# Graph-Based Construction for Non-Malleable Codes

Shohei Satake * Yujie Gu † Kouichi Sakurai ‡

February 20, 2021

**Abstract**

Non-malleable codes protect communications against adversarial tampering of data, which can be seen as a relaxation of error-correcting codes and error-detecting codes. Recently, Rasmussen and Sahai (ITC2020) explicitly constructed non-malleable codes in the split-state model using expander graphs. In this paper we extend their construction by means of bipartite expander graphs. The resulted codes can have flexible parameters and reduce the encoding space cost in comparison with the explicit codes by Rasmussen and Sahai.

## 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 $\mathrm{enc} : \mathcal{M} \to \mathcal{X}$, 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 guarantee the reliable communication with respect to the family $\mathcal{F}$ such 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

*Faculty of Advanced Science and Technology, Kumamoto University, 2-39-1, Kurokami, Chuo, Kumamoto, Japan, 860-8555. e-mail: shohei-satake@kumamoto-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: 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

$\mathcal{F}$), which ensure that either the tampered codeword can be 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, 22], the concept of non-malleable codes is in a sprit of non-malleablity proposed by Dolev, Dwork and Naor [17] in cryptographic primitives. Informally speaking, 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 [14, 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 an *equally-sized* split-state model.

Prior to a recent work [35], all known constructions of non-malleable codes for the split-state model have relied on complex mathematical proofs based on two-source extractors and additive combinatorics, see [1, 2, 3, 4, 5, 6, 12, 13, 18, 28, 29] for example. Moreover, Dziembowski, Kazana and Obremski 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" [18]. Very recently, Rasmussen and Sahai [35] discovered that expander graphs could provide non-malleable codes for the split-state model and single-bit messages. Their proof relies on the edge-counting technique for the underlying graph, together with eigenvalue evaluation of its adjacency matrix, see Remark 8 and Appendix. However, their non-malleable codes are only for equally-sized split-state model where the size of each part is equal to the number of vertices of the underlying graph. We noticed that the construction in the paper [35] cannot directly be transferred to the general split-state model since the symmetry of two states is required in their proofs which, however, does not hold in the general case. Inspired by this, we attempted to instantiate Rasmussen and Sahai's construction based on bipartite graphs, which are a typical class of expander graphs having been developed in various areas such as coding theory, number theory and combinatorics, see e.g. [9, 10, 16, 25, 26, 38, 40, 44]. However, we noticed that the proofs in [35] cannot be used to verify the non-malleability if the underlying graph is a bipartite graph in general (see Remark 8).

In this work, we establish a coding scheme based on bipartite graphs. In particular we prove that when the underlying bipartite graph is an $(r, s)$-biregular graph with the second largest eigenvalue $\mu$, 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. Our construction can be seen as an extension of the construction in [35] in the sense that we could deduce the codes for equally-sized split-state model in [35] as special cases (see Remark 16). As in [35] our proof is based on edge-counting technique for the underlying biregular graph as well. However, we apply the expander mixing lemma for biregular graphs, which is different from the fact used in [35]. Also the analysis of non-malleability in [35] needs to be modified to deal with biregular graphs, see Remark 21. Moreover, we instantiate our construction based on some specific bipartite graphs (see Table 1), obtaining explicit non-malleable codes which reduce the encoding space cost in comparison with the explicit codes in [35]. Specifically, for given $0 < \varepsilon < 1$, our instantiation provides explicit $\varepsilon$-non-malleable codes using $20\log(1/\varepsilon) + O(\log\log(1/\varepsilon))$ space for encoding each messeage, while the instantiation in [35] only gives explicit codes using $24\log(1/\varepsilon) + O(1)$ space, see Example 17.

| Ref. | $\lvert\mathcal{L}\rvert$ | $\lvert\mathcal{R}\rvert$ | equally-sized? | encoding space cost | comments |
|---|---|---|---|---|---|
| [35, Section C] | $q^3$ | $q^3$ | Yes | $24\log(1/\varepsilon) + O(1)$ | $q = p^2$, $p$ is a prime |
| Example 17 | $\Theta(p^{5/2}\log(p))$ | $\Theta(p^{5/2}\log(p))$ | Yes | $20\log(1/\varepsilon) + O(\log\log(1/\varepsilon))$ | $p$ is a prime |
| Example 19 | $(q+2)q^2$ | $q^3$ | No | $24\log(1/\varepsilon) + O(1)$ | $q$ is a prime power |

Table 1: Parameters of explicit $\varepsilon$-non-malleable codes in this paper and [35]

The remainder of this paper is organized as follows. Section 2 briefly reviews non-malleable codes and the basics in graph theory. Section 3 provides the coding scheme and shows the main theorems in this paper (Theorems 12 and 13). Section 4 presents some explicit non-malleable codes. Sections 5 and 6 prove the two main theorems respectively. Section 7 concludes this paper. Finally Appendix proves a technical lemma used in Section 6 and briefly reviews the code designed by Rasmussen and Sahai [35].

## 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.

### 2.1 Non-malleable codes

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 $\mathrm{dec} : \mathcal{X} \to \mathcal{M} \cup \{\perp\}$ 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 1** (Split-state non-malleable codes)**.** In the *split-state model*, assume $\mathcal{X}=\mathcal{L}\times\mathcal{R}$ is the product set of $\mathcal{L}$ and $\mathcal{R}$. Let

$$\mathcal{F} := \{f=(g,h) : g:\mathcal{L}\to\mathcal{L},\ h:\mathcal{R}\to\mathcal{R}\},$$

and for each $(L,R)\in\mathcal{L}\times\mathcal{R}$, $f(L,R) := (g(L),h(R))$. 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 *$\varepsilon$-non-malleable code with respect to $\mathcal{F}$* if for every $f=(g,h)\in\mathcal{F}$, there exists a distribution $D_f$ on $\mathcal{M}\cup\{\text{same}^*,\perp\}$ such that for every $m\in\mathcal{M}$ and the following two random variables $A_f^m, B_f^m$, we have $\Delta(A_f^m,B_f^m)\le\varepsilon$.

$$A_f^m := \left\{\begin{array}{c}(L,R)\leftarrow\mathrm{enc}(m);\\ \text{Output } \mathrm{dec}(g(L),h(R))\end{array}\right\},$$
$$B_f^m := \left\{\begin{array}{c}\tilde{m}\leftarrow D_f;\\ \text{If } \tilde{m}=\text{same}^* \text{ output } m, \text{ else output } \tilde{m}\end{array}\right\}.$$

Hereafter, as in [18] and [35], the symbol "$\perp$" from Definition 1 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.

This paper focuses on *single-bit* non-malleable codes, i.e., $\mathcal{M}=\{0,1\}$. It is shown by Dziembowski, Kazana and Obremski [18] that single-bit non-malleable codes can also be formulated as in the following Theorem 2.

**Theorem 2** ([18, 19])**.** *Let* $(\mathrm{enc},\mathrm{dec})$ *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* $(\mathrm{enc},\mathrm{dec})$ *is an $\varepsilon$-non-malleable code with respect to $\mathcal{F}$ if and only if it holds for every $f\in\mathcal{F}$ that*

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

*where the probability is taken over the randomness of* enc.

## 2.2 Expander graphs

Throughout this paper, we assume that all graphs are undirected and simple, i.e., with no 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* $G$.

A graph $G$ is called a *$d$-regular graph* if every vertex of $G$ connects exactly $d$ edges. A bipartite graph $G = (V_1, V_2, E)$ is called an *$(r, s)$-biregular graph* if every vertex of $V_1$ and $V_2$ connects exactly $r$ and $s$ edges, respectively.

**Lemma 3** (Handshaking lemma)**.** *For an $(r, s)$-biregular graph $G = (V_1, V_2, E)$ and its associated orientation $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$, it holds that*

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

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

**Lemma 4** (e.g. [11], [15])**.** *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 $\lambda_n = -\sqrt{rs}$.*

According to Lemma 4, the largest eigenvalue of a (bi-)regular graph is always determined. However, the second largest eigenvalue usually has rich properties. For a $d$-regular graph $G$, denote $\lambda(G) := \max\limits_{2 \leq i \leq n} |\lambda_i|$. 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 *$\mu$-spectral expander* if $\mu(G) \leq \mu$. It has the following nice property.

**Proposition 5** ([40])**.** *Let $G = (V_1, V_2, E)$ be an $(r, s)$-biregular graph which is a $\mu$-spectral expander. For a subset $S \subset V_1$, define the neighbour of $S$ as*

$$N(S) := \{u \in V_2 : u \text{ is adjacent to some vertex in } S\},$$

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

**Lemma 6** ([42]). *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 $\mu(G) = \Omega(\sqrt{r+s})$.*

It is recently proved ([45]) that for $s \geq r$ with $s = O(|V_1|^{2/3})$, the random $(r,s)$-biregular graphs with vertex set $V_1 \cup V_2$ are $O(\sqrt{s})$-spectral expanders with high probability, which are optimal (up to a constant) with respect to Lemma 6. On the other hand, it is in general non-trivial to *explicitly* construct such spectral expanders.

We will make use of the following lemma in our proofs.

**Lemma 7** (Expander mixing lemma, [15], [23], [24]). *Let $G = (V_1, V_2, E)$ be an $(r,s)$-biregular graph with $n$ vertices, $\mu(G) = \mu$ and $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$ be the associated orientation of $G$. For any pair of subsets $S \subset V_1$ and $T \subset V_2$, let*

$$\begin{aligned} E(S,T) &:= \left|\{(s,t) \in \overrightarrow{E} : s \in S, t \in T\}\right|, \\ D(S,T) &:= \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot |S||T| - E(S,T). \end{aligned} \tag{2.1}$$

*Then we have*

$$|D(S,T)| \leq \mu\sqrt{|S||T|}. \tag{2.2}$$

**Remark 8.** The non-malleable codes from [35] 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 $S, T \subset V$,

$$\left|\frac{d}{n}|S||T| - e(S,T)\right| \leq \lambda\sqrt{|S||T|}. \tag{2.3}$$

Here $e(S,T)$ denotes the number of edges between $S$ and $T$. However, if $G$ is a bipartite graph, the estimation (2.3) cannot be used to prove the non-malleability for the coding schemes in [35] (see Appendix) and the coding scheme in this paper (see Definition 10), since in this case $\lambda(G) = d$ (see Lemma 4), which only implies $O(\sqrt{d})$-non-malleable codes. However we will see in Theorem 13 that Lemma 7 can produce $o(1)$-non-malleable codes.

**Remark 9.** In the literature, extractors are the main pseudo-random objects applied for constructing non-malleable codes, see [1, 4, 5, 6, 12, 13, 18, 28, 29] for example. It is worth noting that extractors can provide bipartite graphs with "stronger" expansion properties (see [37, 43]) than the one guaranteed in Proposition 5 based on spectral expanders. However it turns out in Section 3 that the bipartite spectral expanders are in fact enough to obtain non-malleable codes.

# 3 Construction

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

## 3.1 Candidate codes

First we propose a coding scheme based on bipartite graphs.

**Definition 10** (Candidate codes)**.** Let $G = (V_1, V_2, E)$ be a bipartite graph and $\overrightarrow{G} = (V_1, V_2, \overrightarrow{E})$ be the associated orientation of $G$. Then the associated graph code $(\mathrm{enc}_G, \mathrm{dec}_G)$ 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 11.** In [35], Rasmussen and Sahai designed a coding scheme based on a graph $G = (V, E)$ so that the space of codewords is $V \times V$ (see Appendix), but it works only for equally-sized split-state model with $|\mathcal{L}| = |\mathcal{R}| = |V|$. Note that here our scheme can be applied to a more flexible split-state model in the sense that $|\mathcal{L}| = |V_1|$ is not necessarily equal to $|\mathcal{R}| = |V_2|$.

## 3.2 Non-malleable codes from biregular graphs

In this subsection we aim to verify the non-malleability of the codes in Definition 10. First, the following theorem, together with Theorem 2, shows that for a given $(r, s)$-biregular graph $G = (V_1, V_2, E)$, checking the non-malleability of the codes in Definition 10 can be done by the edge-counting for each subgraph of $G$. In particular, this theorem will provide a feasible way of using the expander mixing lemma (Lemma 7) to prove the non-malleability of the coding scheme in Definition 10.

**Theorem 12.** *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*

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

The proof of Theorem 12 is deferred to Section 5. Applying Lemma 7 to Theorem 12, we obtain the following theorem.

**Theorem 13.** *Let $G = (V_1, V_2, E)$ be a sufficiently large $(r, s)$-biregular graph which is a $\mu$-spectral expander. Suppose that $|E| = \Omega\Big(\frac{(rs)^2 \log(rs)}{\mu}\Big)$, equivalently, $\sqrt{|V_1||V_2|} = \Omega\Big(\frac{(rs)^{\frac{3}{2}} \log(rs)}{\mu}\Big)$. Let $\mathcal{F}$ be the set of all functions $f = (g, h)$ with $g : V_1 \to V_1$ and $h : V_2 \to V_2$. 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 $\mathcal{F}$.*

The proof of Theorem 13 can be found in Section 6.

**Remark 14.** Suppose that $G$ is an $(r, s)$-biregular graph with $s \geq r = \omega(\sqrt{s})$ and $\mu(G) = O(\sqrt{s})$. Then Theorem 13 guarantees that $(\mathrm{enc}_G, \mathrm{dec}_G)$ 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 4, the quantity $O(s^{1/4}/r^{1/2})$ in Theorem 13 is best possible up to a constant.

The following corollary follows from Theorem 13 directly.

**Corollary 15.** *Let $G = (V_1, V_2, E)$ be a bipartite $d$-regular graph with $|V_1| = |V_2| = n$ which is a $\mu$-spectral expander. Suppose that $n = \Omega\Big(\frac{\log(d)\cdot d^3}{\mu}\Big)$ and $\mathcal{F}$ is as in Theorem 12. Then $(\mathrm{enc}_G, \mathrm{dec}_G)$ is an $O\Big(\frac{\mu^{3/2}}{d}\Big)$-non-malleable code with respect to $\mathcal{F}$.*

As a special case of Theorem 13, Corollary 15, together with Lemma 3, guarantees the non-malleability only for equally-sized split model.

**Remark 16.** Corollary 15 actually can deduce the non-malleable codes based on Cayley graphs in [35, Section C]. In fact, 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 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* [34]. For a prime $p$ let $\mathbb{F}_p$ denote the $p$-element field and $q = p^2$. In [35], Rasmussen and Sahai constructed $O(q^{-1/4})$-non-malleable codes from a non-bipartite graph $Cay(\mathbb{F}_p^6, S)$ with some $S \subset \mathbb{F}_p^6$ such that $|S| = q$. According to Corollary 15, the corresponding bi-Cayley graph provides the same non-malleable code as in [35].

# 4 Instantiation

In this section, for a given real number $0 < \varepsilon < 1$, we present explicit $\varepsilon$-non-malleable codes based on specific spectral expanders.

We note that in the presented codes below, both the encoding and decoding for each message costs time $O(\log(1/\varepsilon))$, which is same (up to a constant) as the explicit codes from [35, Section C]. However, this work improves upon [35] in the following two folds. First, our codes can reduce the encoding space cost in comparison of the explicit codes from [35, Section C]. Second, the parameters of our codes can be taken more flexibly (see Example 19).

We briefly explain the first improvement regarding the encoding space cost. Recall from Corollary 15 that an $\varepsilon$-non-malleable code can be derived from a bipartite $d$-regular graph $G = (V_1, V_2, E)$ with $d = \Theta(1/\varepsilon)^4$, $\mu(G) = \Theta(1/\varepsilon)^2$ and $|V_1| = |V_2| = \Theta(1/\varepsilon)^{10}\log(1/\varepsilon)$, assuming $\mu(G) = \Theta(\sqrt{d})$ which is optimal with respect to Lemma 6. Since $|V_1 \times V_2| = \Theta((1/\varepsilon)^{20}(\log(1/\varepsilon))^2)$, in this scheme, encoding each bit uses $20\log(1/\varepsilon) + O(\log\log(1/\varepsilon))$ space, which coincides to the observation in [35, Section 1.3]. However [35] only provided explicit $\varepsilon$-non-malleable codes using encoding space $24\log(1/\varepsilon) + O(1)$ (see also Remark 16). Here the following Example 17 can produce explicit $\varepsilon$-non-malleable codes achieving the encoding space cost $20\log(1/\varepsilon) + O(\log\log(1/\varepsilon))$, by means of explicit bipartite $(p+1)$-regular graphs $G$ with $\Theta(p^{5/2}\log(p))$ vertices and $\mu(G) \leq 2\sqrt{p}$, where $p$ is a given prime.

**Example 17** (LPS Ramanujan graphs)**.** Let $p_1, p_2$ be two distinct primes such that $p_2 > 2\sqrt{p_1}$ and $p_1$ is a quadratic non-residue modulo $p_2$. Then in [32], Lubotzky, Phillips and Sarnak explicitly constructed a bipartite $(p_1+1)$-regular graph $X^{p_1,p_2}$ with $p_2((p_2)^2-1)$ vertices and $\mu(X^{p_1,p_2}) \leq 2\sqrt{p_1}$.

Now for each sufficiently large prime $p$, by Bertrand's postulate, there exists a prime $p' = \Theta(p^{5/6}\log^{1/3}(p)) > 2\sqrt{p}$, which could be found in poly$(p)$-time. Note that for primes $p$ and $p' = \Theta(p^{5/6}\log^{1/3}(p))$, the adjacency list of $X^{p,p'}$ is computable in poly$(\log(p))$-time, and hence the graph can be constructed in poly$(p)$-time, see [33] for example. If $p$ is a quadratic non-residue modulo $p'$, then $X^{p,p'}$ is a bipartite $(p+1)$-regular graph with $\Theta(p^{5/2}\log(p))$ vertices. Thus according to Corollary 15 the graph $X^{p,p'}$ with $p' = \Theta(p^{5/6}\log^{1/3}(p))$ provides an $O(p^{-1/4})$-non-malleable code for an equally-sized split-state model with $|\mathcal{L}| = |\mathcal{R}| = \Theta(p^{5/2}\log(p))$.

**Remark 18.** In Example 17, if $p_1$ is a quadratic residue modulo $p_2$, one can instead explicitly construct a non-bipartite $(p_1+1)$-regular graph $Y^{p_1,p_2}$ with $p_1((p_1)^2-1)/2$ vertices and $\lambda(Y^{p_1,p_2}) \leq 2\sqrt{p_1}$, see [32]. These graphs can be applied to the coding scheme in [35] (see Appendix).

Next, to show the second improvement, we introduce a typical example of expander biregular graphs, see e.g. [31, 40, 42]. Note that for given $0 < \varepsilon < 1$, these graphs provide $\varepsilon$-non-malleable codes with encoding space cost $24\log(1/\varepsilon)+O(1)$, which is same as the instantiations

in [35]. However, unlike the explicit codes from [35] which are only for equally-sized scenario, these explicit non-malleable codes are valid for the *non*-equally-sized split state model as well.

**Example 19** (Generalized quadrangles)**.** A *generalized quadrangle* of order $(\alpha, \beta)$ is an $(\alpha + 1, \beta + 1)$-biregular graph $GQ(\alpha, \beta) = (V_1, V_2, E)$ such that

1. For all $x, y \in V_1 \cup V_2$, there exists a path of length $\leq 4$ connecting $x$ and $y$;
2. 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 $y$;
3. 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 $y$.

It is known that $|V_1| = (\alpha+1)(\alpha\beta+1)$, $|V_2| = (\beta+1)(\alpha\beta+1)$ and $\mu(GQ(\alpha, \beta)) = \sqrt{\alpha + \beta}$; see [36, Section 1.2], [41, Corollary 1.5.5] and [40]. For every prime power $q$, there exists an explicit $GQ(q - 1, q + 1)$ (see [8, Sections 4 and 5]). Since $|E| = r|V_1| = \Theta(q^4)$ and $\frac{(rs)^2 \log(rs)}{\mu} = \Theta(q^{7/2} \log q)$, we could conclude by Theorem 13 that $(\mathrm{enc}_G, \mathrm{dec}_G)$ is an $O(q^{-1/4})$-non-malleable code for a split-state model with $|\mathcal{L}| = (q + 2)q^2$ and $|\mathcal{R}| = q^3$.

# 5 Proof of Theorem 12

In Sections 5 and 6, 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 $f^{-1}(S) := \cup_{s \in S} f^{-1}(s)$.

The proof here is analogous to the proof of [35, 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 $Q_b$.

**Case 1** $(b = 0)$**.** For any $e = (v, w) \in \overrightarrow{E}$, the total number of non-edges of $G$ mapped by $f$ to $e$ is

$$\begin{aligned} &|\{(x, y) \in (V_1 \times V_2) \setminus \overrightarrow{E} : f(x, y) = (g(x), h(y)) = (v, w)\}| \\ &= |g^{-1}(v)||h^{-1}(w)| - E\big(g^{-1}(v), h^{-1}(w)\big). \end{aligned}$$

Since Lemma 3 implies $|(V_1 \times V_2) \setminus \overrightarrow{E}| = (|V_2| - r)|V_1| = (|V_1| - s)|V_2|$, we have

$$Q_0 = \frac{\sum_{(v,w)\in\overrightarrow{E}} \Big\{ |g^{-1}(v)||h^{-1}(w)| - E\big(g^{-1}(v), h^{-1}(w)\big) \Big\}}{(|V_2| - r)|V_1|}. \tag{5.1}$$

**Case 2** ($b = 1$)**.** 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 $E\big(g^{-1}(v), h^{-1}(w)\big)$. By Lemma 3, we have

$$\begin{aligned} Q_1 &= \frac{\sum_{(v,w)\notin\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|} \\ &= \frac{|\overrightarrow{E}| - \sum_{(v,w)\in\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|} \\ &= 1 - \frac{\sum_{(v,w)\in\overrightarrow{E}} E\big(g^{-1}(v), h^{-1}(w)\big)}{r|V_1|}, \end{aligned} \tag{5.2}$$

where the last equality follows from Lemma 3.

Summing up (5.1) and (5.2) completes the proof.

# 6 Proof of Theorem 13

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 $v \in V_i$, $N(v) = \{u \in V_j : u, v \text{ are adjacent in } G\}$.

Now define the following partitions of $V_1$ and $V_2$.

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

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

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

Since Theorem 12 shows that $T = 1/2 + \sum_{1\le i,j\le 2} R_{i,j}$, it suffices to evaluate $R_{i,j}$ for each pair of $1 \le i, j \le 2$.

**Case 1** ($i = 2$) By the definition of $D(S, T)$ in (2.1),

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

$$\leq \frac{|V_1|}{2s(|V_1|-s)|V_2|} \cdot s \cdot \frac{\sqrt{rs}}{\sqrt{|V_1||V_2|}} \cdot \frac{|V_1|}{rs} \cdot |V_2|$$
$$= O\left(\frac{1}{\sqrt{rs}} \cdot \sqrt{\frac{|V_1|}{|V_2|}}\right) = O\Big(\frac{1}{r}\Big)$$

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

**Case 2** ($i = 1$, $j = 2$) Similarly to Case 1, we have

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

**Case 3** ($i = j = 1$) This is the most complicate case to evaluate $R_{i,j}$. Now 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 $R_{1,1} = \sum_{1\leq k,l\leq\lceil\log_2(rs)\rceil} S_{k,l}$, the proof completes by showing that

$$\sum_{1\leq k,l\leq\lceil\log_2(rs)\rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{6.1}$$

To that end, we discuss in the following cases.

*Case 3-1* ($k \leq l$) By applying Lemma 7,

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

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

where the second and last inequalities follow from Lemma 7.

By Jensen's inequality and Lemma 3, we obtain

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

To complete the discussions for Case 3-1, we need the following lemma, which is proved in the Appendix.

**Lemma 20.**

$$
\sum_{1\leq k\leq l\leq \lceil \log_2(rs)\rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{6.2}
$$

*Case 3-2* ($k > l$) We deal with the following equation.

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

By an analogous calculation as in Case 3-1 and Lemma 20, we have

$$
\sum_{1\leq l<k\leq \lceil \log_2(rs)\rceil} S_{k,l} = O\Big(\frac{\mu^{\frac{3}{2}}}{\sqrt{rs}}\Big). \tag{6.3}
$$

Combining (6.2) and (6.3) yields (6.1). This completes the proof.

**Remark 21.** The presented proof here is along with the analysis in [35]. However the discussion in [35] relies on the symmetry of $\mathcal{L}$ and $\mathcal{R}$, which does not hold for biregular graphs in general by Lemma 3. This difference mainly involves the discussion of Case 3. Particularly, compared to Case 3-1, the role of $G(k)$ and $H(l)$ for evaluating $\delta^{-1}S_{k,l}$ needs to be switched in Case 3-2.

## 7 Conclusion, remarks and problems

In this paper, we proposed a coding scheme based on bipartite graphs as an extension of the construction in [35]. In particular we showed that the non-malleability can be satisfied if the underlying bipartite graph is a biregular $\mu$-spectral expander with sufficiently small $\mu$. Moreover, we instantiated the coding scheme on specific spectral expanders and obtained explicit and efficient non-malleable codes. Our results provide more evidence, other than the work in [35], to that constructing non-malleable codes for the split-state model is not necessarily based on extractors, which answers the question by Dziembowski, Kazana and Obremski [18].

We would like to mention that low-density parity-check (LDPC) codes can be applied to construct non-malleable codes according to Theorem 13. In fact, an LDPC code is associated to a Tanner graph. The *Tanner graph* of an LDPC code with parity-check matrix $H = (h_{ij})$ is a bipartite graph $I = (V_1, V_2, E)$, where $V_1$ and $V_2$ are indexes of rows and columns of $H$ respectively, and two vertices $i \in V_1$ and $j \in V_2$ are adjacent if and only if $h_{ij} \neq 0$, see [39]. On the one hand, it is known ([27, 38, 44]) that if a Tanner graph is an expander biregular graph, the corresponding LDPC code has fast decoding algorithms. On the other hand, as shown in [16, 25, 30], the algebraic or combinatorial constructions of LDPC codes often provide Tanner graphs which are $\mu$-spectral expanders with optimally small $\mu$ with respect to Lemma 6. We remark that the generalized quadrangles in Example 19 also appear as Tanner graphs of specific LDPC codes, see [25, 30, 40]. Thus if the Tanner graphs of LDPC codes are spectral expanders, we can apply these graphs (accordingly LDPC codes) to construct non-malleable codes by Theorem 13.

Finally in terms of practical applications, it is desirable to construct split-state non-malleable codes for $k$-bit messages with $k \geq 1$. As far as we know, there is no known graph-theoretic constructions of non-malleable codes for the split-state model and $k$-bit messages. It would be of interest to generalize the codes in this paper and [35] for $k$-bit messages in the split state model.

## Acknowledgement

S. Satake has been supported by Grant-in-Aid for JSPS Fellows 20J00469 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, 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.

[7] S. Agrawal, D. Gupta, H. K. Maji, O. Pandey and M. Prabhakaran, "A rate-optimizing compiler for non-malleable codes against bit-wise tampering and permutations," in *Twelfth IACR Theory of Cryptography Conference (TCC 2015)*, pp. 375–397, 2015.

[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] C. Ballantine and D. Ciubotaru, "Ramanujan bigraphs associated with $SU(3)$ over a $p$-adic field," *Proc. Amer. Math. Soc.*, vol. 139, no. 6, pp. 1939–1953, 2011.

[10] C. Ballantine, B. Feigon, R. Ganapathy, J. Kool, K. Maurischat and A. Wooding, "Explicit construction of Ramanujan bigraphs," in *Women in numbers Europe*, Assoc. Women Math. Ser., 2, Springer, Cham, pp. 1–16, 2015.

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

[12] 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.

[13] 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.

[14] 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.

[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, Jun. 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, 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] S. Hoory, N. Linial and A. Wigderson, "Expander graphs and their applications," *Bull. Amer. Math. Soc. (N.S.)*, vol. 43, no. 4, pp. 439–561, 2006.

[27] H. Janwa and A. K. Lal, "On Tanner codes: minimum distance and decoding," *Appl. Algebra Engrg. Comm. Comput.*, vol. 13, no. 5, pp. 335–347, 2003.

[28] 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.

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

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

[31] A. Lubotzky, Discrete Groups, Expanding Graphs and Invariant Measures, Birkhäuser Verlag, Basel, 2010.

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

[33] 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.

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

[35] 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.

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

[37] R. Shaltiel, "Recent developments in explicit constructions of extractors," *Bull. Eur. Assoc. Theor. Comput. Sci. EATCS*, no. 77, pp. 67–95, 2002.

[38] M. Sipser and D. A. Spielman, "Expander codes," *IEEE Trans. Inform. Theory*, vol. 42, no. 6, pp. 1710–1722, Nov. 1996.

[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] 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.

[43] A. Wigderson and D. Zuckerman, "Expanders that beat the eigenvalue bound: Explicit construction and applications," *Combinatorica*, vol. 19, no. 1, pp. 125–138, 1999.

[44] Z. Zémor, "On expander codes," *IEEE Trans. Inform. Theory*, vol. 47, no. 2, pp. 835–837, Feb. 2001.

[45] Y. Zhu, "On the second eigenvalue of random bipartite biregular graphs," *arXiv:2005.08103*, 2020.

# A Proof of Lemma 20

This section proves Lemma 20, which is employed in the proof of Theorem 13 (see Case 3-1 in Section 6). The proof here is an analogue of the discussion on [35, Theorem 10].

To bound $\sum_{1\le k\le l\le \lceil \log_2(rs)\rceil} S_{k,l}$, let

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

By the definitions of $L$ and $K$,

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

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

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

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.2). On the other hand, the definition of $H^1(l)$ implies that

$$|h^{-1}(H^1(l))| \ge \frac{|V_2||H^1(l)|}{2^l}. \tag{A.3}$$

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

$$L \le O\Big(\sqrt{rs\log(rs)}\Big) \cdot \sqrt{\sum_{1\le l\le \lceil \log_2(rs)\rceil} \frac{|h^{-1}(H^1(l))|}{|V_2|}} = O\Big(\sqrt{rs\log(rs)}\Big), \tag{A.4}$$

where the last equation follows from (A.3).

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

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

where the first inequality follows from (A.3) and the following inequality.

$$|g^{-1}(G^1(k))| \ge \frac{|V_1||G^1(k)|}{2^k}. \tag{A.5}$$

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

$$\begin{aligned} K &\le \sum_{t=0}^{\lceil \log_2(rs)\rceil} \frac{2^{-\frac{t}{4}}}{(|V_1|^3|V_2|)^{\frac{1}{4}}} \left(\sum_{l=t}^{\lceil \log_2(rs)\rceil} \Big(|g^{-1}(G^1(l-t))|\Big)^{\frac{3}{4}} \cdot \left(\sum_{l=t}^{\lceil \log_2(rs)\rceil} |h^{-1}(H^1(l))|\right)^{\frac{1}{4}}\right. \\ &\le \sum_{t=0}^{\lceil \log_2(rs)\rceil} 2^{-\frac{t}{4}} = O(1). \end{aligned} \tag{A.6}$$

By (A.1), (A.4) and (A.6), we get

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

where the last equality follows from the condition in Theorem 13 that $|E| = \Omega\Big(\frac{(rs)^2\log(rs)}{\mu}\Big)$. This completes the proof of Lemma 20.

We remark that one can also prove (6.3) for Case 3-2 in Section 6 by conducting a similar discussion as the above proof for Lemma 20.

# B The graph code by Rasmussen and Sahai

In this section, we briefly review the graph code proposed in [35]. To define it more rigorously, we associate the underlying graph with a *digraph* as described below.

**Definition 22** ([35])**.** Let $G = (V, E)$ be an undirected graph with no multiple edges (but each vertex may have at most one loop). Let $D_G = (V, E')$ be the associated symmetric digraph with vertex set $V$ and edge set $E' \subset V \times V$ such that

$$E' = \{(u, w), (w, u) \in V \times V : \{u, w\} \in E\}.$$

Then the associated graph code $(\mathrm{enc}'_G, \mathrm{dec}'_G)$ consists of the functions

$$\mathrm{enc}'_G : \{0, 1\} \to V \times V, \qquad \mathrm{dec}'_G : V \times V \to \{0, 1\}$$

such that

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

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

**Theorem 23** ([35])**.** *Let $G = (V, E)$ be a $d$-regular graph with $n$ vertices and $\lambda(G) = \lambda$. Assume that $n = \Omega(\frac{d^3 \log(d)}{\lambda})$. Let $\mathcal{F}$ be the set of all functions $f = (g, h)$ with $g : V \to V$ and $h : V \to V$. Then $(\mathrm{enc}'_G, \mathrm{dec}'_G)$ is an $O\left(\frac{\lambda^{\frac{3}{2}}}{d}\right)$-non-malleable code with respect to $\mathcal{F}$.*