.T., Tessaro, S.: Key-alternating ciphers and key-length extension: Exact bounds and multi-user security. In: Robshaw, M., Katz, J. (eds.) Advances in Cryptology - CRYPTO 2016, Proceedings, Part I. Lecture Notes in Computer Science, vol. 9814, pp. 3–32. Springer (2016). https://doi.org/10.1007/978-3-662-53018-4_1
24. Iwata, T., Kurosawa, K.: OMAC: one-key CBC MAC. In: Johansson, T. (ed.) Fast Software Encryption - FSE 2003, Revised Papers. Lecture Notes in Computer Science, vol. 2887, pp. 129–153. Springer (2003). https://doi.org/10.1007/978-3-540-39887-5_11
25. Jha, A.: The generalized sum-capture problem for abelian groups (2024)

26. Jha, A., Nandi, M.: A survey on applications of h-technique: Revisiting security analysis of PRP and PRF. Entropy **24**(4), 462 (2022). https://doi.org/10.3390/E24040462

27. Kim, S., Lee, B., Lee, J.: Tight security bounds for double-block hash-then-sum macs. In: Advances in Cryptology - EUROCRYPT 2020, Proceedings, Part I. pp. 435–465 (2020). https://doi.org/10.1007/978-3-030-45721-1_16

28. Leurent, G., Nandi, M., Sibleyras, F.: Generic attacks against beyond-birthday-bound macs. In: Shacham, H., Boldyreva, A. (eds.) Advances in Cryptology - CRYPTO 2018. Proceedings, Part I. Lecture Notes in Computer Science, vol. 10991, pp. 306–336. Springer (2018). https://doi.org/10.1007/978-3-319-96884-1_11

29. Lucks, S.: The sum of prps is a secure PRF. In: Preneel, B. (ed.) Advances in Cryptology - EUROCRYPT 2000, Proceeding. Lecture Notes in Computer Science, vol. 1807, pp. 470–484. Springer (2000). https://doi.org/10.1007/3-540-45539-6_34

30. Luykx, A., Preneel, B., Tischhauser, E., Yasuda, K.: A MAC mode for lightweight block ciphers. In: Peyrin, T. (ed.) Fast Software Encryption - FSE 2016, Revised Selected Papers. Lecture Notes in Computer Science, vol. 9783, pp. 43–59. Springer (2016). https://doi.org/10.1007/978-3-662-52993-5_3

31. Mennink, B., Reyhanitabar, R., Vizár, D.: Security of full-state keyed sponge and duplex: Applications to authenticated encryption. In: Iwata, T., Cheon, J.H. (eds.) Advances in Cryptology - ASIACRYPT 2015, Proceedings, Part II. Lecture Notes in Computer Science, vol. 9453, pp. 465–489. Springer (2015). https://doi.org/10.1007/978-3-662-48800-3_19

32. Naito, Y.: Blockcipher-based macs: Beyond the birthday bound without message length. In: Advances in Cryptology - ASIACRYPT 2017, Proceedings, Part III. pp. 446–470 (2017). https://doi.org/10.1007/978-3-319-70700-6_16

33. Patarin, J.: Introduction to mirror theory: Analysis of systems of linear equalities and linear non equalities for cryptography. IACR Cryptol. ePrint Arch. p. 287 (2010), http://eprint.iacr.org/2010/287

34. Patarin, J.: Mirror theory and cryptography. Appl. Algebra Eng. Commun. Comput. **28**(4), 321–338 (2017). https://doi.org/10.1007/S00200-017-0326-Y

35. Sibleyras, F., Todo, Y.: Keyed sum of permutations: A simpler rp-based PRF. In: Rosulek, M. (ed.) Topics in Cryptology - CT-RSA 2023, Proceedings. Lecture Notes in Computer Science, vol. 13871, pp. 573–593. Springer (2023)

36. Taylor, R.: An integrity check value algorithm for stream ciphers. In: Stinson, D.R. (ed.) Advances in Cryptology - CRYPTO 1993. Proceedings. Lecture Notes in Computer Science, vol. 773, pp. 40–48. Springer (1993). https://doi.org/10.1007/3-540-48329-2_4

37. Yasuda, K.: The sum of CBC macs is a secure PRF. In: Pieprzyk, J. (ed.) Topics in Cryptology - CT-RSA 2010. Proceedings. Lecture Notes in Computer Science, vol. 5985, pp. 366–381. Springer (2010). https://doi.org/10.1007/978-3-642-11925-5_25

38. Yasuda, K.: A new variant of PMAC: beyond the birthday bound. In: Rogaway, P. (ed.) Advances in Cryptology - CRYPTO 2011. Proceedings. Lecture Notes in Computer Science, vol. 6841, pp. 596–609. Springer (2011). https://doi.org/10.1007/978-3-642-22792-9_34

A Residual Calculations

We aim to show:

$$\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) - \frac{f^2_{\leq(i-1)}}{2^n}\right| \leq \frac{2s^2+8q(s+2q)^2+8q^2(s+2q)}{2^{2n}} \tag{45}$$

$$\sqrt{\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right)} \leq \frac{\sqrt{2}(s+2q)}{2^{n/2}} + \frac{20(s+2q)^{5/2}}{2^{3n/2}} \tag{46}$$

First consider $\left|\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) - \frac{f^2_{\leq(i-1)}}{2^n}\right|$. We need both lower and upper bounds on $\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right)$. Let $\mathcal{I} = \{i_1,\ldots,i_s\}$ be an arbitrary indexing of $\mathcal{R}$ and $\mathcal{J} = \{j_1,\ldots,j_{r_{\leq(i-1)}}\}$ denote the indexing corresponding to $\mathsf{Y}_{\leq(i-1)}$. Then, $\mathcal{I}\sqcup\mathcal{J}$ gives an indexing of $\mathcal{F}$.

For all $j,j' \in \mathcal{I}\sqcup\mathcal{J}$, let $1_{j,j'}$ denote the indicator random variable corresponding to the event $\mathsf{A}_j \oplus \mathsf{B}_{j'} = \mathsf{T}^{(i)}$, where $\mathsf{A}_j, \mathsf{B}_{j'} \in \mathcal{F}$. Then, we have

$$\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) = \sum_{j\neq j'\in\mathcal{I}\sqcup\mathcal{J}} \Pr\left(1_{j,j'}\right). \tag{47}$$

Now, we can have four cases depending upon where j and j' come from:

Case A: $j,j' \in \mathcal{I}$. In this case, for any pair of (j,j'), $\Pr\left(1_{j,j'}\right) = 1/(2^n-1)$ and there are at most $s(s-1)$ such pairs, which results in

$$\sum_{j\neq j'\in\mathcal{I}} \Pr\left(1_{j,j'}\right) = \frac{s(s-1)}{2^n-1}. \tag{48}$$

Case B: $j \in \mathcal{I} \wedge j' \in \mathcal{J}$. In this case, using the fact that there are at least $(2^n - s - 2q)$ and at most 2^n solutions for any equation, we have

$$\frac{2s(i-1)}{2^n} \leq \sum_{j\in\mathcal{I},j'\in\mathcal{J}} \Pr\left(1_{j,j'}\right) \leq \frac{2s(i-1)}{2^n-s-2q} \tag{49}$$

Case C: $j \in \mathcal{I} \wedge j' \in \mathcal{J}$. This case is symmetrical to Case B above.

$$\frac{2s(i-1)}{2^n} \leq \sum_{j'\in\mathcal{I},j\in\mathcal{J}} \Pr\left(1_{j,j'}\right) \leq \frac{2s(i-1)}{2^n-s-2q} \tag{50}$$

Case D: $j,j' \in \mathcal{J}$. Using similar argumentation as above, we have

$$\frac{4(i-1)^2-2(i-1)}{2^n} \leq \sum_{j,j'\in\mathcal{J}} \Pr\left(1_{j,j'}\right) \leq \frac{4(i-1)^2-2(i-1)}{2^n-s-2q} \tag{51}$$

Recall that

$$\frac{f^2_{\leq(i-1)}}{2^n} = \frac{(s+2(i-1))^2}{2^n}.$$

Then, (45) follows from (47)-(51).

Now, consider the second claim. We have to compute the variance of $\mu(\mathsf{T}^{(i)},\mathcal{F})$. First, using the above formulation, we have

$$\begin{aligned}\mathbb{V}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) &= \mathbb{V}\left(\sum_{j,j'\in\mathcal{I}\sqcup\mathcal{J}} 1_{j,j'}\right)\\ &= \sum_{j,j'\in\mathcal{I}\sqcup\mathcal{J}} \mathbb{V}\left(1_{j,j'}\right) + \sum_{\substack{j_1,j_2,j_3,j_4\in\mathcal{I}\sqcup\mathcal{J}\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right)\\ &\leq \sum_{j,j'\in\mathcal{I}\sqcup\mathcal{J}} \mathbb{E}\left(1_{j,j'}\right) + \sum_{\substack{j_1,j_2,j_3,j_4\in\mathcal{I}\sqcup\mathcal{J}\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right)\\ &\leq \mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) + \sum_{\substack{j_1,j_2,j_3,j_4\in\mathcal{I}\sqcup\mathcal{J}\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) \qquad (52)\end{aligned}$$

Now, from (47)-(51), we have

$$\mathbb{E}\left(\mu(\mathsf{T}^{(i)},\mathcal{F})\right) \leq \frac{2(s+2q)^2}{2^n}. \qquad (53)$$

All that remains is to bound the covariances for every choice of $(j_1,j_2)\neq(j_3,j_4)$. First, we have

$$\mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) = \Pr\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) - \Pr\left(1_{j_1,j_2}\right)\Pr\left(1_{j_3,j_4}\right)$$

Given the above discussion on $\Pr\left(1_{j,j'}\right)$ for arbitrary j,j', it is sufficient to upper bound $\Pr\left(1_{j_1,j_2}, 1_{j_3,j_4}\right)$, and use lower bound on $\Pr\left(1_{j_1,j_2}\right)$ (and $\Pr\left(1_{j_3,j_4}\right)$) from the above discussion. Depending upon $j_k\in\mathcal{I}$ or $j_k\in\mathcal{J}$, for all $k\in[4]$, we can have 16 cases, that we group into 5 supercases depending upon the size of $\{j_1,j_2,j_3,j_4\}\cap\mathcal{I}$. We will skip most of the details of computation for each case, and instead discuss the most important subcases.

Case A: $|\{j_1,j_2,j_3,j_4\}\cap\mathcal{I}| = 4$: In this case it is easy to see that $\Pr\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) \leq 1/(2^n-1)(2^n-3)$, and thus

$$\begin{aligned}\sum_{\substack{j_1,j_2,j_3,j_4\in\mathcal{I}\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) &\leq s^4\left(\frac{1}{(2^n-1)(2^n-3)} - \frac{1}{(2^n-1)^2}\right)\\ &\leq \frac{16s^4}{2^{3n}}. \qquad (54)\end{aligned}$$

Case B: $|\{j_1,j_2,j_3,j_4\}\cap\mathcal{I}| = 3$: Wlog assume $j_1\notin\mathcal{I}$. Then, first $\Pr\left(1_{j_3,j_4}\right) = 1/(2^n-1)$ and $\Pr\left(1_{j_1,j_2}\mid 1_{j_3,j_4}\right) \leq 1/(2^n-s-2q)$ (since the j_1 variable is sampled out of a set of size at least (2^n-s-2q)). Thus, in this case, we have

$$\begin{aligned}\sum_{\substack{|\{j_1,j_2,j_3,j_4\}\cap\mathcal{I}|=3\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) &\leq 8s^3q\left(\frac{1}{(2^n-1)(2^n-s-2q)} - \frac{1}{2^n(2^n-1)}\right)\\ &\leq \frac{32(s+2q)^4q}{2^{3n}}. \qquad (55)\end{aligned}$$

Case C: $|\{j_1, j_2, j_3, j_4\} \cap \mathcal{I}| = 2$: The most interesting subcase here is $|\{j_1, j_2\} \cap \mathcal{I}| = 1$, $|\{j_3, j_4\} \cap \mathcal{I}| = 1$. Wlog assume $j_1, j_3 \in \mathcal{I}$ and $j_2, j_4 \in \mathcal{J}$. Let $\mathtt{R}_1, \mathtt{R}_3, \mathtt{Y}_2, \mathtt{Y}_4$ denote the corresponding values in $\mathcal{F}$. We have two equations:

$$\mathtt{R}_1 \oplus \mathtt{Y}_2 = \mathsf{T}^{(i)}$$
$$\mathtt{R}_3 \oplus \mathtt{Y}_4 = \mathsf{T}^{(i)}$$

Now, if $\mathtt{Y}_2$ and $\mathtt{Y}_4$ come from different equations, then the above holds with at most $1/(2^n - s - 2q)^2$ probability as each of $\mathtt{Y}_2$ and $\mathtt{Y}_4$ are sampled from a set of size at least $(2^n - s - 2q)$. The interesting case arises when they are from the same equation, say (k). In this case the above equation holds if and only if $\mathtt{R}_1 \oplus \mathtt{R}_3 = \mathsf{T}^{(i)} \oplus \mathsf{T}^{(k)}$. Thus, we have a modified system

$$\mathtt{R}_1 \oplus \mathtt{R}_3 = \mathsf{T}^{(i)} \oplus \mathsf{T}^{(k)}$$
$$\mathtt{R}_1 \oplus \mathtt{Y}_2 = \mathsf{T}^{(i)}$$

Now, once we fix j_1, j_3 and (k) all other indices are fixed (remember, (i) is fixed throughout). Thus, we have at most $2s^2q$ choices and each choice holds with at most $1/(2^n - 1)(2^n - s - 2q)$ probability, which is less than the probability in other cases. All in all, by taking the maximum probability, in this case we have

$$\sum_{\substack{|\{j_1,j_2,j_3,j_4\} \cap \mathcal{I}|=2 \\ \{j_1,j_2\} \neq \{j_3,j_4\}}} \mathbb{V}(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}) \leq 24s^2q^2 \left(\frac{1}{(2^n - s - 2q)^2} - \frac{1}{2^{2n}} \right)$$
$$\leq \frac{96(s + 2q)^3 q^2}{2^{3n}}. \tag{56}$$

Case D: $|\{j_1, j_2, j_3, j_4\} \cap \mathcal{I}| = 1$: Wlog assume $j_1 \in \mathcal{I}$. The most interesting case here would be if j_3 and j_4 correspond to the same equation index say (k), in which case $\mathbb{1}_{j_3,j_4}$ happens if and only if $\mathsf{T}^{(i)} = \mathsf{T}^{(k)}$. But since $\mathsf{T}^{(i)}$ is uniform and independent of $\mathsf{T}^{(k)}$, the overall probability in this subcase is still $1/2^n(2^n - s - 2q) \leq 1/(2^n - s - 2q)(2^n - s - 2q)$. Again by taking the maximum probability across all subcases, we have

$$\sum_{\substack{|\{j_1,j_2,j_3,j_4\} \cap \mathcal{I}|=1 \\ \{j_1,j_2\} \neq \{j_3,j_4\}}} \mathbb{V}(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}) \leq 48sq^3 \left(\frac{1}{(2^n - s - 2q)^2} - \frac{1}{2^{2n}} \right)$$
$$\leq \frac{192(s + 2q)^2 q^3}{2^{3n}}. \tag{57}$$

Case E: $|\{j_1, j_2, j_3, j_4\} \cap \mathcal{I}| = 0$: Using a similar argumentation as above, we have

$$\sum_{\substack{|\{j_1,j_2,j_3,j_4\} \cap \mathcal{I}|=0 \\ \{j_1,j_2\} \neq \{j_3,j_4\}}} \mathbb{V}(\mathbb{1}_{j_1,j_2}, \mathbb{1}_{j_3,j_4}) \leq 16q^4 \left(\frac{1}{(2^n - s - 2q)^2} - \frac{1}{2^{2n}} \right)$$
$$\leq \frac{64(s + 2q)q^4}{2^{3n}}. \tag{58}$$

A cursory look shows that the covariance across all the cases is in $O((s+2q)^5/2^{3n})$. In particular, after appropriate simplifications, we have

$$\sum_{\substack{j_1,j_2,j_3,j_4\in\mathcal{I}\sqcup\mathcal{J}\\ \{j_1,j_2\}\neq\{j_3,j_4\}}} \mathbb{V}\left(1_{j_1,j_2}, 1_{j_3,j_4}\right) \leq \frac{400(s+2q)^5}{2^{3n}} \tag{59}$$

Then, (46) follows by taking square root on both sides of (52) after appropriate substitutions from (53) and (59).