Electronic Colloquium on Computational Complexity, Report No. 4 (2022)

Analyzing Ta-Shma's Code via the Expander Mixing Lemma

Silas Richelson* Sourya Roy†

Abstract

Random walks in expander graphs and their various derandomizations (*e.g.*, replacement/zig-zag product) are invaluable tools from pseudorandomness. Recently, Ta-Shma used s-wide replacement walks in his breakthrough construction of a binary linear code almost matching the Gilbert-Varshamov bound (STOC 2017). Ta-Shma's original analysis was entirely linear algebraic, and subsequent developments have inherited this viewpoint. In this work, we rederive Ta-Shma's analysis from a combinatorial point of view using repeated application of the *expander mixing lemma*. We hope that this alternate perspective will yield a better understanding of Ta-Shma's construction. As an additional application of our techniques, we give an alternate proof of the *expander hitting set lemma*.

1 Introduction

Error correcting codes (ECCs) allow a sender to encode a message so that the receiver can recover the full message even if several codeword bits are lost or flipped during transmission. ECCs are incredibly useful, both in theory and in practice [Sha79, STV01, CJW19] (and many, many more). Formally, a binary code is a map $\mathcal{C} : \{0,1\}^k \to \{0,1\}^n$ which sends a message $m \in \{0,1\}^k$ to the codeword $\mathcal{C}(m) \in \{0,1\}^n$. Two important parameters of a code are the *distance* and *rate*, which are respectively measures of the code's quality and efficiency. *Rate* is the ratio k/n, the number of message bits per codeword bit while *distance* refers to the minimum fraction of coordinates (in $[n]$) on which two distinct codewords disagree. One of the holy grails in coding theory is to find the best tradeoff between the distance and rate of a binary code. It is known that codes with optimal distance $\delta = 1/2$ must have exponentially small rate [Plo60]. The Gilbert-Varshamov (GV) bound [Gil52, Var57] states for any $\delta \in (0, 1/2)$, there exists a code C_n with blocklength n and distance d with rate $1 - H(\delta) - o_n(1)$ where $H(\cdot)$ is Shannon's binary entropy function. Unfortunately, this is a probabilistic (or greedy) construction and we do not know of explicit binary codes matching this bound. For distances δ close to $1/2$, the GV bound states that there exists a code with distance $(1-\varepsilon)/2$ and rate $\Omega(\varepsilon^2)$. On the other hand, it is known that any code with distance $(1-\varepsilon)/2$ must have rate $\mathcal{O}\big(\varepsilon^2 \cdot \log(1/\varepsilon)\big)$ [ABN+92]. Constructing an explict code matching the GV bound even for these distance parameters is a major open problem.

A few years ago, in a breakthrough result, Ta-Shma [TS17] described an explicit construction which got very close: he constructed a family of codes $\{C_n\}_n$ with rate $\Omega(\varepsilon^{2+o_\varepsilon(1)})$ and distance $(1-\varepsilon)/2$. The core of his construction is an amplification procedure which increases the distance of the code using

*UC Riverside. Email: silas@cs.ucr.edu.
†UC Riverside. Email: sourya.roy@email.ucr.edu.

ISSN 1433-8092

certain special types of random walks on expander graphs. Specifically, Ta-Shma encodes a message $m \in \{0,1\}^k$ as follows.

1. Use a "base code" $\mathcal{C}_0 : \{0,1\}^k \to \{0,1\}^n$ with a good (but not optimal) rate/distance tradeoff, to encode message $m \in \{0,1\}^k$ into a n-bit codeword $\mathcal{C}_0(m)$ which we will equivalently interpret as function $f : [n] \to \{0,1\}$.

2. Identify the coordinate set $[n]$ with the vertices of an expander graph A.[1]

3. Let $W \subset A^t = [n]^t$ be a *special* subset of the set of all t-length walks in A. Define $g : W \to \{0,1\}$ by $g(a_1, \ldots, a_t) = f(a_1) \oplus \cdots \oplus f(a_t)$, where $\oplus$ is the bit XOR. Output $g \in \{0,1\}^{|W|}$.

The ingenious component in TaShma's construction is the choice of the subset W. As we will soon see, choosing W to be the set of all t-length walks in A does not yield an optimal distance/rate tradeoff. TaShma, instead, uses a derandomized subset of walks, resulting from taking an *s-wide replacement product walk* on A. In the ordinary replacement product, another expander B is chosen with $|B| = \deg(A)$ so that given $a \in A$, each $b \in B$ corresponds to some $a' \in N(a)$. A t-length replacement product walk in A chooses a random $a \sim A$ and a $(t-1)$-length walk $(b_1, \ldots, b_{t-1})$ in B and outputs the walk $(a_1, \ldots, a_t)$ in A where $a_1 = a$ and a_{i+1} is the b_i-th neighbor of a_i for $i = 1, \ldots, t-1$. Note the set of replacement product walks in A is a proper subset of the set of all walks. The s-wide replacement product is a parametrized version of the ordinary replacement product. We explain the s-wide replacement product in detail in Section 2.

1.1 Our Contribution

In this note, we rederive the analysis of TaShma [TS17] using repeated applications of the *Expander Mixing Lemma*. TaShma's original analysis, as well as subsequent developments, convey a strongly linear algebraic viewpoint. In this writeup, we take the expander mixing lemma as our starting point and proceed from there in a combinatorial fashion. Thus, we demonstrate that no linear algebra is needed for the analysis of Ta-Shma's code beyond that which is needed to prove the expander mixing lemma. We would like to be forthcoming and stress that **our analysis is completely equivalent to Ta-Shma's original analysis**. So if you are hoping to read about a new code with improved parameters, you should read something else. This paper is for those researchers who have had difficulty penetrating the intuition behind Ta-Shma's construction. We believe that this alternate perspective will appeal to a wider audience and make it easier for the scientific community to innovate on Ta-Shma's breakthrough work.

Our proof is the same as the original proof insofar as a random walk on a graph can be modelled both as a random process and as a linear operator. The original analysis takes the linear operator view, we take the random process view. In theory, the linear operator view is convenient for quantitatively reason about random walks because it reduces the task to understanding repeated multiplication by a fixed matrix. However, when analyzing replacement product walks from the linear operator perspective, the adjacency matrices of the outer and inner expander graphs have to be combined using some kind of tensor product. The situation is worse for the $s-$wide replacement product since then one has to keep track of s different tensor product matrices and the iterated matrix product needs to alternate over these s matrices. Thus, it seems there are diminishing returns in terms of the simplicity afforded by the linear operator perspective when the set of all random walks is to be derandomized. By using the random

[1] We abuse notation by refering to A both as the graph and the vertex set.

process view, we are able to express the same ideas in a much simpler way. This, in turn, makes it easier to see what is going on in certain key steps of the argument.

1.2 Techniques: Expander Mixing Lemma and consequences

Notation. Throughout this paper, we refer to graphs by their vertex sets, and use $\sim$ to indicate that two vertices are connected with an edge. So for example, if A is a graph and $a, a' \in A$ are vertices, we write $a \sim a'$ if there is an edge between a and a'. We write RW_A^t (resp. $\mathsf{RW}_A^t(a)$) for the distribution which outputs a $t-$length random walk in A (resp. a $t-$length random walk in A which begins at a). Given two distributions $\mathcal{D}$ and $\mathcal{D}'$, we will write $\mathcal{D} \equiv \mathcal{D}'$ to denote that they are same.

In order to get a sense for our technique, let us analyze the distance amplification procedure resulting from taking a random walk on an expander. Typically expander graphs are defined via the second largest eigenvalue of the adjacency matrix of the graph; in this paper we will use the following equivalent definition (similar definitions have been used in other works, *e.g.*, [DK17]).

Definition 1. *We say that a graph A is a $\lambda-$*expander *if for all $f, g : A \to \mathbb{R}$, the following holds:*

$$\left|\mathbb{E}_{a\sim a'}\big[f(a)\cdot g(a')\big] - \mu_f\mu_g\right| \le \lambda\sigma_f\sigma_g,$$

where μ_f and σ_f are the expectation and standard deviation of the random variable $f(a)$ (namely, $\mu_f = \mathbb{E}_a\big[f(a)\big]$ and $\sigma_f^2 + \mu_f^2 = \mathbb{E}_a\big[f(a)^2\big]$, and similarly for μ_g and σ_g).

Now consider the distance amplification framework above instantiated with A being a constant degree, $d-$regular $\lambda-$expander, and W being the set of all $t-$length random walks in A. Note that $|W| = n \cdot d^{t-1}$, and so the rate of the resulting code is $\mathcal{O}(d^{-t})$. If A is Ramanujan (*i.e.*, an expander with the best possible relationship between λ and d) then $\lambda \approx 2/\sqrt{d}$ which makes the rate $\mathcal{O}\big((\lambda/2)^{2t}\big)$. Regarding the distance, note that for any $n-$bit string $f : [n] \to \{0,1\}$, if the fraction of non-zero coordinates is $\frac{1-\varepsilon}{2}$, then $\varepsilon = -\mathbb{E}_{v\sim[n]}\big[(-1)^{f(v)}\big]$. For this reason, we show that the amplification framework above decreases *bias*, where

$$\mathrm{Bias}(f) := \left|\mathbb{E}_{v\sim[n]}\big[(-1)^{f(v)}\big]\right|.$$

The claim below shows that when W is the set of all $t-$length walks in A, a regular Ramanujan expander graph with expansion λ, and when $\mathrm{Bias}(f) \le \sqrt{\lambda}$, then $\mathrm{Bias}(g) \le \frac{1}{2}\cdot(4\lambda)^{t/2}$. It follows that if the distance of the amplified code is $\frac{1-\varepsilon}{2}$, then the rate is $\Omega\big(\varepsilon^4 \cdot 8^{-2t}\big)$. For any constant $\alpha > 0$, it is possible to choose parameters so that $\varepsilon^\alpha \le 8^{-2t}$, in which case the rate is $\Omega(\varepsilon^{4+\alpha})$.

Claim 1. *Let A be a regular $\lambda-$expander, $f : A \to \{0,1\}$ a function of bias $\left|\mathbb{E}_a\big[(-1)^{f(a)}\big]\right| \le \sqrt{\lambda}$. For $k \ge 1$, define $h_k : A \to \mathbb{R}$ as*

$$h_k(a) := \mathbb{E}_{(a_1,\ldots,a_k)\sim\mathsf{RW}_A^k(a)}\Big[(-1)^{f(a_1)\oplus\cdots\oplus f(a_k)}\Big].$$

Let $\varepsilon_k := \left|\mathbb{E}_a\big[h_k(a)\big]\right|$ and σ_k be such that $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\big[h_k(a)^2\big]$. Then for all $k \ge 1$:

$$\varepsilon_k \le \frac{1}{2}\cdot(4\lambda)^{k/2};\ \sigma_k \le \sqrt{\mathbb{E}_a\big[h_k(a)^2\big]} \le (4\lambda)^{\frac{k-1}{2}}.$$

We will actually prove the following slight generalization of Claim 1, which will be more useful in our analysis later on. Note Claim 1 is recovered from Claim 2 by letting H be the constant function which always outputs 1, and noting that $\hat{\varepsilon}_1 \le \sqrt{\lambda}$ and $\hat{\sigma}_1 \le 1$.

Claim 2. *Let A be a regular $\lambda-$expander, $f : A \to \{0,1\}$ a function of bias $\left|\mathbb{E}_a\left[(-1)^{f(a)}\right]\right| \leq \sqrt{\lambda}$, and $H : A \to \mathbb{R}$ any function. For $k \geq 1$, let $\hat{h}_k : A \to [0,1]$ be defined by*

$$\hat{h}_k(a) = \mathbb{E}_{(a_1,\ldots,a_k)\sim \mathsf{RW}^k(a)}\left[(-1)^{f(a_1)\oplus\cdots\oplus f(a_k)} \cdot H(a_k)\right].$$

Let $\hat{\varepsilon}_k := \left|\mathbb{E}_a\left[\hat{h}_k(a)\right]\right|$ and $\hat{\sigma}_k$ such that $\hat{\sigma}_k^2 + \hat{\varepsilon}_k^2 = \mathbb{E}_a\left[\hat{h}_k(a)^2\right]$. Then for $k \geq 2$,

$$\hat{\varepsilon}_k \leq 2^{k-2} \cdot (\lambda^{\frac{k-1}{2}}\hat{\varepsilon}_1 + \lambda^{\frac{k}{2}}\hat{\sigma}_1); \text{ and } \hat{\sigma}_k \leq \sqrt{\mathbb{E}_a\left[\hat{h}_k(a)^2\right]} \leq 2^{k-2} \cdot (\lambda^{\frac{k-2}{2}}\hat{\varepsilon}_1 + \lambda^{\frac{k-1}{2}}\hat{\sigma}_1).$$

Proof. The key observation is that for $k \geq 2$, $\hat{h}_k(a) = (-1)^{f(a)} \cdot \mathbb{E}_{a'\sim N(a)}\left[\hat{h}_{k-1}(a')\right]$. This lets us bound $\hat{\varepsilon}_k$ and $\hat{\sigma}_k$ in terms of $\hat{\varepsilon}_{k-1}$ and $\hat{\sigma}_{k-1}$ using the expander mixing lemma (Definition 1) as follows:

$$\cdot\ \hat{\varepsilon}_k = \left|\mathbb{E}_a\left[\hat{h}_k(a)\right]\right| = \left|\mathbb{E}_{a\sim a'}\left[(-1)^{f(a)} \cdot \hat{h}_{k-1}(a')\right]\right| \leq \sqrt{\lambda}\hat{\varepsilon}_{k-1} + \lambda\hat{\sigma}_{k-1};$$

$$\cdot\ \hat{\sigma}_k^2 \leq \hat{\sigma}_k^2 + \hat{\varepsilon}_k^2 = \mathbb{E}_a\left[\hat{h}_k(a)^2\right] = \mathbb{E}_a\left[\mathbb{E}_{a'\sim N(a)}\left[\hat{h}_{k-1}(a')\right]^2\right] = \mathbb{E}_{a'\sim_{A^2} a''}\left[\hat{h}_{k-1}(a') \cdot \hat{h}_{k-1}(a'')\right]$$

$$\leq \hat{\varepsilon}_{k-1}^2 + \lambda^2\hat{\sigma}_{k-1}^2,$$

where $a' \sim_{A^2} a''$ indicates that (a', a'') is a uniform edge in A^2 (a λ^2-expander). We have used that the distribution which draws $a \sim A$, $a', a'' \sim N(a)$ and outputs (a', a'') is identical to the uniform edge distribution on A^2. The claim follows by induction. □

1.3 Improving the rate via s-wide replacement product walks

The rate of the above code is roughly ε^4, which is too low. In order for it to have rate $\approx \varepsilon^2$, we would have needed $\varepsilon_t \leq \lambda^t$ rather than what we got which was $\varepsilon_t \leq \lambda^{t/2}$ (actually we got something weaker, we are oversimplifying to clarify the discussion). The recursive formulas which appeared in the proof were:

- $\varepsilon_k \leq \mathrm{Bias}(f) \cdot \varepsilon_{k-1} + \lambda\sigma_{k-1} \leq \sqrt{\lambda}\varepsilon_{k-1} + \lambda\sigma_{k-1}$ (we assumed $\mathrm{Bias}(f) \leq \sqrt{\lambda}$);
- $\sigma_k \leq \varepsilon_{k-1} + \lambda\sigma_{k-1}$ (implied by $\sigma_k^2 \leq \varepsilon_{k-1}^2 + \lambda^2\sigma_{k-1}^2$).

The problem here is the bound $\sigma_k \leq \varepsilon_{k-1} + \lambda\sigma_{k-1}$, specifically the ε_{k-1} term on the right since we are moving from a $k-$th level term to a $(k-1)-$th level term without gaining a factor of λ. Plugging this into the first equation gives $\varepsilon_k \leq \sqrt{\lambda}\varepsilon_{k-1} + \lambda\varepsilon_{k-2} + \lambda^2\sigma_{k-2}$, where the first two terms are problematic (we are moving from level k to level $k-1$ and $k-2$ but gaining only one factor of $\sqrt{\lambda}$ and λ, respectively). The first problematic term could be fixed by choosing λ such that $\mathrm{Bias}(f) \leq \lambda$; but the second problematic term cannot be easily fixed. This phenomenon was observed in [TS17] where the problem is summarized by saying "one out of every two steps works".

A natural idea for derandomizing W is to work with a set of replacement (or zig-zag) product walks. Unfortunately this yields no improvement as the "one out of every two steps works" problem persists. Ben-Aroya and Ta-Shma [BATS11] solved this problem in a different context by using an expander graph B on a slightly larger vertex set of size d^s for $s \geq 2$, and by analyzing the resulting walk s steps at a time. This is called the s-wide replacement product. Ta-Shma was then able to successfully argue that "$s-4$ out of every s steps work". When interpreted in our language, this observation translates to a recursive formula like $\varepsilon_k \leq \lambda^{s-4} \cdot \varepsilon_{k-s}$, where we move from a $k-$th level term to a $(k-s)-$th level term, while gaining $(s-4)$ factors of λ. Gaining s factors of λ would have let us solve to the optimal $\varepsilon_k \leq \lambda^k$, obtaining rate of $\approx \varepsilon^2$; gaining $(s-4)$ factors of λ lets us solve instead to $\varepsilon_k \leq \lambda^{k(1-4/s)}$ which is almost as good when s is large.

2 Preliminaries

Random Walks on Graphs. Let A be the vertex set of a graph. Given $a, a' \in A$, we write $a \sim a'$ if a and a' are connected by an edge. For $a \in A$, let $N(a) \subset A$ denote the *neighborhood* of A, *i.e.*, $N(a) := \{a' \in A : a \sim a'\}$. For an integer $d \geq 1$, we say that A is $d-$regular if $|N(a)| = d$ for all $a \in A$. For an integer $k \geq 1$, let

$$\mathsf{RW}_A^k := \{(a_1, \ldots, a_k) \in A^k : a_i \sim a_{i+1} \; \forall \, i = 1, \ldots, k-1\}$$

denote the set of $k-$length random walks in A. Similarly, for $a \in A$, $\mathsf{RW}_A^k(a)$ is the set of $k-$length random walks in A which begin at a, so $\mathsf{RW}_A^k(a) := \{(a_1, \ldots, a_k) \in \mathsf{RW}_A^k : a_1 = a\}$. We will often view RW_A^k as a distribution, where $(a_1, \ldots, a_k) \sim \mathsf{RW}_A^k$ means that $a_1 \sim A$ is drawn uniformly and then $a_{i+1} \sim N(a_i)$ is drawn for $i = 1, \ldots, k-1$.

Expander Graphs. Graph expansion is usually defined as the second largest eigenvalue of the graph's adjacency matrix,[2] *i.e.*,

$$\lambda := \max_{x, y \perp \mathbb{1}} \frac{|\langle x, My \rangle|}{|x||y|}, \tag{1}$$

where the max is over all nonzero $x, y \in \mathbb{R}^{|A|} - \{0\}$ which are perpendicular to the all 1s vector $\mathbb{1}$. Our Definition 1 can be recovered from (1) for any $f, g : A \to \mathbb{R}$ by setting $x, y \in \mathbb{R}^{|A|}$ to be $x_a = f(a) - \mu_f$ and $y_a = g(a) - \mu_g$.

Cayley Graphs. Given a finite group G and a subset $U \subseteq G$, the Cayley graph $\mathrm{Cayley}(G, U)$ has vertex set G with $g \sim g'$ iff $g^{-1}g' \in U$. Note that $\mathrm{Cayley}(G, U)$ is $|U|-$regular; additionally, if U is closed under inversion, then $\mathrm{Cayley}(G, U)$ is undirected. Cayley graphs play a key role in many explicit constructions of expander graphs. Ta-Shma's original construction used two Cayley graphs as explicit expander constructions. The first Cayley graph was over $\mathbb{F}_2^k$, and the second was over $\mathsf{PGL}_2(\mathbb{F}_q)$, the projective general linear group over a large finite field. The use of this second Cayley graph put restrictions on some of the parameters, which required some care in order to navigate. Subsequently to Ta-Shma's original paper, new constructions of expanders based on Cayley graphs have been given. We will use a new construction, due to Alon [Alo21], instead of the $\mathsf{PGL}_2(\mathbb{F}_q)$ construction as it will give us more flexibility.

Theorem 1. *We have the following expander constructions from [Alo21] and [AGHP92], respectively.*

- **The Outer Graph:** *For all integers $n, d \in \mathbb{N}$ there is an explicit construction of a $d-$regular Cayley graph with $n \cdot (1 + o_n(1))$ vertices and expansion $\lambda \leq \frac{8}{\sqrt{d}}$.*
- **The Inner Graph:** *For all integers $r, \ell \in \mathbb{N}$ such that $\ell \leq r/2$, there exists an explicit[3] construction of an undirected $2^{2\ell}-$regular Cayley graph over $\mathbb{F}_2^r$ which is a $(r-1)2^{-\ell}-$expander.*

[2] The *adjacency matrix* of the graph A is $M \in \{0,1\}^{|A| \times |A|}$, where $M(a, a') = 1$ iff $a \sim a'$.

[3] This Cayley graph construction is actually *fully explicit*, in the sense that given any vertex, the $i-$th neighbor can be computed in polylogarithmic time.

The Shifted Neighborhood Distribution. Let B be a Cayley graph on $\mathbb{F}_2^{ms}$, and let $d = 2^m$. For any $b = \big(b[1], \ldots, b[s]\big) \in B \cong [d]^s$, let $\mathsf{shift}(b) = \big(b[2], \ldots, b[s], b[1]\big) \in B$ be the element obtained by circularly shifting the coordinates of b. Given $b \in B$, the *shifted neighborhood distribution of* b, denoted $\tilde{N}(b)$, draws $u \sim U$ (the generator set of the Cayley graph) and outputs $\mathsf{shift}(b+u)$ (note $b+u$ is a random neighbor of b in B). It is clear that the expansion of B is not affected by using the shifted neighborhood distribution instead of the original neighborhood distribution. Indeed,

$$\left|\mathbb{E}_{\substack{b\sim B\\ b'\sim\tilde{N}(b)}}\big[f(b)\cdot g(b')\big] - \mu_f\mu_g\right| = \left|\mathbb{E}_{\substack{b\sim B\\ b'\sim N(b)}}\big[f(b)\cdot \tilde{g}(b')\big] - \mu_f\mu_{\tilde{g}}\right| \le \lambda\sigma_f\sigma_{\tilde{g}} = \lambda\sigma_f\sigma_g,$$

where $\tilde{g} = g \circ \mathsf{shift}$; clearly $(\mu_{\tilde{g}}, \sigma_{\tilde{g}}) = (\mu_g, \sigma_g)$. Let $\tilde{\mathsf{RW}}_B^k$ denote the set of $k-$length shifted random walks in B. We prove the following claim about $\tilde{\mathsf{RW}}_B^k$, when k is small.

Claim 3. *For all $k \le s$, the distribution that chooses $(b_1, \ldots, b_k) \sim \tilde{\mathsf{RW}}_B^k$ and outputs the tuple $(b_1[1], b_2[1], \ldots, b_k[1]) \in [d]^k$ is identical to the uniform distribution on $[d]^k$.*

Proof. It suffices to prove the claim for $k = s$, since when $k < s$, the distribution $\tilde{\mathsf{RW}}_B^k$ is identical to the distribution which draws $(b_1, \ldots, b_s) \sim \tilde{\mathsf{RW}}_B^s$ and outputs $(b_1, \ldots, b_k)$. Note that $\tilde{\mathsf{RW}}_B^s$ draws $u_1, \ldots, u_{s-1} \sim U$, $b_1 \sim B$ and outputs $(b_1, \ldots, b_s) \in B^s$, where $b_i = \mathsf{shift}(b_{i-1}+u_{i-1})$ for $i = 2, \ldots, s$. This means that for all $i = 1, \ldots, s$, $b_i[1] = b_1[i] + \sum_{j<i} u_j[i-j+1]$ (addition over $\mathbb{F}_2^m$). Uniformity of $\big(b_1[1], b_2[1], \ldots, b_t[1]\big)$ follows from the uniformity of $b_1 = \big(b_1[1], \ldots, b_1[s]\big) \sim [d]^s$. □

2.1 The s-wide Replacement Product

Let A and B denote, respectively, the outer and inner graphs promised by Theorem 1. So A is a $d-$regular graph on (roughly) n vertices, while B is a Cayley graph over $\mathbb{F}_2^{ms}$, where $2^m = d$, so that vertices of B are identified with $s-$tuples of elements in $[d]$: $b = \big(b[1], \ldots, b[s]\big) \in [d]^s$. Given $a \in A$, a vertex $b \in B$ can be identified with an $s-$tuple of neighbors of a since $|N(a)| = d$. Define the *rotation map* $\phi : A \times B \to A$ via $\phi(a, b) = a'$ where a' is the $b[1]-$th neighbor of a. Since ϕ only depends on the first coordinate of b, we write $\phi(a, \hat{b})$ where $\hat{b}$ is shorthand for $b[1]$. For any $k \ge 1$, the $k-$length $s-$wide replacement walk distribution, denoted $s\mathsf{RW}_{A,B}^k$ draws $a \sim A$ and $(b_1, \ldots, b_{k-1}) \sim \tilde{\mathsf{RW}}_B^{k-1}$, and outputs $(a_1, \ldots, a_k) \in A^k$ where $a_1 = a$ and $a_{i+1} = \phi(a_i, \hat{b}_i)$ for $i = 1, \ldots, k-1$. Since the graphs A and B will be fixed throughout this paper, we write $s\mathsf{RW}^k$ rather than $s\mathsf{RW}_{A,B}^k$. Given $a \in A$, the distribution $s\mathsf{RW}^k(a)$ outputs a sample from $s\mathsf{RW}^k$ conditioned on $a_1 = a$. Likewise, given $(a, b) \in A \times B$, $s\mathsf{RW}^k(a, b)$ outputs a sample from $s\mathsf{RW}^k$ conditioned on $(a_1, b_1) = (a, b)$. The $s-$wide replacement walk is shown in Figure 1.

For our graphs A and B (specifically, since A is $d-$regular and B is a Cayley graph over $\mathbb{F}_2^{ms} \cong [d]^s$) the next fact follows immediately from Claim 3.

Fact 1 (Pseudorandomness). *For all $k = 1, 2, \ldots, s, s+1$ and all $a \in A$, $s\mathsf{RW}^k(a) \equiv \mathsf{RW}_A^k(a)$.*

Following Ta-Shma's nomenclature, we will refer to the fact above as the *pseudorandomness* property. This property will play a crucial role in our proofs below as it will allow us to transform a short $s-$wide walk into a pure random walk on A, thus eliminating the dependency on the graph B.

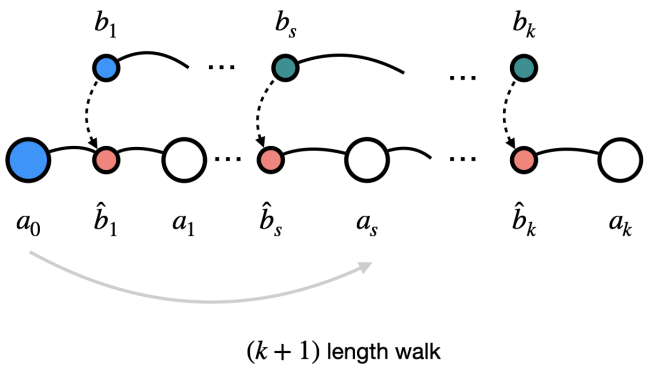

Figure 1: Illustration of s-wide random walk on A using a graph B.

Local Invertibility. Since A is undirected, its edge relation is symmetric. This means that whenever $a, a' \in A$ and $b \in B$ are such that $a' = \phi(a, \hat{b})$, there must exist some $\hat{b}' \in [d]$ such that $a = \phi(a', \hat{b}')$. In this case we say that $(\hat{b}, \hat{b}')$ are inverses with respect to the $A-$edge (a, a'). Local invertibility in our context means that these inverse relations are independent of the A edges. So, specifically, for all $\hat{b}$ there exists $\hat{b}'$ such that $(\hat{b}, \hat{b}')$ are inverses with respect to all A edges. This means, for example that for all $a \in A$, if you walk to $a' = \phi(a, \hat{b})$ and then continue to $a'' = \phi(a', \hat{b}')$, then $a'' = a$. This property is easy to establish in our situation because A is a Cayley graph.

Practically speaking, what this means for us is that $s-$wide replacement walks can be "started in the middle". For standard random walks, the distribution RW_A^k which outputs $(a_1, \ldots, a_k)$ is identical to the distribution which first chooses $a_i \sim A$ randomly, and then draws $(a_i, a_{i+1}, \ldots, a_k) \sim \mathsf{RW}_A^{k-i+1}(a_i)$ and $(a_i, a_{i-1}, \ldots, a_1) \sim \mathsf{RW}_A^{i}(a_i)$, outputting $(a_1, \ldots, a_k)$. This follows from the regularity of A. Likewise, because of local invertibility, the $s-$wide replacement walk distribution $s\mathsf{RW}^k$ is identical to the following "start in the middle" version which draws $a_i \sim A$ and $b_i \sim B$, then draws $(b_i, \ldots, b_{k-1}) \sim \tilde{\mathsf{RW}}_B^{k-i}(b_i)$ and $(b_i, \ldots, b_1) \sim \tilde{\mathsf{RW}}_B^{i}(b_i)$ (in this case the shifted neighborhood distribution needs to shift the other way), then sets $a_{j+1} = \phi(a_j, \hat{b}_j)$ for $j = i, \ldots, k-1$ and $a_{j-1} = \phi(a_j, \hat{b}'_j)$ for $j = i, \ldots, 2$, where $\hat{b}'_j$ is the inverse of $\hat{b}_j$; finally $(a_1, \ldots, a_k)$ is output.

3 Main theorem

Theorem 2. *For every $\varepsilon > 0$ there exists an explicit linear code $\{\mathcal{C}_k\}_k$ that has distance $\geq \frac{1}{2} - \varepsilon$ and rate $= \Omega(\varepsilon^{2+o(1)})$.*

Proof. Fix $k \in \mathbb{N}$. The construction of $\mathcal{C}_k$ uses the following building blocks.

- **The Base Code:** Let $\mathcal{C}_0 : \{0,1\}^k \to \{0,1\}^{n_0}$ be an explicit code of bias ε_0 and rate R_0. We use the construction in [ABN+92], so that $R_0 = \mathcal{O}(\varepsilon_0^{-3})$.
- **The Outer Graph:** Let A be the d_A-regular Cayley graph with expansion λ_A. We use the construction of Theorem 1, so that $\lambda_A \leq 8/\sqrt{d_A}$ and $|A| = n_0 \cdot \big(1 + o_{n_0}(1)\big)$.

- **The Inner Graph:** Let B be a d_B-regular Cayley graph over $\mathbb{F}_2^r$ with expansion λ_B. We use the construction of Theorem 1 so that $\lambda_B = (r-1)\cdot 2^{-\ell}$ and $d_B = 2^{2\ell}$ for integers $\ell, r \in \mathbb{N}$ such that $\ell \le r/2$.

The building blocks carry several parameters which we now connect. In order to set up the $s-$wide replacement product, define additional parameters $s, m \in \mathbb{N}$ such that $r = ms$, and let $d_A = 2^m$, so $B \simeq [d_A]^s$. It will be important for our analysis to have $\lambda_A \le \lambda_B^2$; in order to arrange this, set $m = s$ and $\ell = s/5$. This gives

$$\lambda_A \le \frac{8}{\sqrt{d_A}} = 8\cdot 2^{-m/2} = \frac{8}{2^{\ell/2}}\cdot 2^{-2\ell} \le (ms-1)^2\cdot 2^{-2\ell} = \lambda_B^2,$$

where the final inequality holds whenever $s \ge 2$. We will also require $\varepsilon_0 \le \lambda_B/2$ which we ensure by setting $\varepsilon_0 = \frac{s^2-1}{2}\cdot 2^{-s/5}$. At this point, all parameters so far have been defined in terms of s; we will specify s later. Note that our setup allows us to use B to take $s-$wide replacement walks in A. We now describe the code. Given $x \in \{0,1\}^k$, $\mathcal{C}_k(x)$ is computed as follows.

- Compute $\mathcal{C}_0(x) \in \{0,1\}^{n_0}$, and define $f : A \to \{0,1\}$ by setting

$$f(a) = \begin{cases} \mathcal{C}_0(x)_i, & a = \iota(i) \\ 0, & \text{otherwise} \end{cases}$$

  where $\iota : [n_0] \hookrightarrow A$ is some fixed embedding.

- Define $g : s\mathsf{RW}^t \to \{0,1\}$ by setting $g(a_0,\dots,a_t) = f(a_0)\oplus\cdots\oplus f(a_t)$. Output $g \in \{0,1\}^{s\mathsf{RW}^t}$.

The rate of $\mathcal{C}_k$ is

$$\mathsf{Rate}_k = \frac{k}{|s\mathsf{RW}^t|} \ge \frac{k}{|A|}\cdot\frac{1}{|B|}\cdot\frac{1}{d_B^{t-1}} = \Omega(\varepsilon_0^{-3})\cdot 2^{-s^2}\cdot d_B^{-(t-1)} = \Omega\big(s^{-6}\cdot 2^{-s^2}\big)\cdot d_B^{-(t-1)}.$$

To bound the bias of $\mathcal{C}_k$, we use the following lemma which is proved in the next section.

Lemma 1 (Bias Reduction of Wide Replacement Product Walks). *Let integers $s, t \in \mathbb{N}$ and graphs A and B be as above; so in particular A and B are λ_A and λ_B expanders with $\lambda_A \le \lambda_B^2$. Let $f : A \to \{0,1\}$ be any function such that $\left|\mathbb{E}_a\big[(-1)^{f(a)}\big]\right| \le \lambda_B$. Then*

$$\left|\mathbb{E}_{(a_0,\dots,a_t)\sim s\mathsf{RW}^t}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_t)}\Big]\right| \le (2\lambda_B)^{t(1-4/s)}.$$

Note that the function $f : A \to \{0,1\}$ defined in the first step of computing $\mathcal{C}_k(x)$ satisfies

$$\left|\mathbb{E}_a\big[(-1)^{f(a)}\big]\right| \le 2\cdot\left|\mathbb{E}_{i\sim[n_0]}\big[(-1)^{\mathcal{C}_0(x)_i}\big]\right| \le 2\varepsilon_0 \le \lambda_B,$$

and so Lemma 1 ensures that $\mathsf{Bias}_k \le (2\lambda_B)^{t(1-4/s)}$. Putting the calculations of Rank_k and Bias_k together and using $\lambda_B = (s^2-1)/\sqrt{d_B}$ gives

$$\mathsf{Rate}_k = \Omega\Big(s^{-6}\cdot(s^2-1)^{-2t}\cdot 2^{-2t-s^2+2s/5}\cdot(2\lambda_B)^{8t/s}\Big)\cdot\mathsf{Bias}_k^2 = \Omega\Big(s^{-5t}\cdot(2\lambda_B)^{8t/s}\Big)\cdot\mathsf{Bias}_k^2,$$

where the right most equality holds whenever $6\log s \le 2s/5$ (implied by $s \ge 100$) and $t \ge s^2$. Note, therefore, that for $\eta \in \big(0, 1/2\big)$, $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}_k^{2+\eta}\big)$ holds whenever $(2\lambda_B)^{t(\eta-4\eta/s-8/s)} \le s^{-5t}$ which,

if $\eta \geq 24/s$ is implied by $(2\lambda_B)^{\eta/2} \leq s^{-5}$. Finally, by plugging in $\lambda_B = (s^2-1)\cdot 2^{-s/5}$, we see that this holds whenever $\eta s \geq 60 \log s$.

So finally, let us prove the theorem. Suppose that we are given $\varepsilon > 0$ and $\eta \in \big(0, 1/2\big)$, and we want to construct $\mathcal{C}_k$ such that $\mathsf{Bias}_k \leq \varepsilon$ and $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}^{2+\eta}\big)$. We let $\mathcal{C}_k$ be the construction defined above with s chosen large enough so that $\eta s \geq 60 \log s$; this ensures $\mathsf{Rate}_k = \Omega\big(\mathsf{Bias}_k^{2+\eta}\big)$ as noticed above. Finally, let us choose t large enough so that $t \geq s^2$ and $(2\lambda_B)^{t(1-4/s)} \leq \varepsilon$; this ensures $\mathsf{Bias}_k \leq \varepsilon$, as desired. □

4 Proof of Lemma 1

In this section we prove the key bias reduction lemma that was the core of Theorem 2. Our proof will be by induction, just like Claim 2, so we will need to modify the statement of Lemma 1 so it adheres to an inductive argument.

4.1 Lemma Statement

Let A and B be the graphs from Section 3. Write λ instead of λ_B for the expansion of B and recall that $\lambda_A \leq \lambda^2$. Let $f : A \to \{0,1\}$ be a function such that $\big|\mathbb{E}_a\big[(-1)^{f(a)}\big]\big| \leq \lambda$. For any $k \geq 0$, define $g_k : A \times B \to \mathbb{R}$ by

$$g_k(a,b) = \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k(a,b)}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\Big]. \tag{2}$$

Let $\varepsilon_k = \big|\mathbb{E}_{a,b}\big[g_k(a,b)\big]\big|$ and let σ_k be such that $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_{a,b}\big[g_k(a,b)^2\big]$. We prove the following.

Lemma 2 (Implies Lemma 1). *Assume the above setup. For all $k \geq 0$*

$$\varepsilon_k \leq (2\lambda)^{k(1-4/s)};\ \sigma_k \leq (2\lambda)^{(k-2)(1-4/s)}.$$

As mentioned, we prove Lemma 2 by induction. The following two claims combine to easily prove Lemma 2; we will prove them in Sections 4.3 and 4.4.

Claim 4 (Base Case.). *Assume the above setup. For all $k = 0, 1, \ldots, s$:*

$$\varepsilon_k \leq \frac{1}{2}\cdot(2\lambda)^{k+1};\ \sigma_k \leq 2\cdot(2\lambda)^{k-1}.$$

Claim 5 (Induction Step.). *Assume the above setup. For all $k > s$:*

- $\varepsilon_k \leq \frac{1}{2}(2\lambda)^s(\varepsilon_{k-s} + 3\sigma_{k-s})$;
- $\sigma_k^2 \leq \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2} + \lambda\sigma_{k-1})\big(\varepsilon_{k-s} + (2+\lambda)\sigma_{k-s}\big) + \lambda^s\sigma_{k-s}\sigma_{k-1} + \lambda^2\sigma_{k-1}^2$

Proof of Lemma 2. Claim 4 clearly establishes the base cases since $\frac{1}{2}\cdot(2\lambda)^{k+1} \leq (2\lambda)^{k(1-4/s)}$ and $2\cdot(2\lambda)^{k-1} \leq (2\lambda)^{(k-2)(1-4/s)}$. For the first part of the induction step, we have

$$\begin{aligned}\varepsilon_k &\leq \frac{1}{2}\cdot(2\lambda)^s\cdot(\varepsilon_{k-s}+3\sigma_{k-s}) \leq \frac{1}{2}\cdot(2\lambda)^s\cdot\Big[(2\lambda)^{(k-s)(1-4/s)} + 3\cdot(2\lambda)^{(k-s-2)(1-4/s)}\Big]\\ &= 8\lambda^4\cdot\Big[(2\lambda)^{k(1-4/s)} + 3\cdot(2\lambda)^{(k-2)(1-4/s)}\Big] \leq 2\lambda^2(4\lambda^2+3)\cdot(2\lambda)^{k(1-4/s)} \leq (2\lambda)^{k(1-4/s)}.\end{aligned}$$

The bound $2\lambda^2(4\lambda^2+3) \le 1$ holds because $\lambda \le 1/3$. The second part of the induction step is similar:

$$\begin{aligned}
\sigma_k^2 &\le \frac{1}{2}\cdot(2\lambda)^{s-2}\cdot(\varepsilon_{k-2}+\lambda\sigma_{k-1})\big(\varepsilon_{k-s}+(2+\lambda)\sigma_{k-s}\big)+\lambda^s\sigma_{k-s}\sigma_{k-1}+\lambda^2\sigma_{k-1}^2 \\
&\le \frac{1}{2}\cdot(2\lambda)^2\cdot\Big[(2\lambda)^{(k-2)(1-4/s)}+\lambda(2\lambda)^{(k-3)(1-4/s)}\Big]\cdot\Big[(2\lambda)^{k(1-4/s)}+(2+\lambda)(2\lambda)^{(k-2)(1-4/s)}\Big]+ \\
&+ \lambda^s(2\lambda)^{(k-s-2)(1-4/s)}(2\lambda)^{(k-3)(1-4/s)}+\lambda^2(2\lambda)^{2(k-3)(1-4/s)} \\
&= 2\lambda^2(2\lambda)^{(2k-2)(1-4/s)}+2\lambda^3(2\lambda)^{(2k-3)(1-4/s)}+(4\lambda^2+2\lambda^3)(2\lambda)^{(2k-4)(1-4/s)}+ \\
&+ (4\lambda^3+2\lambda^4)(2\lambda)^{(2k-5)(1-4/s)}+2^{4-s}\lambda^4(2\lambda)^{(2k-5)(1-4/s)}+\lambda^2(2\lambda)^{(2k-6)(1-4/s)} \\
&\le \left[2\lambda^2+2\lambda^3+(4\lambda^2+2\lambda^3)+(2\lambda^2+\lambda^3)+2^{3-s}\lambda^3+\frac{1}{4}\right]\cdot(2\lambda)^{(2k-4)(1-4/s)} \le (2\lambda)^{(2k-4)(1-4/s)},
\end{aligned}$$

where the last bound has used $8\lambda^2+6\lambda^3 \le 3/4$ which holds because $\lambda \le 1/4$. □

4.2 Key Intuition

In this section we zoom in on some of the key steps in the coming proofs in order to give extra explanations and intuitions.

$s-$wide Replacement Product Walks in A. Recall that a random $s-$wide replacement product walk in A (*i.e.*, a random sample from $s\mathsf{RW}^k$) is produced as follows:

1. choose base points $(a,b) \sim A \times B$;
2. generate $(b_1,\ldots,b_k) \in B^k$ as follows:
   - (i) set $b_1 = b$;
   - (ii) for $i \ge 2$, draw $b_i \sim N(b_{i-1})$ and set $b_i = \mathsf{shift}(b_i)$, where shift cycles the coordinates of an element of $B \simeq [d]^s$, so $\mathsf{shift}\big(b_i[1],\ldots,b_i[s]\big) = \big(b_i[2],\ldots,b_i[s],b_i[1]\big)$.
3. generate and output $(a_0,\ldots,a_k) \in A^{k+1}$ as follows:
   - (i) set $a_0 = a$;
   - (ii) for $i \ge 1$, set $a_i = \phi(a_{i-1},\hat{b}_i)$ where $\hat{b}_i = b_i[1] \in [d]$ denotes the first coordinate of $b_i \in [d]^s$, and where ϕ is the rotation map of A.

Pseudorandomness. As mentioned in Section 2, when $k \le s$ the distributions $s\mathsf{RW}^k$ and RW_A^{k+1} are identical. That is, a random $k-$step $s-$wide replacement product walk in A is just a random $(k+1)-$step random walk in A. The following is an example of how this concept manifests itself in the next section. Let $\varepsilon_k(a) = \mathbb{E}_b\big[g_k(a,b)\big]$.

$$\varepsilon_k(a) = \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k(a)}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\Big] = \mathbb{E}_{(a_0,\ldots,a_k)\sim \mathsf{RW}_A^{k+1}}\Big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_k)}\Big] = h_{k+1}(a),$$

whenever $k \le s$, where h_{k+1} is the function defined and analyzed in Claim 1.

The Ignore First Step Trick. This refers to a key step in the proof that for all $k \geq 1$,

$$\sigma_k^2 \leq \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \sigma_{k-1}^2. \tag{3}$$

This bound is useful as it reduces the task of bounding σ_k^2 to the task of bounding $\mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right]$, which will turn out to be much easier. The proof of (3) requires other ideas as well. Recall from the previous paragraph the definition of $\varepsilon_k(a)$; additionally let $\sigma_k(a)$ be such that $\sigma_k(a)^2 + \varepsilon_k(a)^2 = \mathbb{E}_b\left[g_k(a,b)^2\right]$.

$$\begin{aligned}\sigma_k^2 &\leq \sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_{a,b}\left[g_k(a,b)^2\right] = \mathbb{E}_{a,b}\left[\mathbb{E}_{b' \sim N(b)}\left[g_{k-1}(a',b')\right]^2\right] = \mathbb{E}_{\substack{a \sim A \\ b \sim_{B^2} b'}}\left[g_{k-1}(a,b) \cdot g_{k-1}(a,b')\right] \\ &\leq \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \mathbb{E}_a\left[\sigma_{k-1}(a)^2\right] \leq \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right] + \lambda^2 \sigma_{k-1}^2.\end{aligned}$$

The second equation on the first line holds because $g_k(a,b) = (-1)^{f(a)} \cdot \mathbb{E}_{b' \sim N(b)}\left[g_{k-1}(a',b')\right]$, where $a' = \phi(a, \hat{b})$; the first inequality on the second line follows from the expander mixing lemma (Definition 1) on B^2 (a λ^2-expander); the final inequality has used $\mathbb{E}_a\left[\sigma_{k-1}(a)^2\right] \leq \sigma_{k-1}^2$ which holds because

$$\mathbb{E}_a\left[\sigma_{k-1}(a)^2 + \varepsilon_{k-1}(a)^2\right] = \mathbb{E}_{a,b}\left[g_{k-1}(a,b)^2\right] = \sigma_{k-1}^2 + \varepsilon_{k-1}^2,$$

and $\varepsilon_{k-1}^2 \leq \mathbb{E}_a\left[\varepsilon_{k-1}(a)^2\right]$ (Jensen's inequality). The ignore first step trick is the reasoning behind the final equation on the first line. The observation is that the distribution which draws $(a,b) \sim A \times B$ and $b', b'' \sim N(b)$ and outputs (a', b', b'') where $a' = \phi(a, \hat{b})$ is identical to the distribution which draws $a' \sim A$ and a random edge $b' \sim_{B^2} b''$ in B^2 and outputs (a', b', b''). See Figure 2 for intuition.

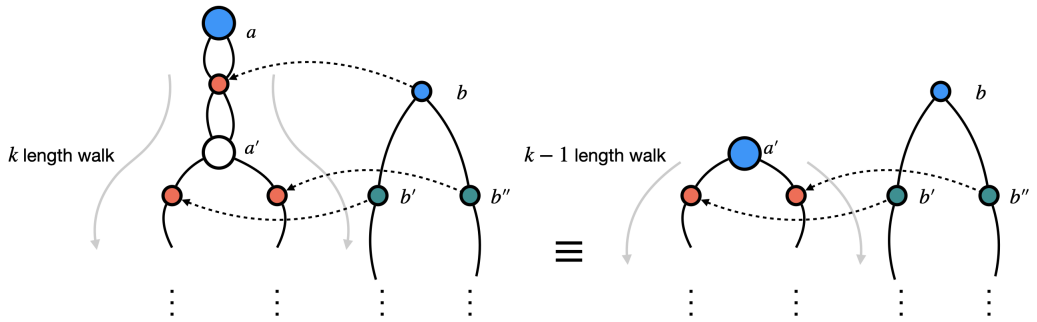

Figure 2: "Ignore first step" trick.

Starting the Replacement Walk in the Middle. A useful feature of random walks on an undirected $d-$regular graph is that the steps can be generated out of order. Specifically, the vertices in a $k-$step random walk can be generated by choosing $a_i \sim A$ first for any $i \in [k]$ and then drawing two walks $(a_i, a_{i+1}, \ldots, a_k) \sim \mathsf{RW}_A^{k-i+1}(a_i)$, $(a_i, a_{i-1}, \ldots, a_1) \sim \mathsf{RW}_A^{i}(a_i)$ and outputting $(a_1, \ldots, a_k)$. Replacement product walks also have this feature, though correctly formulating it requires precision. We will use that the following distribution is identical to $s\mathsf{RW}^k$ for any $i \in \{0, 1 \ldots, k-1\}$:

1. $a_i \sim A$ and a random edge $b_i \sim b_{i+1}$ in B; set $b_{i+1} = \mathsf{shift}(b_{i+1})$;
2. generate $(b_1, \ldots, b_k) \in B^k$ as follows:
   - (i) for $j \geq i+2$, draw $b_j \sim N(b_{j-1})$ and set $b_j = \mathsf{shift}(b_j)$;
   - (ii) for $j \leq i-1$, draw $b_j \sim N(b_{j+1})$ and set $b_j = \mathsf{shift}^{-1}(b_j)$;

3. generate and output $(a_0, \ldots, a_k) \in A^{k+1}$ as follows:

   (i) for $i \geq i+1$, set $a_i = \phi(a_{i-1}, \hat{b}_i)$ where $\hat{b}_i = b_i[1] \in [d]$ denotes the first coordinate of $b_i \in [d]^s$, and where ϕ is the rotation map of A;

   (ii) for $j \leq i-1$, set $a_j = \phi^{-1}(a_{j+1}, \hat{b}_j)$ where $\phi^{-1}(a, \hat{b}) = \phi(a, \hat{b}')$ where $\hat{b}'$ is the local inverse of $\hat{b}$.

An example of how this is used is the first step of the bound for ε_k when $k > s$:

$$\begin{aligned} \varepsilon_k &= \left| \mathbb{E}_{(a_0,\ldots,a_k)\sim s\mathsf{RW}^k} \left[(-1)^{f(a_s)} \cdot (-1)^{f(a_0)\oplus\cdots\oplus f(a_s)} \cdot (-1)^{f(a_s)\oplus\cdots\oplus f(a_k)} \right] \right| \\ &= \left| \mathbb{E}_{\substack{a_s\sim A \\ b_s\sim b_{s+1}}} \left[(-1)^{f(a_s)} \cdot \overleftarrow{g}_s(a_s, b_s) \cdot g_{k-s}(a_s, b_{s+1}) \right] \right|, \end{aligned}$$

where $\overleftarrow{g}_s(a, b)$ indicates that the repalcement walk is drawn in the "backwards" fashion according to Steps 2(ii) and 3(ii) above. Equivalently, $\overleftarrow{g}_s(a, b)$ is the expectation of $(-1)^{f(a_0)\oplus\cdots\oplus f(a_s)}$ over $(a_0, \ldots, a_s) \sim s\mathsf{RW}^s$ conditioned on $(a_s, b_s) = (a, b)$.

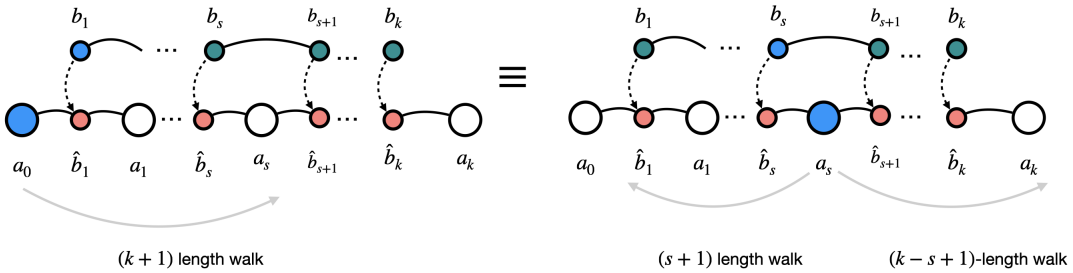

Figure 3: Starting the Replacement Walk in the Middle.

4.3 Bounding the ε_k Terms

In this section we bound the ε_k terms in Claims 4 and 5, thereby proving half of each claim. We bound the σ_k terms in the next section.

The Base Case. This follows directly from the pseudorandomness property, and the analysis already done in Section 1.2 (Claim 1). Specifically, when $k \leq s$, we have

$$\varepsilon_k = \left| \mathbb{E}_a \left[\varepsilon_k(a) \right] \right| = \left| \mathbb{E}_a \left[h_{k+1}(a) \right] \right| \leq \frac{1}{2} \cdot (2\lambda)^{k+1},$$

where $\varepsilon_k(a) = h_{k+1}(a)$ by pseudorandomness (h_{k+1} is the function defined in Claim 1).

The Induction Step. Fix $k > s$. We have

$$\varepsilon_k = \left| \mathbb{E}_{\substack{a\sim A \\ b\sim b'}} \left[(-1)^{f(a)} \cdot \overleftarrow{g}_s(a, b) \cdot g_{k-s}(a, b') \right] \right| \leq \left| \mathbb{E}_{a\sim A} \left[(-1)^{f(a)} \cdot \overleftarrow{\varepsilon}_s(a) \cdot \varepsilon_{k-s}(a) \right] \right| + \lambda\sigma_s\sigma_{k-s},$$

where the equality holds by starting the replacement walk in the middle, and the inequality is the expander mixing lemma (Definition 1) on B. We are using the shorthand $\overleftarrow{\varepsilon}_s(a)$ for $\mathbb{E}_b\big[\overleftarrow{g}_s(a,b)\big]$, and we have used Cauchy-Schwarz to bound the standard deviation terms, just as we did in the computation in the "ignore first step trick" paragraph in Section 4.2. Specifically,

$$\mathbb{E}_a\big[\overleftarrow{\sigma}_s(a)\cdot\sigma_{k-s}(a)\big] \le \sqrt{\mathbb{E}_a[\overleftarrow{\sigma}_s(a)^2]}\sqrt{\mathbb{E}_a[\sigma_{k-s}(a)^2]} \le \overleftarrow{\sigma}_s\sigma_{k-s} = \sigma_s\sigma_{k-s}.$$

By pseudorandomness, $(-1)^{f(a)}\cdot\overleftarrow{\varepsilon}_s(a) = (-1)^{f(a)}\cdot h_{s+1}(a) = \mathbb{E}_{a'\sim N(a)}\big[h_s(a')\big] = \mathbb{E}_{a'\sim N(a)}\big[\varepsilon_{s-1}(a')\big]$, and so we get the desired bound on ε_k via the expander mixing lemma on A, as follows:

$$\begin{aligned}\varepsilon_k &\le \Big|\mathbb{E}_{a\sim a'}\big[\varepsilon_{s-1}(a)\cdot\varepsilon_{k-s}(a')\big]\Big| + \lambda\sigma_s\sigma_{k-s} \le \varepsilon_{s-1}\varepsilon_{k-s} + \lambda^2\sigma_{s-1}\sigma_{k-s} + \lambda\sigma_s\sigma_{k-s}\\ &\le \frac{1}{2}(2\lambda)^s(\varepsilon_{k-s} + 3\sigma_{k-s}).\end{aligned}$$

4.4 Bounding the σ_k Terms

The Base Case. We have already noted that when $1 \le k \le s$, $\varepsilon_{k-1}(a) = h_k(a)$ by pseudorandomness. Thus, $\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] = \mathbb{E}_a\big[h_k(a)^2\big] \le (2\lambda)^{2k-2}$, by Claim 1. It follows from the first step trick that $\sigma_k^2 \le (2\lambda)^{2k-2} + \lambda^2\sigma_{k-1}^2$, which implies $\sigma_k \le (2\lambda)^{k-1} + \lambda\sigma_{k-1}$. Iterating this bound gives

$$\sigma_k \le \lambda^{k-1}\cdot\big(2^{k-1} + 2^{k-2} + \cdots + 2 + 1\big) \le 2\cdot(2\lambda)^{k-1}.$$

The Induction Step. Fix $k > s$. As mentioned in the "ignore first step trick" paragraph in Section 4.2, $\sigma_k^2 \le \mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] + \lambda^2\sigma_{k-1}^2$ holds and so it suffices to bound $\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big]$. By starting the replacement walk in the middle, we get

$$\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] = \mathbb{E}_{\substack{a_{s-1}\sim A\\ b_{s-1}\sim b_s}}\Big[(-1)^{f(a_{s-1})}\cdot g_{k-s}(a_{s-1},b_s)\cdot G(a_{s-1},b_{s-1})\Big],$$

where $G : A\times B\to\mathbb{R}$ is defined by $G(a,b) := \mathbb{E}_{(a_0,\ldots,a_{s-1})}\big[(-1)^{f(a_{s-1})\oplus\cdots\oplus f(a_0)}\cdot\varepsilon_{k-1}(a_0)\big]$, where the expectation is over $(a_0,\ldots,a_{s-1})$ drawn as follows:

- set $b_{s-1} = b$; for $1\le i\le s-2$, draw $b_i\sim N(b_{i+1})$ and then set $b_i = \mathsf{shift}^{-1}(b_i)$;
- set $a_{s-1} = a$; for $0\le i\le s-2$ set $a_i = \phi^{-1}(a_{i+1},\hat{b}_{i+1})$.

The expander mixing lemma (Definition 1) on B gives

$$\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] \le \mathbb{E}_a\Big[(-1)^{f(a)}\cdot\varepsilon_{k-s}(a)\cdot\mu_G(a)\Big] + \lambda\sigma_{k-s}\sigma_G,$$

where $\mu_G := \mathbb{E}_{a,b}\big[G(a,b)\big]$, $\mu_G(a) := \mathbb{E}_b\big[G(a,b)\big]$ and σ_G is such that $\sigma_G^2 + \mu_G^2 = \mathbb{E}_{a,b}\big[G(a,b)^2\big]$. By pseudorandomness, $\mu_G(a) = \mathbb{E}_{(a_0,\ldots,a_{s-1})\sim\mathsf{RW}_A^s(a)}\big[(-1)^{f(a_0)\oplus\cdots\oplus f(a_{s-1})}\cdot\varepsilon_{k-1}(a_{s-1})\big] = \hat{h}_s(a)$, where $\hat{h}_s : A\to\mathbb{R}$ is given by $\hat{h}_s(a) = \mathbb{E}_{(a_1,\ldots,a_s)\sim\mathsf{RW}_A^s}\big[(-1)^{f(a_1)\oplus\cdots\oplus f(a_s)}\cdot\varepsilon_{k-1}(a_s)\big]$. Note this is the function defined in Claim 2, instantiated with $H(a) = \varepsilon_{k-1}(a)$. We have $(-1)^{f(a)}\cdot\mu_G(a) = \mathbb{E}_{a'\sim N(a)}\big[\hat{h}_{s-1}(a')\big]$, and so by the expander mixing lemma on A and Claim 2 we have

$$\begin{aligned}\mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] &\le \mathbb{E}_{a\sim a'}\big[\varepsilon_{k-s}(a)\cdot\hat{h}_{s-1}(a')\big] + \lambda\sigma_{k-s}\sigma_G\\ &\le \varepsilon_{k-s}\cdot 2^{s-3}(\lambda^{s-2}\cdot\hat{\varepsilon}_1 + \lambda^{s-1}\hat{\sigma}_1) + \lambda^2\sigma_{k-s}\cdot 2^{s-3}(\lambda^{s-3}\hat{\varepsilon}_1 + \lambda^{s-2}\hat{\sigma}_1) + \lambda\sigma_{k-s}\sigma_G,\end{aligned}$$

where $\hat{\varepsilon}_1$ and $\hat{\sigma}_1$ are the notations from Claim 2. In our case, $\hat{\varepsilon}_1 = \mathbb{E}_a\big[(-1)^{f(a)} \cdot \varepsilon_{k-1}(a)\big] = \varepsilon_{k-2}$, and $\hat{\sigma}_1 = \sqrt{\mathbb{E}_a[\varepsilon_{k-1}(a)^2] - \hat{\varepsilon}_1^2} \leq \sqrt{\mathbb{E}_{a,b}[g_{k-1}(a,b)^2] - \hat{\varepsilon}_1^2} = \sqrt{\sigma_{k-1}^2 + \varepsilon_{k-1}^2 - \varepsilon_{k-2}^2} \leq \sigma_{k-1}$. We have used Jensen's inequality and that $\varepsilon_{k-2} \geq \varepsilon_{k-1}$. Using these values and remembering the bound $\sigma_k^2 \leq \mathbb{E}_a\big[\varepsilon_{k-1}(a)^2\big] + \lambda^2\sigma_{k-1}^2$ gives

$$\sigma_k^2 \leq \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2} + \lambda\sigma_{k-1})(\varepsilon_{k-s} + \lambda\sigma_{k-s}) + \lambda\sigma_{k-s}\sigma_G + \lambda^2\sigma_{k-1}^2. \tag{4}$$

This is almost the required bound except we still need to simplify σ_G. For this purpose, let us add a parameter to our notation for G, writing G_{s-1} instead of G, since it is an expectation over a length $(s-1)$ "backwards" replacement walk. For $r \leq s-1$, let $\mu_r := \mathbb{E}_{a,b}\big[G_r(a,b)\big]$, let $\mu_r(a) := \mathbb{E}_b\big[G_r(a,b)\big]$ and τ_r such that $\tau_r^2 + \mu_r^2 = \mathbb{E}_{a,b}\big[G_r(a,b)^2\big]$. We need to bound$\tau_{s-1}$. By the ignore first step trick and expander mixing lemma on B^2,

$$\tau_{s-1}^2 \leq \mathbb{E}_{a,b}\big[G_{s-1}(a,b)^2\big] = \mathbb{E}_{\substack{a\sim A \\ b\sim_{B^2} b'}}\Big[G_{s-2}(a,b) \cdot G_{s-2}(a,b')\Big] \leq \mathbb{E}_a\big[\mu_{s-2}(a)^2\big] + \lambda^2\tau_{s-2}^2.$$

We have already seen that $\mu_{s-2}(a) = \hat{h}_{s-1}(a)$, and so by Claim 2 and our computation of $\hat{\varepsilon}_1$ and $\hat{\sigma}_1$ above, $\tau_{s-1}^2 \leq (2\lambda)^{2s-6}(\varepsilon_{k-2}+\lambda\sigma_{k-1})^2 + \lambda^2\tau_{s-2}^2$, which implies $\tau_{s-1} \leq (2\lambda)^{s-3}(\varepsilon_{k-2}+\lambda\sigma_{k-1}) + \lambda\tau_{s-2}$. Iterating this bound (and using $\tau_0 \leq \sigma_{k-1}$) gives

$$\tau_{s-1} \leq \lambda^{s-3}(\varepsilon_{k-2} + \lambda\sigma_{k-1})(2^{s-3} + 2^{s-4} + \cdots) + \lambda^{s-1}\tau_0 \leq 2 \cdot (2\lambda)^{s-3}(\varepsilon_{k-2} + \lambda\sigma_{k-1}) + \lambda^{s-1}\sigma_{k-1}.$$

Plugging this into (4) gives the desired bound:

$$\sigma_k^2 \leq \frac{1}{2}(2\lambda)^{s-2}(\varepsilon_{k-2} + \lambda\sigma_{k-1})\big(\varepsilon_{k-s} + (2+\lambda)\sigma_{k-s}\big) + \lambda^s\sigma_{k-s}\sigma_{k-1} + \lambda^2\sigma_{k-1}^2.$$

5 Expander Hitting Set Lemma

Just for fun, we include a new proof of the classical expander hitting set lemma.

Lemma 3. *Let A be a $\lambda-$expander, and let $S \subset A$ be a set of size $|S| = \rho|A|$. Then for all $t \geq 1$,*

$$\Pr_{(a_1,\ldots,a_t)\sim \mathsf{RW}^t}\Big[a_i \in S \;\forall\; i = 1,\ldots,t\Big] \leq \rho \cdot \big(\rho + \lambda(1-\rho)\big)^{t-1}.$$

Proof. Let $\mathbb{1}_S : A \to \{0,1\}$ be the indicator function of S. For $k \geq 1$, define $g_k : A \to \mathbb{R}$ by

$$g_k(a) = \Pr_{(a_1,\ldots,a_k)\sim \mathsf{RW}^k(a)}\Big[a_i \in S \;\forall\; i = 1,\ldots,k\Big].$$

Let $\varepsilon_k := \mathbb{E}_a\big[g_k(a)\big]$ and σ_k be so $\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\big[g_k(a)^2\big]$. Our proof is by induction on t; it is clear that the lemma holds in the base case. For $k \geq 2$, note that $g_k(a) = \mathbb{1}_S(a) \cdot \mathbb{E}_{a'\sim N(a)}\big[g_{k-1}(a')\big]$ holds, and so

$$\sigma_k^2 + \varepsilon_k^2 = \mathbb{E}_a\big[g_k(a)^2\big] = \mathbb{E}_{\substack{a\sim A \\ a',a''\sim N(a)}}\Big[\mathbb{1}_S(a) \cdot g_{k-1}(a') \cdot g_{k-1}(a'')\Big] = \varepsilon_{2k-1}.$$

We have used that $\mathbb{1}_S(a)^2 = \mathbb{1}_S(a)$ holds for all $a \in A$, and that choosing $a \sim A$ and then two $(k-1)$ length walks starting at a is identical to simply choosing a random walk of length $(2k-1)$. Now, fix $t \geq 2$ and $k, \ell \geq 1$ such that $t = k + \ell$. We have

$$\begin{aligned}\varepsilon_t &= \mathbb{E}_{(a_1,\ldots,a_t)\sim \mathsf{RW}^t}\Big[\mathbb{1}_S(a_1)\cdots\mathbb{1}_S(a_t)\Big] = \mathbb{E}_{a\sim a'}\big[g_k(a)\cdot g_\ell(a')\big] \leq \varepsilon_k\varepsilon_\ell + \lambda\sigma_k\sigma_\ell \\ &\leq \sqrt{\varepsilon_k^2+\lambda\sigma_k^2}\cdot\sqrt{\varepsilon_\ell^2+\lambda\sigma_\ell^2} = \sqrt{(1-\lambda)\varepsilon_k^2+\lambda\varepsilon_{2k-1}}\cdot\sqrt{(1-\lambda)\varepsilon_\ell^2+\lambda\varepsilon_{2\ell-1}}\end{aligned}$$

where the last inequality on the first line is the expander mixing lemma on A and the first inequality on the second line is Cauchy-Schwarz. Note that if $2k-1 < t$ then we can use induction to bound the terms on the right hand side:

$$(1-\lambda)\varepsilon_k^2 + \lambda\varepsilon_{2k-1} \leq \rho\cdot\big(\rho+\lambda(1-\rho)\big)^{2k-2}\cdot\big[(1-\lambda)\rho+\lambda\big] = \rho\cdot\big(\rho+\lambda(1-\rho)\big)^{2k-1}.$$

Therefore, if t is even, we can set $k = \ell = t/2$ to obtain $\varepsilon_t \leq \rho\cdot\big(\rho+\lambda(1-\rho)\big)^{t-1}$, as desired. This does not fully work if t is odd since if we set $k = \lceil t/2\rceil$ and $\ell = \lfloor t/2\rfloor$, then $2k-1 = t$ and so we cannot use induction to bound ε_{2k-1}. However, we can bound ε_k, ε_ℓ, $\varepsilon_{2\ell-1}$ by induction; this gives

$$\varepsilon_t^2 \leq \Big((1-\lambda)\rho^2\big(\rho+\lambda(1-\rho)\big)^{2k-2} + \lambda\varepsilon_t\Big)\cdot\Big(\rho\big(\rho+\lambda(1-\rho)\big)^{2\ell-1}\Big) = 2A\cdot\varepsilon_t + B,$$

where $A = \frac{\lambda\rho}{2}\cdot\big(\rho+\lambda(1-\rho)\big)^{t-2}$ and $B = (1-\lambda)\rho^3\big(\rho+\lambda(1-\rho)\big)^{2t-3}$. Collecting the terms in this way allows us to proceed by completing the square. We get $\varepsilon_t \leq A + \sqrt{A^2+B}$ and we complete the proof by showing that $A + \sqrt{A^2+B} = \rho\big(\rho+\lambda(1-\rho)\big)^{t-1}$. For this last calculation, set the shorthand $\Phi := \rho + \lambda(1-\rho)$. We have

$$A + \sqrt{A^2+B} = \rho\cdot\Phi^{t-2}\cdot\left[\frac{\lambda}{2} + \sqrt{\frac{\lambda^2}{4}+\rho(1-\lambda)\Phi}\right] = \rho\cdot\Phi^{t-1},$$

where the final equation holds because $\Phi = \lambda/2 + \sqrt{\lambda^2/4+\rho(1-\lambda)\Phi}$, which is verified by a simple calculation. □

Acknowledgement

The authors would like to thank Prahladh Harsha and Aparna Shankar for many helpful discussions.

References

[ABN+92] Noga Alon, Jehoshua Bruck, Joseph Naor, Moni Naor, and Ron M Roth. Construction of asymptotically good low-rate error-correcting codes through pseudo-random graphs. *IEEE Transactions on information theory*, 38(2):509–516, 1992.

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Struct. Algorithms*, 3(3):289–304, 1992.

[Alo21] Noga Alon. Explicit expanders of every degree and size. *Combinatorica*, pages 1–17, 2021.

[BATS11] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-ramanujan graphs using the zig-zag product. *SIAM Journal on Computing*, 40(2):267–290, 2011.

[CJW19] Lijie Chen, Ce Jin, and R Ryan Williams. Hardness magnification for all sparse np languages. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1240–1255. IEEE, 2019.

[DK17] Irit Dinur and Tali Kaufman. High dimensional expanders imply agreement expanders. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 974–985. IEEE, 2017.

[Gil52] E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, 1952.

[Plo60] Morris Plotkin. Binary codes with specified minimum distance. *IRE Transactions on Information Theory*, 6(4):445–450, 1960.

[Sha79] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[STV01] Madhu Sudan, Luca Trevisan, and Salil Vadhan. Pseudorandom generators without the xor lemma. *Journal of Computer and System Sciences*, 62(2):236–266, 2001.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251, 2017.

[Var57] R. R. Varshamov. Estimate of the number of signals in error correcting codes. *Docklady Akad. Nauk, S.S.S.R.*, 117:739–741, 1957.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il