Graph-Based Construction for Non-Malleable Codes

Shohei Satake [*] Yujie Gu [†] Kouichi Sakurai [‡]

Abstract

Non-malleable codes are introduced to protect the communication against adversarial tampering of data, as a relaxation of the error-correcting codes and error-detecting codes. To explicitly construct non-malleable codes is a central and challenging problem which has drawn considerable attention and been extensively studied in the past few years. Recently, Rasmussen and Sahai built an interesting connection between non-malleable codes and (non-bipartite) expander graphs, which is the first explicit construction of non-malleable codes based on graph theory other than the typically exploited extractors. So far, there is no other graph-based construction for non-malleable codes yet. In this paper, we aim to explore more connections between non-malleable codes and graph theory. Specifically, we first extend the Rasmussen-Sahai construction to bipartite expander graphs. Accordingly, we establish several explicit constructions for non-malleable codes based on Lubotzky-Phillips-Sarnak Ramanujan graphs and generalized quadrangles, respectively. It is shown that the resulting codes can either work for a more flexible split-state model or have better code rate in comparison with the existing results.

1 Introduction

Non-malleable codes, introduced by Dziembowski, Pietrzak and Wichs [21, 22], are resilient to adversarial tampering on *arbitrary* number of symbols which is beyond the scope of error-correcting and error-detecting codes. Consider the following "tampering experiment". A message $m \in \mathcal{M}$ is encoded via a (randomized) encoding function enc : $\mathcal{M} \to \mathcal{X}$, yielding a codeword $c = \mathrm{enc}(m)$. However the codeword c is modified by an adversary using some tampering function $f \in \mathcal{F}$ with $f : \mathcal{X} \to \mathcal{X}$ to an erroneous word $\tilde{c} = f(c)$, and $\tilde{c}$ is decoded using a deterministic function dec, resulting $\tilde{m} = \mathrm{dec}(\tilde{c})$. In terms of the practical application, the reliability $\tilde{m} = m$ is desired. An error-correcting code with minimum distance d can

[*]School of Interdisciplinary Mathematical Sciences, Meiji University 4-21-1 Nakano, Nakano-ku, Tokyo 164-8525, Japan e-mail: shohei_satake@meiji.ac.jp

[†]Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan, 819-0395 e-mail: gu@inf.kyushu-u.ac.jp

[‡]Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka, Japan, 819-0395 e-mail: sakurai@inf.kyushu-u.ac.jp

guarantee the reliable communication with respect to the family $\mathcal{F}$ which satisfies that for $f \in \mathcal{F}$ the Hamming distance between $\tilde{c} = f(c)$ and c is at most $\lfloor (d-1)/2 \rfloor$. However it is impossible to achieve the reliability using error-correcting codes if the tampering family $\mathcal{F}$ is large. In order to deal with this, Dziembowski *et al.* [21] proposed the non-malleable codes (with respect to $\mathcal{F}$), which ensure that either the tampered codeword is correctly decoded, i.e., $\tilde{m} = m$, or the decoded message $\tilde{m}$ is completely unrelated to the original message m. As remarked in [21] and [22], the concept of non-malleable codes is in a spirit of non-malleablity proposed by Dolev, Dwork and Naor [17] in cryptographic primitives. Informally, the non-malleability in the context of encryption requires that given the ciphertext it is impossible to generate a different ciphertext so that the respective plaintexts are related [17].

It is known that no non-malleable code exists if the tampering family $\mathcal{F}$ is the entire space of functions. Thus the study on non-malleable codes has focused on the specific families $\mathcal{F}$. One typical tampering family is with the *split-state model*, which has also been investigated in the context of leakage cryptography [13, 20]. Roughly speaking, this model assumes that the encoded memory/state of the system is partitioned into two parts and adversaries can arbitrarily tamper the data stored in each part independently. More precisely, each message is encoded into a word $c = (L, R) \in \mathcal{L} \times \mathcal{R}$ and adversaries try to tamper it using some functions $g : \mathcal{L} \to \mathcal{L}$ and $h : \mathcal{R} \to \mathcal{R}$ which change c to $\tilde{c} = (g(L), h(R)) \in \mathcal{L} \times \mathcal{R}$. Moreover, if $|\mathcal{L}| = |\mathcal{R}|$, we call it *equally-sized* split-state model.

To explicitly construct non-malleable codes is a fundamental and challenging problem. In the literature, explicit non-malleable codes for the split-state model have been derived based on two-source extractors and additive combinatorics, see [1], [2], [3], [4], [5], [6], [7], [10], [11], [18], [29], [30] for example. Notably, Dziembowski, Kazana and Obremski [18] pointed out: "This brings a natural question if we could show some relationship between the extractors and the non-malleable codes in the split-state model. Unfortunately, there is no obvious way of formalizing the conjecture that non-malleable codes need to be based on extractors". Recently, Rasmussen and Sahai [36] discovered that (non-bipartite) expander graphs could provide non-malleable codes for the split-state model, which in some sense answers Dziembowski-Kazana-Obremski's question in [18]. Inspired by [36], we are interested with exploring more graph-theoretic constructions for split-state non-malleable codes. More precisely, we shall study the following problem.

Problem 1. *Based on graph theory, provide explicit constructions of non-malleable codes for the split-state model.*

Indeed, Rasmussen and Sahai [36] provided an elegant answer to Problem 1. However we noticed that the construction in [36] cannot be directly transferred to the general split-state model. Inspired by this, we initially extend the construction in [36] to *bipartite graphs*. Specifically, in this paper, we first establish a coding scheme based on bipartite graphs. Then

Ref.	$\vert\mathcal{L}\vert$	$\vert\mathcal{R}\vert$	code rate	comments
[36, Sec. C]	q^3	q^3	$\frac{1}{24\log_2(1/\varepsilon)+O(1)}$	$q=p^2$, p is a prime
Cor. 17	$\Theta(p^{5/2}\log(p))$	$\Theta(p^{5/2}\log(p))$	$\frac{1}{20\log_2(1/\varepsilon)+O(\log\log(1/\varepsilon))}$	p is an odd prime

Table 1: Explicit graph-based ε-non-malleable codes in this paper and [36]

Ref.	#vertices n	$\vert\mathcal{L}\vert$	$\vert\mathcal{R}\vert$	code rate	minimum rate for n ($\vert\mathcal{L}\vert=\vert\mathcal{R}\vert=n/2$)
Thm. 18	$2q^2(q+1)$	$(q+2)q^2$	q^3	$\frac{1}{\log_2(q^6+2q^5)}$	$\frac{1}{\log_2(q^6+2q^5+q^4)}$
Thm. 20	$(q^3+q^2+2)(q^5+1)$	$(q^2+1)(q^5+1)$	$(q^3+1)(q^5+1)$	$\frac{1}{\log_2(q^{15}+O(q^{14}))}$	$\frac{1}{\log_2(q^{16}+O(q^{15}))}$

Table 2: Code rates of non-malleable codes based on bipartite graphs with n vertices, where q is a prime power

we prove that when the underlying bipartite graph is an (r,s)-*biregular graph* with the second largest eigenvalue μ, our coding scheme provides $O\big(\frac{\mu^{3/2}}{\sqrt{rs}}\big)$-non-malleable codes for the split-state model which is not necessarily to be equally-sized (see Theorem 10). This can be seen as an extension of the coding scheme in [36] in the sense that we could deduce the codes for equally-sized split-state model in [36] as special cases (see Remark 9). Based on this, we provide several more solutions to Problem 1 by means of Lubotzky-Phillips-Sarnak Ramanujan graphs and generalized quadrangles (see Tables 1 and 2). In particular, the resulting non-malleable codes can either work for more flexible *non-equally-sized* split-state model (see Theorems 18, 20) or have *better* code rate (see Theorem 16, Corollary 17) in comparison with the non-malleable codes in [36, Section C]. In particular, for a given size of graphs, codes for non-equally-sized split-state model in general realize larger code rate than the rate of codes for equally-sized split-state model (see Table 2 and Section 4).

The remainder of this paper is organized as follows. Section 2 briefly reviews non-malleable codes and basics in graph theory. Section 3 provides the coding scheme based on bipartite graphs and discusses its non-malleability. Section 4 analyzes the code rate of the established non-malleable codes. Section 5 presents several explicit constructions for non-malleable codes. Section 6 concludes this paper. Appendix proves Theorem 10.

2 Preliminaries

In this section we recall the notion of non-malleable codes and some useful basics in graph theory. Throughout this paper, let $x \leftarrow \mathcal{X}$ denote that the random variable x sampled uniformly from a set $\mathcal{X}$. Let $\perp$ denote a special symbol.

For positive-valued functions f and g over $\mathbb{N}$, we say $f = O(g)$ as $n \to \infty$ if there exists a constant $C > 0$ that $f(n) \leq Cg(n)$ holds for any sufficiently large n. Similarly $f = \Omega(g)$ as $n \to \infty$ if there exists a constant $C > 0$ that $f(n) \geq Cg(n)$ holds for any sufficiently large n. In particular $f = \Theta(g)$ as $n \to \infty$ if $\Omega(g) = f = O(g)$ holds. Also $f = o(g)$ as $n \to \infty$ if $\lim_{n\to\infty} f(n)/g(n) = 0$.

2.1 Non-malleable codes

Let $\mathcal{M}$ be a set of messages and $\mathcal{X}$ a set of codewords. A *coding scheme* is a pair of functions $(\mathrm{enc}, \mathrm{dec})$, where $\mathrm{enc} : \mathcal{M} \to \mathcal{X}$ is a randomized encoding function, and $dec : \mathcal{X} \to \mathcal{M} \cup \{\perp\}$ is a deterministic decoding function. Assume that for all $m \in \mathcal{M}$,

$$\Pr[\mathrm{dec}(\mathrm{enc}(m)) = m] = 1,$$

where the probability is taken over the randomness of enc.

Let A, B be two random variables over the same set $\mathcal{X}$. Then the *statistical distance* between A and B is defined as

$$\Delta(A, B) := \frac{1}{2} \sum_{x \in \mathcal{X}} \Big| \Pr[A = x] - \Pr[B = x] \Big|.$$

Definition 2 (Split-state non-malleable codes)**.** In the split-state model, assume $\mathcal{X} = \mathcal{L} \times \mathcal{R}$ is the product set of sets $\mathcal{L}$ and $\mathcal{R}$. Let $\mathcal{F}$ be a set of functions from $\mathcal{L} \times \mathcal{R}$ to itself, where each $f \in \mathcal{F}$ can be represented as $f(L, R) = (g(L), h(R))$ for all $(L, R) \in \mathcal{L} \times \mathcal{R}$ with some $g : \mathcal{L} \to \mathcal{L}$ and $h : \mathcal{R} \to \mathcal{R}$. Then a coding scheme $(\mathrm{enc}, \mathrm{dec})$ such that $\mathrm{enc} : \mathcal{M} \to \mathcal{L} \times \mathcal{R}$ and $\mathrm{dec} : \mathcal{L} \times \mathcal{R} \to \mathcal{M} \cup \{\perp\}$ is called an *ε-non-malleable code with respect to $\mathcal{F}$* if for every $f \in \mathcal{F}$, there exists a probability distribution D_f on $\mathcal{M} \cup \{\mathrm{same}^*, \perp\}$ such that for every $m \in \mathcal{M}$, we have $\Delta(A_f^m, B_f^m) \leq \varepsilon$, where for $m \in \mathcal{M}$ and $f \in \mathcal{F}$, let A_f^m and B_f^m be events defined as follows.

$$A_f^m := \left\{ \begin{array}{c} (L, R) \leftarrow \mathrm{enc}(m); \\ \text{Output } \mathrm{dec}(g(L), h(R)) \end{array} \right\},$$

$$B_f^m := \left\{ \begin{array}{c} \tilde{m} \leftarrow D_f; \\ \text{If } \tilde{m} = \mathrm{same}^* \text{ output } m \text{ else output } \tilde{m} \end{array} \right\}.$$

Hereafter, as in [18] and [36], the symbol "$\perp$" from Definition 2 will be dropped since it usually denotes the situation when the decoding function detects tampering and outputs an error message, which is not dealt in this paper. As mentioned in [18], this would be not so problematic for practical applications.

Definition 3 (Code rate)**.** For a coding scheme $\mathcal{C}$ with the set of messages $\mathcal{M}$ and the set of codewords $\mathcal{X}$, the code rate $R(\mathcal{C})$ is defined as

$$R(\mathcal{C}) := \frac{\log_2 |\mathcal{M}|}{\log_2 |\mathcal{X}|}.$$

In particular if $\mathcal{M} = \{0,1\}^\kappa$ and $\mathcal{X} = \{0,1\}^n$ then $R(\mathcal{C})$ is the ratio of the message length κ and codeword length n.

This paper focuses on *single-bit* non-malleable codes, i.e., $\mathcal{M} = \{0,1\}$. It is shown in [18] that single-bit non-malleable codes can be formulated as in the following Theorem 4 as well.

Theorem 4 ([18], [19])**.** *Let* (enc, dec) *be a coding scheme with* $\mathrm{enc} : \{0,1\} \to \mathcal{X}$ *and* $\mathrm{dec} : \mathcal{X} \to \{0,1\}$*. Let* $\mathcal{F}$ *be a set of functions from* $\mathcal{X}$ *to itself. Then* (enc, dec) *is an* ε*-non-malleable code with respect to* $\mathcal{F}$ *if and only if it holds for every* $f \in \mathcal{F}$ *that*

$$\frac{1}{2} \sum_{b \in \{0,1\}} \Pr\Big[\mathrm{dec}\big(f(\mathrm{enc}(b))\big) = 1 - b\Big] \leq \frac{1}{2} + \varepsilon$$

where the probability is over the uniform choice of b *and the randomness of* enc.

2.2 Expander graphs

Throughout this paper, we assume that all graphs are undirected and simple, i.e., without multiple edges and loops. Let $G = (V, E)$ denote a graph G with vertex set V and edge set E. Let $G = (V_1, V_2, E)$ be a bipartite graph with a partition (V_1, V_2) of vertex set and edge set $E \subset \{\{v_1, v_2\} : v_1 \in V_1, v_2 \in V_2\}$. For convenience, we identify $G = (V_1, V_2, E)$ with an orientation $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$ where

$$\overrightarrow{E} = \{(v_1, v_2) : \{v_1, v_2\} \in E\} \subset V_1 \times V_2.$$

We call $\overrightarrow{G}$ the *associated orientation of* G.

We say a vertex has *degree* d if it connects exactly d edges. A graph G is called a d-*regular graph* if every vertex has degree d. A bipartite graph $G = (V_1, V_2, E)$ is called an (r, s)-*biregular graph* if every vertex of V_1 and V_2 has degree r and s, respectively. Clearly, for an (r, s)-biregular graph $G = (V_1, V_2, E)$ and its associate orientation $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$, the following equation holds.

$$|E| = |\overrightarrow{E}| = r|V_1| = s|V_2|. \tag{2.1}$$

Let $G = (V, E)$ be a graph with n vertices. Then the *adjacency matrix* of G, denoted by $A(G)$, is a $|V| \times |V|$ binary matrix such that the (u, w)-entry is 1 if and only if $\{u, w\} \in E$. Clearly, $A(G)$ is a symmetric matrix and thus has exactly n real eigenvalues with multiplicity, denoted by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

Lemma 5 (e.g. [9]). *Let G be a graph with n vertices.*

1. *If G is d-regular, then $\lambda_1 = d$ and $\lambda_n \geq -d$, where $\lambda_n = -d$ if and only if G is bipartite.*
2. *If G is (r, s)-biregular, then $\lambda_1 = \sqrt{rs}$ and $\lambda_n = -\sqrt{rs}$.*

By Lemma 5, the largest eigenvalue of a (bi-)regular graph is always determined. However, the second largest eigenvalue typically has rich properties. For a d-regular graph G, denote $\lambda(G) := \max_{2 \leq i \leq n} |\lambda_i|$. For an (r, s)-biregular graph G, denote

$$\mu(G) := \max_{2 \leq i \leq n-1} |\lambda_i|.$$

An (r, s)-biregular graph G is a *μ-spectral expander* if $\mu(G) \leq \mu$. It has the following nice expansion property.

Proposition 6 ([40]). *Let $G = (V_1, V_2, E)$ be an (r, s)-biregular graph which is a μ-spectral expander. For a subset $S \subset V_1$, define the neighbour of S as $N(S) := \{u \in V_2 : u$ is adjacent to some vertex in $S\}$, and let $\rho(S) := \frac{|S|}{|V_1|}$. Then for every subset $S \subset V_1$,*

$$\frac{|N(S)|}{|S|} \geq \frac{r^2}{\rho(S)(rs - \mu^2) + \mu^2}.$$

By Proposition 6, it is readily seen that if G is a μ-spectral expander with small μ, then G has a good expansion property and thus we are interested in how $\mu(G)$ can be small.

Lemma 7 ([27]). *Suppose that G is a sufficiently large graph. Then the followings hold.*

(1) *If G is d-regular, then $\lambda(G) = \Omega(\sqrt{d})$.*

(2) *If G is (r, s)-biregular, then $\mu(G) = \Omega(\sqrt{r + s})$.*

3 Codes from bipartite graphs

In this section we provide a bipartite graph based coding scheme and show that it produces non-malleable codes.

3.1 A coding scheme

First we propose a coding scheme based on bipartite graphs.

Construction 8. Let $G = (V_1, V_2, E)$ be a bipartite graph and $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$ the associated orientation of G. Then the associated graph code $\mathcal{C}_G := (\mathrm{enc}_G, \mathrm{dec}_G)$ consists of the functions

$$\mathrm{enc}_G : \{0,1\} \to V_1 \times V_2, \qquad \mathrm{dec}_G : V_1 \times V_2 \to \{0,1\}$$

such that

$$\mathrm{enc}_G(b) := \begin{cases} (u,w) \leftarrow (V_1 \times V_2) \setminus \overrightarrow{E} & \text{if } b = 0; \\ (u,w) \leftarrow \overrightarrow{E} & \text{if } b = 1, \end{cases}$$

$$\mathrm{dec}_G(v_1, v_2) := \begin{cases} 0 & \text{if } (v_1, v_2) \notin \overrightarrow{E}; \\ 1 & \text{if } (v_1, v_2) \in \overrightarrow{E}. \end{cases}$$

Remark 9. Rasmussen and Sahai [36] designed a coding scheme based on a graph $G = (V, E)$ so that the space of codewords is $V \times V$, but it works only for equally-sized split-state model with $|\mathcal{L}| = |\mathcal{R}| = |V|$. On the other hand, our code can be applied to a more flexible split-state model, i.e. $|\mathcal{L}| = |V_1|$ may not be necessary equal to $|\mathcal{R}| = |V_2|$. An advantage of such model is discussed in Section 4 afterwards.

3.2 Non-malleability

We show that the coding scheme in Construction 8 based on biregular spectral expanders can produce non-malleable codes for the split-state model.

Theorem 10. *Let $G = (V_1, V_2, E)$ be an (r,s)-biregular graph with n vertices which is a μ-spectral expander. Suppose that $r = r(n), s = s(n)$ with $r, s \to \infty$ as $n \to \infty$ (hence $\mu = \mu(n) \to \infty$). Assume that $|E| = \Omega\left(\frac{(rs)^2 \log(rs)}{\mu}\right)$ $(n \to \infty)$. Let $\mathcal{F}$ be the set of all functions $f = (g,h)$ with $g : V_1 \to V_1$ and $h : V_2 \to V_2$, where $f(v_1, v_2) := (g(v_1), h(v_2))$ for any $(v_1, v_2) \in V_1 \times V_2$. Then the code $\mathcal{C}_G$ is an $O\left(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\right)$-non-malleable code with respect to $\mathcal{F}$ as $n \to \infty$.*

The proof of Theorem 10 is referred to Appendix.

Remark 11. Suppose that G is an (r,s)-biregular graph with $s \geq r = \omega(\sqrt{s})$ and $\mu(G) = O(\sqrt{s})$. Then Theorem 10 guarantees that the code $\mathcal{C}_G$ is an $O(s^{1/4}/r^{1/2})$-non-malleable code, where $s^{1/4}/r^{1/2} = o(1)$ by the assumption on r and s. On the other hand, according to Lemma 7, the quantity $O(s^{1/4}/r^{1/2})$ in Theorem 10 is best possible up to a constant.

The following corollary follows from Theorem 10 and (2.1).

Corollary 12. *Let $G = (V_1, V_2, E)$ be a bipartite d-regular graph with $|V_1| = |V_2| = n/2$ which is a μ-spectral expander. Suppose that $n = \Omega\left(\frac{\log(d)\cdot d^3}{\mu}\right)$ and $\mathcal{F}$ is as in Theorem 10. Then the code $\mathcal{C}_G$ is an $O\left(\frac{\mu^{3/2}}{d}\right)$-non-malleable code with respect to $\mathcal{F}$.*

Remark 13. Corollary 12 actually includes the explicit construction of non-malleable codes by Rasmussen and Sahai [36, Section C]. Indeed, for a finite abelian group X and a subset S of X, the Cayley graph $Cay(X, S)$ is an $|S|$-regular graph with vertex set X in which two vertices x and y are adjacent if and only if $xy^{-1} \in S$. Note that from $Cay(X, S)$, a bipartite $|S|$-regular graph can be easily obtained as follows. Take two disjoint copies X_1 and X_2 of X and construct a bipartite graph so that $x_1 \in X_1$ and $x_2 \in X_2$ are adjacent if and only if $x_1 x_2^{-1} \in S$; such a bipartite regular graph is called a bi-Cayley graph [35]. For a prime p let $\mathbb{F}_p$ denote the p-element field and $q = p^2$. Rasmussen and Sahai [36] constructed an $O(q^{-1/4})$-non-malleable code from a non-bipartite graph $Cay(\mathbb{F}_p^6, S)$ with some $S \subset \mathbb{F}_p^6$ such that $|S| = q$. In terms of Corollary 12, the corresponding bi-Cayley graph provides the same non-malleable code as in [36, Section C].

4 Non-equally-sized split-state model

In this section we discuss the code rate of non-malleable codes for *non-equally-sized* split-state model. Recall from Theorem 10 that the non-malleable code $\mathcal{C}_G$ is established for a given integer n and a bipartite graph $G = (V_1, V_2, E)$ with n vertices. The robustness ε of the non-malleable code $\mathcal{C}_G$ relies on the parameters $|V_1|$, $|V_2|$, r, s and μ, which are functions of n. In terms of the code rate, we also have the following interesting observation.

Lemma 14. *Let n be a positive integer. Let $G = (V_1, V_2, E)$ be a bipartite graph with n vertices. Then*

$$R(\mathcal{C}_G) = \frac{1}{\log_2(|V_1||V_2|)}. \tag{4.1}$$

Moreover assuming $|V_1| \geq |V_2|$ we have

$$R(\mathcal{C}_G) \geq \frac{1}{\log_2(\lceil n/2 \rceil \cdot \lfloor n/2 \rfloor)}, \tag{4.2}$$

where the equality holds if and only if $|V_1| = \lceil n/2 \rceil$ and $|V_2| = \lfloor n/2 \rfloor$.

Proof. The equation (4.1) immediately follows from the construction of $\mathcal{C}_G$ (Construction 8). Also (4.2) can be proved by finding the maximum value of $\log_2(|V_1||V_2|)$ under the conditions that $|V_1| + |V_2| = n$ and $1 \leq |V_1|, |V_2| \leq n - 1$. Notice that $|V_2| = n - |V_1|$, then $|V_1||V_2| = |V_1|(n - |V_1|)$ is maximized if and only if $|V_1| = \lceil n/2 \rceil$ and $|V_2| = \lfloor n/2 \rfloor$. Since the function $\log_2(\cdot)$ is monotone-increasing, the inequality (4.2) follows. □ □

According to Lemma 14, for any bipartite graph $G = (V_1, V_2, E)$ with n vertices, it is readily seen that the smallest code rate occurs when $|V_1| = |V_2|$ (i.e. equally-sized). In other words, the non-malleable codes derived from Construction 8 can have *better* code rate for *non-*equally-sized split-state model in comparison with equally-sized split-state model in general. Furthermore it is worth noting that the larger the ratio of $|V_1|$ and $|V_2|$ is, the better the code rate is. In addition to the above analysis on code rate, one could also derive code rate according to the robustness parameter ε of the established code in Construction 8 (see Section 5.1, for example).

5 Explicit constructions

In this section, we present explicit non-malleable codes based on specific biregular spectral expanders.

5.1 Via Lubotzky-Phillips-Sarnak Ramanujan graphs

In this subsection, we construct non-malleable codes based on suitably chosen graphs from known families of bipartite regular graphs. The resulting codes can have better code rate in comparison with the codes in [36] (see also Remark 13).

To show our construction based on Corollary 12, we need the following claim.

Claim 15. For a given large prime p, there exist explicit (bipartite) $(p+1)$-regular graphs G with $\Theta(p^{5/2}\log(p))$ vertices and $\mu(G) \le 2\sqrt{p}$.

Our construction of graphs is based on the following *Ramanujan graphs* due to Lubotzky, Phillips and Sarnak [32], and Margulis [33] (see also [14, Theorem 4.2.2]). Let p, r be two distinct odd primes such that $r > 2\sqrt{p}$ and p is a *quadratic non-residue* modulo r. Then one can explicitly construct a bipartite $(p+1)$-regular graph $X^{p,r}$ with $r(r^2-1)$ vertices and $\mu(X^{p,r}) \le 2\sqrt{p}$ for every $r > 2\sqrt{p}$. Indeed the graph $X^{p,r}$ is constructed as a Cayley graph $Cay(\mathrm{PGL}_2(\mathbb{F}_r), S)$ with some explicit generating set $S \subset \mathrm{PGL}_2(\mathbb{F}_r)$ of size $p+1$, where $\mathrm{PGL}_2(\mathbb{F}_r)$ denotes the projective general linear group of rank 2 over the r-element field $\mathbb{F}_r$. The details of the construction can be found in [14]. Note that (e.g. [34]) for each prime p, one could check whether two given vertices are adjacent in $X^{p,r}$ in $O(\log(r))$-time, and hence the graph can be constructed in poly(r)-time.

Proof of Claim 15. To prove Claim 15, it suffices to take the graph $X^{p,r}$ with $r = \Theta(p^{5/6}\log^{1/3}(p))$. Indeed for each sufficiently large prime p, by Bertrand's postulate, there exists a prime $r = \Theta(p^{5/6}\log^{1/3}(p)) > 2\sqrt{p}$, which can be found in poly(p)-time. If p is a quadratic non-residue modulo r, then $X^{p,r}$ is a bipartite $(p+1)$-regular graph with $\Theta(p^{5/2}\log(p))$ vertices and $\mu(X^{p,r}) \le 2\sqrt{p}$. □

Thus we obtain the following theorem.

Theorem 16. *For any sufficiently large prime p, suppose that $r = \Theta(p^{5/2}\log(p))$ is a prime such that p is a quadratic non-residue modulo r. Then the code $\mathcal{C}_G$ with $G = X^{p,r}$ is an $O(p^{-1/4})$-non-malleable code for the split-state model with $|\mathcal{L}| = |\mathcal{R}| = \Theta(p^{5/2}\log(p))$.*

Note that Theorem 16 cannot deal with the case when $r = \Theta(p^{5/2}\log(p))$ is a prime such that p is a *quadratic residue* modulo r. However, in this case, one can instead explicitly construct a *non-bipartite* $(p+1)$-regular graph $Y^{p,r}$ with $r(r^2-1)/2$ vertices and $\lambda(Y^{p,r}) \leq 2\sqrt{p}$ (see [14], [32], [33]). By [36, Theorem 7], the graph $Y^{p,r}$ with $r = \Theta(p^{5/2}\log(p))$ provides an $O(p^{-1/4})$-non-malleable code for the split-state model with $|\mathcal{L}| = |\mathcal{R}| = \Theta(p^{5/2}\log(p))$. Moreover, for each pair of primes p and r, one could check whether two given vertices are adjacent or not in $X^{p,r}$ and $Y^{p,r}$ in $O(\log(p))$-time. By these facts and Theorem 16, we immediately obtain the following corollary.

Corollary 17. *For any sufficiently large prime p, there exists an explicit $(p+1)$-regular graph G with $\Theta(p^{5/2}\log(p))$ vertices which provides an $O(p^{-1/4})$-non-malleable code for the split-state model with $|\mathcal{L}| = |\mathcal{R}| = \Theta(p^{5/2}\log(p))$. In particular, for every $0 < \varepsilon < 1$, there exists an explicit ε-non-malleable code with code rate*

$$\frac{1}{20\log_2(1/\varepsilon) + O(\log\log(1/\varepsilon))}.$$

Moreover, both of encoding and decoding can be done in $O(\log(1/\varepsilon))$-time.

The last statement of Corollary 17 directly follows from the discussion in [36, Section 1.3]. Note that the explicit codes derived in [36] (see also Remark 13) have code rate $1/(24\log(1/\varepsilon)+O(1))$ while encoding and decoding time is $O(\log(1/\varepsilon))$. In other words, the resulting codes here can have better code rate in comparison with the codes in [36], while encoding and decoding time are the same (up to constant).

5.2 Via generalized quadrangles

In this subsection, we provide split-state non-malleable codes based on generalized quadrangles. The code rates of these codes can also be found in Table 2 in Introduction.

A *generalized quadrangle* of order (α, β) is an $(\alpha+1, \beta+1)$-biregular graph $GQ(\alpha, \beta) = (V_1, V_2, E)$ such that

- for all $x, y \in V_1 \cup V_2$, there exists a path of length ≤ 4 connecting x and y;
- for all $x, y \in V_1 \cup V_2$, if the length of the shortest path connecting x and y is $h < 4$, then there exists only one path of length h connecting x and y;

- for every $x \in V_1 \cup V_2$, there exists $y \in V_1 \cup V_2$ such that there exists a path of length 4 connecting x and y.

More details of generalized quadrangles can be found in [37], [41]. Now based on generalized quadrangles, we can derive two families of non-malleable codes for the non-equally-sized scenario.

Theorem 18. *For any prime power q, the code $\mathcal{C}_G$ with $G = GQ(q-1, q+1)$ is an $O(q^{-1/4})$-non-malleable code for the split-state model with $|\mathcal{L}| = (q+2)q^2$ and $|\mathcal{R}| = q^3$.*

We need the following lemma to prove Theorem 18.

Lemma 19 ([37], [40], [41])**.** *For the graph $GQ(\alpha, \beta)$, we have*

- $|V_1| = (\alpha+1)(\alpha\beta+1)$,
- $|V_2| = (\beta+1)(\alpha\beta+1)$,
- $\mu(GQ(\alpha,\beta)) = \sqrt{\alpha+\beta}$.

Proof of Theorem 18. To obtain the theorem, we apply an explicit construction of $GQ(q-1, q+1)$ for every prime power q ([8, Sections 4 and 5]). According to (2.1) and Lemma 19, we have $|E| = r|V_1| = \Theta(q^4)$ and $\frac{(rs)^2\log(rs)}{\mu} = \Theta(q^{7/2}\log q)$. Thus by Theorem 10, the code $\mathcal{C}_G$ with $G = GQ(q-1, q+1)$ gives the desired code. □

The following theorem can deal with more unbalanced non-equally-sized scenario.

Theorem 20. *For any prime power q, the code $\mathcal{C}_G$ with $G = GQ(q^2, q^3)$ is an $O(q^{-1/4})$-non-malleable code for the split-state model with $|\mathcal{L}| = (q^2+1)(q^5+1)$ and $|\mathcal{R}| = (q^3+1)(q^5+1)$.*

Proof. For every prime power q, there is an explicit construction of $GQ(q^2, q^3)$ (e.g. [37, Chapter 3]). By (2.1) and Lemma 19, we have $|E| = r|V_1| = \Theta(q^{10})$ and $\frac{(rs)^2\log(rs)}{\mu} = \Theta(q^{17/2}\log q)$. Thus according to Theorem 10, $(\mathrm{enc}_G, \mathrm{dec}_G)$ with $G = GQ(q^2, q^3)$ gives the desired code. □ □

6 Concluding remarks

In this paper, we proposed a coding scheme based on bipartite graphs and showed that the non-malleability can be satisfied if the underlying bipartite graph is a biregular μ-spectral expander with sufficiently small μ. Based on it, we provided explicit non-malleable codes via several types of biregular spectral expanders such as Ramanujan graphs and generalized quadrangles. The established non-malleable codes can either work for a more flexible split-state model or have better code rate in comparison with the existing results.

In addition, our results show that some related error-correcting codes have potential applications to constructing non-malleable codes for the split-state model. For example, it is well-known in coding theory and combinatorics that a low-density parity-check (LDPC) code has an associated bipartite graph called *Tanner graph*. Precisely, the Tanner graph of an LDPC code with parity-check matrix $H = (h_{ij})$ is a bipartite graph such that the vertex set is the index set of rows and columns of H, and two vertices i and j are adjacent if and only if $h_{ij} \neq 0$, see [39]. It is shown that the algebraic or combinatorial constructions of LDPC codes often provide Tanner graphs with small second largest eigenvalue, see [16], [25], [31], [38], [28] for example. According to the bipartite graph based coding scheme proposed in this paper, a connection between LDPC codes and non-malleable codes can be accordingly established. Particularly, the constructions in Theorems 18 and 20 are based on several typical bipartite graphs realized as Tanner graphs of LDPC codes.

In terms of practical applications, it is desirable to construct split-state non-malleable codes for k-bit messages with $k \geq 2$. As far as we know, there is no known graph-theoretic constructions of split-state non-malleable codes for $k \geq 2$. It would be of interest to generalize the graph-based codes in this paper and [36] for k-bit messages in the split-state model (see also [42, Section 2.1.3]).

Acknowledgement

The authors are grateful to Mr. Peter Rasmussen and Prof. Amit Sahai for their helpful comments to an earlier version of this paper. S. Satake has been supported by Grant-in-Aid for JSPS Fellows 20J00469 of the Japan Society for the Promotion of Science. Y. Gu has been supported by Grant-in-Aid for Early-Career Scientists 21K13830 of the Japan Society for the Promotion of Science. K. Sakurai has been supported by Grant-in-Aid for Scientific Research (B) 18H03240 of the Japan Society for the Promotion of Science.

References

[1] D. Aggarwal, S. Agrawal, D. Gupta, H. K. Maji, O. Pandey and M. Prabhakaran, "Optimal computational split-state non-malleable codes," in *Thirteenth IACR Theory of Cryptography Conference (TCC 2016-A)*, pp. 393–417, 2016.

[2] D. Aggarwal and J. Briët, "Revisiting the Sanders-Bogolyubov-Ruzsa theorem in $\mathbb{F}_p^n$ and its application to non-malleable codes," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1322-1326, 2016.

[3] D. Aggarwal, Y. Dodis and S. Lovett, "Non-malleable codes from additive combinatorics," *SIAM J. Comput.* vol. 47, no. 2, pp. 524–546, 2018.

[4] D. Aggarwal, Y. Dodis, T. Kazana and M. Obremski, "Non-malleable reductions and applications," in *47th Annual Symposium on the Theory of Computing (STOC 2015)*, pp. 459–468, 2015.

[5] D. Aggarwal and M. Obremski, "Inception makes non-malleable codes shorter as well!," *Cryptology ePrint Archive*, Report 2019/399, 2019.

[6] D. Aggarwal and M. Obremski, "A constant rate non-malleable code in the split-state model," in *IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS 2020)*, pp.1285–1294, 2020.

[7] D. Aggarwal, M. Obremski, J. L. Ribeiro, M. Simkin and L. Siniscalchi, "Computational and information-theoretic two-source (non-malleable) extractors," *Cryptology ePrint Archive*, Report 2020/259, 2020.

[8] R. W. Ahrens and G. Szekeres, "On a combinatorial generalization of 27 lines associated with a cubic surface," *J. Austral. Math. Soc.*, vol. 10, no. 3–4, pp. 485–492, 1969.

[9] A. E. Brouwer and W. H. Haemers, Spectra of Graphs, Springer, New York, 2012.

[10] E Chattopadhyay, V. Goyal and X. Li, "Non-malleable extractors and codes, with their many tampered extensions," in *48th Annual Symposium on the Theory of Computing (STOC 2016)*, pp. 285–298, 2016.

[11] E. Chattopadhyay and D. Zuckerman, "Non-malleable codes against constant split-state tampering," in *55th Annual Symposium on Foundations of Computer Science (FOCS 2014)*, pp. 306–315, 2014.

[12] S. M. Cioabă, F. Lazebnik and W. Li, On the spectrum of Wenger graphs, *J. Comb. Theory Ser. B*, vol. 107, pp. 132-139, 2014.

[13] F. Davì, S. Dziembowski and D. Venturi, "Leakage-resilient storage," in *Security and Cryptography for Networks*, J. A. Garay and R. De Prisco, eds., Lecture Notes in Comput. Sci. 6280, Springer, Berlin, pp. 121-137, 2010.

[14] G. Davidoff, P. Sarnak, A. Valette, Elementary Number Theory, Group Theory, and Ramanujan Graphs, Cambridge University Press, Cambridge, 2003.

[15] S. De Winter, J. Schillewaert and J. Verstraete, "Large incidence-free sets in geometries," *Electron. J. Combin.*, vol. 19, no. 4, #P24, 2012.

[16] Q. Diao, J. Li, S. Lin and I. F. Blake, New classes of partial geometries and their associated LDPC codes, *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 2947–2965, June, 2016.

[17] D. Dolev, C. Dwork, and M. Naor, "Non-malleable cryptography," *SIAM J. Comput.*, vol. 30, pp. 391–437, 2000.

[18] S. Dziembowski, T. Kazana and M. Obremski, "Non-malleable codes from two-source extractors," in *33rd Annual Cryptology Conference (CRYPTO 2013)*, pp. 239–257, 2013.

[19] S. Dziembowski, T. Kazana and M. Obremski, "Non-malleable codes from two-source extractors," *Cryptology ePrint Archive*, Report 2013/498, 2013.

[20] S. Dziembowski and K. Pietrzak, "Leakage-resilient cryptography," in *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2008)*, pp. 293–302, 2008.

[21] S. Dziembowski, K. Pietrzak and D. Wichs, "Non-malleable codes," in *Innovations in Computer Science (ICS 2010)*, pp. 434–452, 2010.

[22] S. Dziembowski, K. Pietrzak and D. Wichs, "Non-malleable codes," *J. ACM*, vol. 65, no. 4, pp. 20:1–20:32, 2018.

[23] W. Haemers, Eigenvalue Techniques in Design and Graph Theory, Ph.D. thesis, Eindhoven University of Technology, 1979.

[24] W. Haemers, "Interlacing eigenvalues and graphs," *Linear Algebra Appl.*, vol. 226/228, pp. 593–616, 1995.

[25] T. Høholdt and H. Janwa, "Eigenvalues and expansion of bipartite graphs," *Des. Codes Cryptogr.*, vol. 65, no. 3, pp. 259–273, 2012.

[26] F. Lazebnik and V. Ustimenko "New examples of graphs without small cycles and of large size," *European J. Combin.*, vol. 14 no.5, pp. 445–460, 1993.

[27] W.-C. W. Li and P. Solé, "Spectra of regular graphs and hypergraphs and orthogonal polynomials," *European J. Combin.* vol. 17, no. 5, pp. 461–477, 1996.

[28] W.-C. W. Li, M. Lu and C. Wang, "Recent developments in low-density parity-check codes," *Lecture Notes in Comput. Sci.*, vol. 5557, pp. 107–123, 2009.

[29] X. Li, "Improved non-malleable extractors, non-malleable codes and independent source extractors," In *49th Annual ACM Symposium on the Theory of Computing (STOC 2017)*, pp. 1144–1156, 2017.

[30] X. Li, "Non-malleable extractors and non-malleable codes: partially optimal constructions," *Cryptology ePrint Archive*, Report 2018/353, 2018.

[31] Z. Liu and D. A. Pados, "LDPC codes from generalized polygons," *IEEE Trans. Inform. Theory*, vol. 51, no. 11, pp. 3890–3898, Nov. 2005.

[32] A. Lubotzky, R. Phillips and P. Sarnak, “Ramanujan graphs,” *Combinatorica*, vol. 8, no. 3, pp. 261–277, 1988.

[33] G. A. Margulis, “Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators,” *Problems Inform. Transmission*, vol. 24, no. 1, pp. 39–46, 1988.

[34] S. Mohanty, R. O’Donnell and P. Paredes, “Explicit near-Ramanujan graphs of every degree,” in *52nd Annual ACM Symposium on Theory of Computing (STOC 2020)*, pp. 510–523, 2020.

[35] B. Nica, A Brief Introduction to Spectral Graph Theory, European Mathematical Society (EMS), Zürich, 2018.

[36] P. M. R. Rasmussen and A. Sahai, “Expander graphs are non-malleable codes,” in *Information-Theoretic Cryptography (ITC 2020)*, pp. 6:1–6:10, 2020.

[37] S. E. Payne and J. A. Thas, Finite Generalized Quadrangles. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[38] P. Sin, J. Sorci and Q. Xiang, “Linear representations of finite geometries and associated LDPC codes,” *J. Comb. Theory Ser. A.*, vol. 173, 105238, 2020.

[39] R. M. Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. Inform. Theory*, vol. 27, no. 5, pp. 533–547, Sep. 1981.

[40] R. M. Tanner, “Explicit concentrators from generalized N-gons,” *SIAM J. Algebraic Discrete Methods*, vol. 5, no. 3, pp. 287–293, 1984.

[41] H. van Maldeghem, Generalized Polygons, MBirkhäuser Verlag, Basel, 1998.

[42] M. Wang, On the Efficiency of Cryptographic Constructions, Ph.D thesis, Purdue University, 2021.

A Proof of Theorem 10

In this section, we prove Theorem 10. Although the proof is based on the discussion in [36], we give a full proof for reader’s convenience. In the proof we adopt the following notations. Let X, Y be two sets and $f : X \to Y$ be a function. For each $y \in Y$, denote $f^{-1}(y) := \{x \in X : f(x) = y\}$. For a subset $S \subset Y$, denote $f^{-1}(S) := \cup_{s \in S} f^{-1}(s)$. Also let $G = (V_1, V_2, E)$

be an (r,s)-biregular graph with n vertices, $\mu(G)=\mu$ and $\overrightarrow{G}=(V_1,V_2,\overrightarrow{E})$ be the associated orientation of G. Then for any pair of subsets $S\subset V_1$ and $T\subset V_2$, let

$$\begin{aligned} E(S,T) &:= \left|\{(s,t)\in\overrightarrow{E} : s\in S, t\in T\}\right|, \\ D(S,T) &:= \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}}\cdot|S||T| - E(S,T). \end{aligned} \tag{A.1}$$

Now recall the statement of Theorem 10.

(Theorem 10) *Let $G=(V_1,V_2,E)$ be a sufficiently large (r,s)-biregular graph which is a μ-spectral expander. Suppose that $|E|=\Omega\Big(\frac{(rs)^2\log(rs)}{\mu}\Big)$. Let $\mathcal{F}$ be the set of all functions $f=(g,h)$ with $g:V_1\to V_1$ and $h:V_2\to V_2$, where $f(v_1,v_2):=(g(v_1),h(v_2))$ for any $(v_1,v_2)\in V_1\times V_2$. Then $(\mathrm{enc}_G,\mathrm{dec}_G)$ is an $O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big)$-non-malleable code with respect to $\mathcal{F}$.*

To derive the non-malleability in Theorem 10, we shall employ Theorem 4 and the following lemmas.

Lemma 21. *Let $G=(V_1,V_2,E)$ be an (r,s)-biregular graph and $\overrightarrow{G}=(V_1,V_2,\overrightarrow{E})$ the associated orientation of G. For given functions $g:V_1\to V_1$ and $h:V_2\to V_2$, define $f:V_1\times V_2\to V_1\times V_2$ such that $f(v_1,v_2):=(g(v_1),h(v_2))$ for any $(v_1,v_2)\in V_1\times V_2$. Let*

$$T:=\frac{1}{2}\sum_{b\in\{0,1\}}\Pr\Big[\mathrm{dec}\big(f(\mathrm{enc}(b))\big)=1-b\Big].$$

Then we have

$$T=\frac{1}{2}+\delta\cdot\sum_{(v,w)\in\overrightarrow{E}}D\big(g^{-1}(v),h^{-1}(w)\big)$$

where

$$\delta:=\frac{|V_2|}{2r(|V_2|-r)|V_1|}=\frac{|V_1|}{2s(|V_1|-s)|V_2|}.$$

Proof of Lemma 21. The proof here is analogous to the proof of [36, Proposition 6]. For $b\in\{0,1\}$, let

$$Q_b:=\Pr\Big[\mathrm{dec}_G\big(f(\mathrm{enc}_G(b))\big)=1-b\Big].$$

Notice that

$$\begin{aligned} Q_0 &= \Pr_{(v,w)\leftarrow(V_1\times V_2)\setminus\overrightarrow{E}}\Big[(g(v),h(w))\in\overrightarrow{E}\Big], \\ Q_1 &= \Pr_{(v,w)\leftarrow\overrightarrow{E}}\Big[(g(v),h(w))\notin\overrightarrow{E}\Big], \end{aligned}$$

and thus $T=(Q_0+Q_1)/2$. Now we turn to compute Q_b.

First let $b=0$. For any $e=(v,w)\in\overrightarrow{E}$, the total number of non-edges of G mapped by f to e is

$$|\{(x,y)\in(V_1\times V_2)\setminus\overrightarrow{E} : f(x,y)=(g(x),h(y))=(v,w)\}|$$

$$= |g^{-1}(v)||h^{-1}(w)| - E\big(g^{-1}(v), h^{-1}(w)\big).$$

Since (2.1) implies $|(V_1 \times V_2) \setminus \overrightarrow{E}| = (|V_2| - r)|V_1| = (|V_1| - s)|V_2|$, we have

$$Q_0 = \frac{\sum_{(v,w)\in\overrightarrow{E}} \Big\{ |g^{-1}(v)||h^{-1}(w)| - E\big(g^{-1}(v), h^{-1}(w)\big) \Big\}}{(|V_2| - r)|V_1|}. \tag{A.2}$$

Next suppose $b = 1$. For any $e = (v, w) \in (V_1 \times V_2) \setminus \overrightarrow{E}$, the total number of edges of G mapped by f to e is $E\big(g^{-1}(v), h^{-1}(w)\big)$. Thus we have

$$\begin{aligned} Q_1 &= \frac{\sum_{(v,w)\notin\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|} \\ &= \frac{|\overrightarrow{E}| - \sum_{(v,w)\in\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|} \\ &= 1 - \frac{\sum_{(v,w)\in\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|}, \end{aligned} \tag{A.3}$$

where the last equality follows from (2.1).

Summing up (A.2) and (A.3) completes the proof. □

Let $f = (g, h) : V_1 \times V_2 \to V_1 \times V_2$ be a given tampering function from $\mathcal{F}$. Recall that for each pair of $1 \le i \ne j \le 2$ and each vertex $v \in V_i$, $N(v) = \{u \in V_j : u, v \text{ are adjacent in } G\}$. Define the following partitions of V_1 and V_2.

$$G^1 := \Big\{ v \in V_1 : |g^{-1}(v)| > \frac{|V_1|}{rs} \Big\}, \quad G^2 := \Big\{ v \in V_1 : |g^{-1}(v)| \le \frac{|V_1|}{rs} \Big\},$$
$$H^1 := \Big\{ w \in V_2 : |h^{-1}(w)| > \frac{|V_2|}{rs} \Big\}, \quad H^2 := \Big\{ w \in V_2 : |h^{-1}(w)| \le \frac{|V_2|}{rs} \Big\}.$$

For $1 \le i, j \le 2$, let

$$R_{i,j} := \delta \cdot \sum_{(v,w)\in\overrightarrow{E}\cap(G^i\times H^j)} D\Big(g^{-1}(v), h^{-1}(w)\Big).$$

It follows from Lemma 21 that

$$T = \frac{1}{2} + \sum_{1\le i,j\le 2} R_{i,j}. \tag{A.4}$$

By (A.4), the proof of Theorem 10 is completed from the following Lemmas 22, 23 and 24.

Lemma 22 ($i = 2$)**.**

$$R_{2,1} + R_{2,2} = O\Big(\frac{1}{r}\Big).$$

Proof. By the definition of $D(S,T)$ in (A.1),

$$
\begin{aligned}
R_{2,1} + R_{2,2} &\le \delta \cdot \sum_{(v,w)\in \overrightarrow{E}\cap(G^2\times V_2)} \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |g^{-1}(v)||h^{-1}(w)| \\
&\le \delta \cdot s \cdot \sum_{w\in V_2} \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot \frac{|V_1|}{rs} \cdot |h^{-1}(w)| \\
&\le \frac{|V_1|}{2s(|V_1|-s)|V_2|} \cdot s \cdot \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot \frac{|V_1|}{rs} \cdot |V_2| \\
&= O\left(\frac{1}{\sqrt{rs}} \cdot \sqrt{\frac{|V_1|}{|V_2|}}\right) = O\left(\frac{1}{r}\right)
\end{aligned}
$$

where the second inequality follows from the definition of G^2. □

Lemma 23 ($i = 1$, $j = 2$)**.**

$$
R_{1,2} = O\left(\frac{1}{s}\right).
$$

Proof. Similar to Case 1, we have

$$
\begin{aligned}
R_{1,2} &\le \delta \cdot \sum_{(v,w)\in \overrightarrow{E}\cap(G^1\times H^2)} \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |g^{-1}(v)||h^{-1}(w)| \\
&\le \delta \cdot \sum_{(v,w)\in \overrightarrow{E}\cap(V_1\times H^2)} \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |g^{-1}(v)||h^{-1}(w)| \\
&\le \delta \cdot r \cdot \sum_{v\in V_1} \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |g^{-1}(v)| \cdot \frac{|V_2|}{rs} \\
&\le \frac{|V_2|}{2r(|V_2|-r)|V_1|} \cdot r \cdot \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |V_1| \cdot \frac{|V_2|}{rs} \\
&= O\left(\frac{1}{\sqrt{rs}} \cdot \sqrt{\frac{|V_2|}{|V_1|}}\right) = O\left(\frac{1}{s}\right).
\end{aligned}
$$

□

Lemma 24 ($i = j = 1$)**.**

$$
R_{1,1} = O\left(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\right).
$$

Since the proof of Lemma 24 is complicated, we first prove Theorem 10 assuming that Lemma 24 holds.

Proof of Theorem 10. By Theorem 4 and (A.4), Theorem 10 immediately follows from Lemmas 22, 23 and 24. □

Now we are going to prove Lemma 24. We will make use of the following lemma which plays a key role to prove the technical Lemma 24.

Lemma 25 (Expander mixing lemma, [15], [23], [24]). *Let $G = (V_1, V_2, E)$ be an (r, s)-biregular graph with n vertices, $\mu(G) = \mu$. Then for any pair of subsets $S \subset V_1$ and $T \subset V_2$, we have*

$$|D(S,T)| \leq \mu\sqrt{|S||T|}. \tag{A.5}$$

Remark 26. The non-malleable codes from [36] used the following fact. Let $G = (V, E)$ be a d-regular (possibly non-bipartite) graph with $\lambda(G) = \lambda$. Then for any pair of subsets $S, T \subset V$,

$$\left|\frac{d}{n}|S||T| - e(S,T)\right| \leq \lambda\sqrt{|S||T|}. \tag{A.6}$$

Here $e(S,T)$ denotes the number of edges between S and T. However, if G is a bipartite graph, the estimation (A.6) cannot be used to prove the non-malleability for the coding schemes in [36] (see Appendix) and the coding scheme in this paper (see Definition 8), since in this case $\lambda(G) = d$ (see Lemma 5), which only implies $O(\sqrt{d})$-non-malleable codes. However we could see from Theorem 10 that using Lemma 25 can produce $o(1)$-non-malleable codes.

Proof of Lemma 24. Take partitions of G^1 and H^1 so that for each pair of $1 \leq k, l \leq \lceil \log_2(rs) \rceil$,

$$G^1(k) := \Big\{v \in G_1 : \frac{|V_1|}{2^{k-1}} \geq |g^{-1}(v)| \geq \frac{|V_1|}{2^k}\Big\},$$
$$H^1(l) := \Big\{w \in H_1 : \frac{|V_2|}{2^{l-1}} \geq |h^{-1}(w)| \geq \frac{|V_2|}{2^l}\Big\}.$$

For each pair of $1 \leq k, l \leq \lceil \log_2(rs) \rceil$, let

$$S_{k,l} := \delta \cdot \sum_{(v,w) \in \overrightarrow{E} \cap \left(G^1(k) \times H^1(l)\right)} D\Big(g^{-1}(v), h^{-1}(w)\Big).$$

Since $R_{1,1} = \sum_{1 \leq k,l \leq \lceil \log_2(rs) \rceil} S_{k,l}$, Lemma 24 follows from the following.

$$\sum_{1 \leq k,l \leq \lceil \log_2(rs) \rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{A.7}$$

To that end, we divide the sum in (A.7) into two parts, namely,

$$\sum_{1 \leq k \leq l \leq \lceil \log_2(rs) \rceil} S_{k,l} \quad \text{and} \quad \sum_{1 \leq l < k \leq \lceil \log_2(rs) \rceil} S_{k,l}.$$

Case 1. This case is to prove the following estimation.

$$\sum_{1 \leq k \leq l \leq \lceil \log_2(rs) \rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{A.8}$$

First we have

$$
\begin{aligned}
\delta^{-1} S_{k,l} &= \sum_{v \in G^1(k)} D\left(g^{-1}(v), \bigcup_{w \in N(v) \cap H^1(l)} h^{-1}(w)\right) \\
&\le \sum_{v \in G^1(k)} \mu \sqrt{|g^{-1}(v)| \cdot \sum_{w \in N(v) \cap H^1(l)} |h^{-1}(w)|} \\
&\le \mu \sqrt{\frac{|V_1|}{2^{k-1}} \cdot \frac{|V_2|}{2^{l-1}}} \sum_{v \in G^1(k)} \sqrt{|N(v) \cap H^1(l)|} \\
&\le 2\mu \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt{|V_1||V_2|} \cdot \sqrt{|G^1(k)|} \cdot \sqrt{E\big(G^1(k), H^1(l)\big)} \\
&\le 2\mu \cdot 2^{-\frac{l+k}{2}} \cdot \sqrt{|V_1||V_2|} \cdot \sqrt{|G^1(k)|} \cdot \sqrt{\frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |G^1(k)||H^1(l)| + \mu\sqrt{|G^1(k)||H^1(l)|}},
\end{aligned}
$$

where the second and last inequalities follow from Lemma 25.

By Jensen's inequality and (2.1), we obtain

$$
S_{k,l} \le O\Big(\frac{\mu}{\sqrt{|E|}}\Big) \cdot 2^{-\frac{l+k}{2}} \cdot |G^1(k)| \cdot \sqrt{|H^1(l)|} + O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big) \cdot 2^{-\frac{l+k}{2}} \cdot \Big(|G^1(k)|^3|H^1(l)|\Big)^{\frac{1}{4}}.
$$

To obtain (A.8), let

$$
\begin{aligned}
L &:= \sum_{1 \le k \le l \le \lceil \log_2(rs) \rceil} 2^{-\frac{l+k}{2}} \cdot |G^1(k)| \cdot \sqrt{|H^1(l)|}, \\
K &:= \sum_{1 \le k \le l \le \lceil \log_2(rs) \rceil} 2^{-\frac{l+k}{2}} \cdot \Big(|G^1(k)|^3|H^1(l)|\Big)^{\frac{1}{4}}.
\end{aligned}
$$

By the definitions of L and K,

$$
\sum_{1 \le k \le l \le \lceil \log_2(rs) \rceil} S_{k,l} = O\Big(\frac{\mu}{\sqrt{|E|}}\Big) \cdot L + O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big) \cdot K. \tag{A.9}
$$

First we estimate L. Notice that for each $k \le \lceil \log_2(rs) \rceil$,

$$
|G^1(k)| \cdot 2^{-\frac{k}{2}} \le 2^{\frac{k}{2}} \le 2\sqrt{rs}. \tag{A.10}
$$

Then by the Cauchy-Schwartz inequality,

$$
\begin{aligned}
L &\le \sum_{1 \le k,l \le \lceil \log_2(rs) \rceil} 2^{-\frac{l+k}{2}} \cdot |G^1(k)| \cdot \sqrt{|H^1(l)|}, \\
&\le 2\sqrt{rs} \cdot \sum_{1 \le l \le \lceil \log_2(rs) \rceil} \sqrt{2^{-l}|H^1(l)|} \\
&\le O\Big(\sqrt{rs \log(rs)}\Big) \cdot \sqrt{\sum_{1 \le l \le \lceil \log_2(rs) \rceil} 2^{-l}|H^1(l)|},
\end{aligned}
$$

where the second inequality follows from (A.10). On the other hand, the definition of $H^1(l)$ implies that

$$|h^{-1}(H^1(l))| \geq \frac{|V_2||H^1(l)|}{2^l}. \tag{A.11}$$

Since $H^1(1), \ldots, H^1(\lceil \log_2(rs) \rceil)$ are disjoint subsets of V_2, we have

$$L = O\Big(\sqrt{rs \log(rs)}\Big) \cdot \sqrt{\sum_{1 \leq l \leq \lceil \log_2(rs) \rceil} \frac{|h^{-1}(H^1(l))|}{|V_2|}} = O\Big(\sqrt{rs \log(rs)}\Big), \tag{A.12}$$

where the last equation follows from (A.11).

Next we aim to bound K. Since we are assuming that $k \leq l$, setting $t = l - k$, we obtain

$$\begin{aligned} K &\leq \sum_{1 \leq k \leq l \leq \lceil \log_2(rs) \rceil} \frac{2^{\frac{k-l}{4}}}{(|V_1|^3 |V_2|)^{\frac{1}{4}}} \Big(|g^{-1}(G^1(k))|^3 \cdot |h^{-1}(H^1(l))|\Big)^{\frac{1}{4}} \\ &\leq \sum_{t=0}^{\lceil \log_2(rs) \rceil} \frac{2^{-\frac{t}{4}}}{(|V_1|^3 |V_2|)^{\frac{1}{4}}} \sum_{l=t}^{\lceil \log_2(rs) \rceil} \Big(|g^{-1}(G^1(l-t))|^3 \cdot |h^{-1}(H^1(l))|\Big)^{\frac{1}{4}}, \end{aligned}$$

where the first inequality follows from (A.11) and the following inequality.

$$|g^{-1}(G^1(k))| \geq \frac{|V_1||G^1(k)|}{2^k}. \tag{A.13}$$

By the definitions of $G^1(k)$ and $H^1(l)$, for each $0 \leq t \leq \lceil \log_2(rs) \rceil$, the sets $g^{-1}(G^1(l-t))$ and $h^{-1}(G^1(l))$, $t \leq l \leq \lceil \log_2(rs) \rceil$, are disjoint subsets of V_1 and V_2, respectively. Then it follows from Hörder's inequality that

$$\begin{aligned} K &\leq \sum_{t=0}^{\lceil \log_2(rs) \rceil} \frac{2^{-\frac{t}{4}}}{(|V_1|^3 |V_2|)^{\frac{1}{4}}} \left(\sum_{l=t}^{\lceil \log_2(rs) \rceil} \Big(|g^{-1}(G^1(l-t))|\Big) \right)^{\frac{3}{4}} \cdot \left(\sum_{l=t}^{\lceil \log_2(rs) \rceil} |h^{-1}(H^1(l))| \right)^{\frac{1}{4}} \\ &\leq \sum_{t=0}^{\lceil \log_2(rs) \rceil} 2^{-\frac{t}{4}} = O(1). \end{aligned} \tag{A.14}$$

By (A.9), (A.12) and (A.14), we get

$$\begin{aligned} \sum_{1 \leq k \leq l \leq \lceil \log_2(rs) \rceil} S_{k,l} &= O\Big(\frac{\mu}{\sqrt{|E|}} \cdot \sqrt{rs \log(rs)}\Big) + O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big) \\ &= O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big), \end{aligned}$$

where the last equality follows from the condition in Theorem 10 that $|E| = \Omega\Big(\frac{(rs)^2 \log(rs)}{\mu}\Big)$. This concludes the discussion for Case 1.

Case 2. In this case, we aim to show the following.

$$\sum_{1\le l<k\le\lceil\log_2(rs)\rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{A.15}$$

To prove (A.15), we deal with the following equation.

$$\delta^{-1}S_{k,l} = \sum_{w\in H^1(l)} D\Bigg(\bigcup_{v\in N(w)\cap G^1(k)} g^{-1}(v),\ h^{-1}(w)\Bigg).$$

This follows from an analogous calculation as in Case 1.

Combining Cases 1 and 2 yields (A.7). This completes the proof of Lemma 24. □