Electronic Colloquium on Computational Complexity, Report No. 91 (2021)

Expander Random Walks: The General Case and Limitations

Gil Cohen* Dor Minzer † Shir Peleg ‡ Aaron Potechin § Amnon Ta-Shma¶

June 29, 2021

Abstract

Cohen, Peri and Ta-Shma [CPTS20] considered the following question: Assume the vertices of an expander graph are labelled by ± 1. What "test" functions $f : \{\pm 1\}^t \to \{\pm 1\}$ can or cannot distinguish t independent samples from those obtained by a random walk? [CPTS20] considered only balanced labelling, and proved that all symmetric functions are fooled by random walks on expanders with constant spectral gap. Furthermore, it was shown that functions computable by AC^0 circuits are fooled by expanders with vanishing spectral expansion.

We continue the study of this question and, in particular, resolve all open problems raised by [CPTS20]. First, we generalize the result to all labelling, not merely balanced. In doing so, we improve the known bound for symmetric functions and prove that the bound we obtain is optimal (up to a multiplicative constant). Furthermore, we prove that a random walk on expanders with constant spectral gap does not fool AC^0. In fact, we prove that the bound obtained by [CPTS20] for AC^0 circuits is optimal up to a polynomial factor.

*Tel Aviv University. Supported by ERC starting grant 949499 and by the Israel Science Foundation grant 1569/18. Email: gil@tauex.tau.ac.il.

†Department of Mathematics, Massachusetts Institute of Technology. Email: dminzer@mit.edu.

‡Department of Computer Science, Tel Aviv University, Tel Aviv, Israel. The research leading to these results has received funding from the Israel Science Foundation (grant number 514/20) and from the Len Blavatnik and the Blavatnik Family foundation. Email: shirpele@tauex.tau.ac.il.

§Department of Computer Science, University of Chicago Email: potechin@uchicago.edu.

¶Department of Computer Science, Tel Aviv University, Tel Aviv, Israel. The research leading to these results has received funding from the Israel Science Foundation (grant number 952/18). Email: amnon@tauex.tau.ac.il.

ISSN 1433-8092

Contents

1 Introduction

Expander graphs are among the most useful combinatorial objects in theoretical computer science, pivotal in derandomization [INW94, Rei05], complexity theory [Val76, AKS87, Din07] and coding theory [SS96, KMRZS17, TS17] to name a few. Informally, expanders are sparse undirected graphs that have many desirable pseudorandom properties. A formal definition can be given in several equivalent ways and here we consider the algebraic definition. An undirected graph $G = (V, E)$ is a λ-*spectral expander* if the second largest eigenvalue of its normalized adjacency matrix is bounded above by λ. For simplicity, we only consider regular graphs. In this case, M is also the random walk matrix of G. Many works in the literature have studied explicit constructions of expander graphs (see, e.g., [LPS88, Mar88, BL06, RVW00, BATS11, MOP20]) and utilized their pseudorandom properties. We refer the reader to the excellent expositions [HLW06, Tre17] and to Chapter 4 of [Vad12].

Expanders can be thought of as spectral sparsifiers of the clique. Let $\mathbf{J}$ be the normalized adjacency matrix of the n-vertex complete graph with self-loops. That is, $\mathbf{J}$ is the $n \times n$ matrix with all entries equal to $\frac{1}{n}$. One can express the normalized adjacency matrix M of G by $M = (1 - \lambda)\mathbf{J} + \lambda E$ for some operator E with spectral norm bounded by 1. As such, one can hope to substitute a sample of two *independent* vertices with the "cheaper" process of sampling an edge from an expander and using its two (highly correlated) end-points. This is captured, e.g., by the expander mixing lemma [AC88]. This idea also appears in many derandomization results, e.g., [INW94, AEL95, RRV99, Rei05, RV05, BCG20].

1.1 Random walks on expanders

A useful generalization of the above idea is to consider not just an edge but rather a length $t-1$ random walk (where the length is measured in edges) on the expander as a replacement to t independent samples of vertices. For concreteness, consider a labelling $\mathsf{val} : V \to \{\pm 1\}$ of the vertices with mean $\mu = \mathbf{E}[\mathsf{val}(V)]$. Quite a lot is known about random walks on expanders. In particular, both the hitting property of expanders [AKS87, CW89, IZ89] as well as the expander Chernoff bound [AKS87, Gil98, Hea08] on which we now elaborate.

The hitting property states that for every set $A \subset V$, a length $t-1$ random walk is contained in A with probability at most $(\mu + \lambda)^t$. For $\lambda \ll \mu$, this bound is close to μ^t–the probability of the event with respect to t independent samples. The expander hitting property corresponds to a random walk "fooling" the AND function, that is, for every λ-spectral expander and every labelling val as above, the AND function cannot distinguish with good probability labels obtained by t independent samples from labels obtained by taking a length $t-1$ random walk. The fundamental expander Chernoff bound states that the number of vertices in A visited by a random walk is highly concentrated around its measure $|A|/|V|$. The expander Chernoff bound corresponds to fooling functions indicating whether the normalized Hamming weight of the input is concentrated around some number μ. Perhaps surprisingly, it was shown that even the highly sensitive PARITY function is fooled by a random walk on expanders (this was noted independently by Alon in 1993 for arbitrarily long walks, Wigderson and Rozenman in 2004 for length 1 walks, and [TS17] where the result appears).

However, it is clear that sometimes a random walk *is not* a good replacement to independent samples. To see this, suppose G is a λ-spectral expander for some constant λ, that has a cut $A \subset V$ with $|A| = \frac{|V|}{2}$ and $|E(A, \overline{A})| \geqslant \mu|A|$ for $\mu \geqslant \frac{1}{2} + \widetilde{\Omega}(\lambda)$. Such graphs exist (see [**?**, Section 7]). If one samples t independent vertices $(v_1, \ldots, v_t)$ from the graph, we expect (v_i, v_{i+1}) to cross the cut about half the time, and by the Chernoff bound the actual number of cut crossings is highly concentrated around the mean. In contrast, when we take a random walk on the graph we expect to cross the cut a μ-fraction of the time, and intuitively the number of cut crossings should be concentrated around μ.[1] Thus, the simple test function that counts the number of times we cross the cut and apply a threshold at $\frac{1}{2} + \tau$ for some $\tau = \widetilde{\Theta}(\lambda)$ should distinguish with probability close to 1 between a random walk and independent samples.

This brings to the front a natural question that was recently raised by [CPTS20] (see also the work of Guruswami and Kumar [GK21] who considered a related question).

What test functions does a random walk on an expander fool?

Formally, we compare two distributions on the set $\{\pm 1\}^t$. The first "ideal" distribution is obtained by sampling independently and uniformly at random t vertices $v_1, \ldots, v_t$ and returning $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. If we let $\mu = \mathbf{E}[\mathsf{val}(V)]$, the latter induces the distribution U_t^{μ} in which the t bits are independent and each has mean μ. The second distribution, denoted by $\mathrm{RW}_{G,\mathsf{val}}$, is obtained by taking a length $t-1$ random walk on the graph, namely, sample v_1 uniformly at random from V, and then for $i = 2, 3, \ldots, t$, we sample v_i uniformly at random from the set of neighbors of v_{i-1}, and return $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. Denote

$$\mathcal{E}_{G,\mathsf{val}}(f) = \left|\mathbf{E}\, f(\mathrm{RW}_{G,\mathsf{val}}) - \mathbf{E}\, f(U_t^{\mu})\right|.$$

Informally, $\mathcal{E}_{G,\mathsf{val}}(f)$ measures the distinguishability between these two distributions as observed by the test function f on the graph G with respect to the labelling val. We wish to have a discussion that holds uniformly on all λ-spectral expanders (on any number of vertices) and for every labelling. The bound, however, is expected to depend on the expectation μ of the labelling. We denote by $\mathcal{E}_{\lambda,\mu}(f)$ the supremum of $\mathcal{E}_{G,\mathsf{val}}(f)$ over all λ-spectral expanders G, on any number of vertices, and all labelling functions $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = \mu$.

The work [CPTS20] focuses on the case $\mu = 0$. Their main result states that for such balanced labelling, for every symmetric function $f : \{\pm 1\}^t \to \{\pm 1\}$,

$$\mathcal{E}_{\lambda,0}(f) = O(\lambda \cdot \log^{3/2}(1/\lambda)). \tag{1.1}$$

This readily implies, for the specific case of balanced labelling, a central limit theorem with respect to the total variation distance, strengthening the existing results that considered the Kolmogorov distance [KV86, Lez01, Klo17]. [CPTS20] further considers non-symmetric functions. In particular, they analyze test functions that are computable by AC^0 circuits and prove that if f is computable by a size-s depth-d circuit then

$$\mathcal{E}_{\lambda,0}(f) = O(\sqrt{\lambda} \cdot (\log s)^{2(d-1)}). \tag{1.2}$$

Thus, for balanced labelling, every test function in AC^0 cannot distinguish t independent labels from those obtained by a random walk on a λ-spectral expander provided λ is taken sufficiently

[1] To show such a concentration one needs to prove a Chernoff bound for a walk on the corresponding directed line graph.

small. This result can be thought of as an analog of Braverman's celebrated result [Bra10] (see also [Tal17]) that studies the pseudorandomness of k-wise independent distributions with respect to $\mathbf{AC}^0$ test functions.

1.2 Our contribution

The work of [CPTS20] left four open problems:

1. Can the results be extended to unbalanced labelling, namely, $\mu \neq 0$?
2. Is the $\log^{3/2}(1/\lambda)$ factor in Equation (1.1) inherent or rather it is an artifact of the proof?
3. Does a constant spectral expansion suffice to fool $\mathbf{AC}^0$ test functions or rather a bound as given by Equation (1.2) is necessary?
4. Note that the bound given by Equation (1.1) does not vanish with t. However, for all the symmetric functions that were studied, such as majority and parity, the bound does vanish with t. Can a bound that vanishes with t be obtained for *all* symmetric functions?

In this work we resolve all four open problems left by [CPTS20]. First, answering Problem 1 and Problem 2, we generalize their main result to any μ and without incurring the poly-logarithmic factor. [2]

Theorem 1.1. *For every symmetric function $f : \{\pm 1\}^t \to \{\pm 1\}$, all $\mu \in (-1, 1)$ and $0 < \lambda < \frac{1-|\mu|}{128e}$ it holds that*

$$\mathcal{E}_{\lambda,\mu}(f) \leqslant \frac{124}{\sqrt{1-|\mu|}} \cdot \lambda.$$

Theorem 1.1 readily implies a central limit theorem with respect to the total variation distance that holds for all labelling. The proof of Theorem 1.1 can be found in Section 3.2. Furthermore, in Section 3.3 we give bounds that vanish with t for specific functions such as certain threshold functions (Theorem 3.10), generalizing the bound on the majority function obtained by [CPTS20], and weight indicators (Theorem 3.12).

Second, addressing Problem 3, we prove that constant spectral expansion does not suffice to fool AC^0 circuits. In fact, the bound obtained by [CPTS20] is tight up to a polynomial.

Let us denote with $\mathsf{AC}(d)$ the class of all languages with polynomial size boolean circuit of depth at most d.

Theorem 1.2. *There exists a constant $\varepsilon > 0$ such that the following holds. For every integer $d \geqslant 3$ there exist $t_d, c_d \in \mathbb{N}$, and a family of functions $(h_t)_{t_d \leqslant t \in \mathbb{N}} \subset \mathsf{AC}(d)$ such that the following holds. For every $\lambda \geqslant \frac{c_d}{\log^{d-2} t}$ there is a λ-spectral expander $G = (V, E)$ and a labelling $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = 0$ such that $\mathcal{E}_{G,\mathsf{val}}(h_t) \geqslant \varepsilon$.*

[2] Aaron Potechin independently improved the bound obtained in [CPTS20] by removing the poly-logarithmic factor, for the case $\mu = 0$.

The proof of Theorem 1.2 appears in Section 4.2.

Lastly, our third result addresses Problem 4. We show that the bound for general symmetric functions does not vanish with t (even for $\mu = 0$).

Theorem 1.3. *There exists a family of symmetric functions $(f_t)_{t\in\mathbb{N}}$ where $f_t : \{\pm 1\}^t \to \{\pm 1\}$ such that for every λ there is a λ-spectral expander $G = (V, E)$, and a labelling $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = 0$ such that for all t, $\mathcal{E}_{G,\mathsf{val}}(f_t) = \Omega(\lambda)$.*

The proof of Theorem 1.3 appears in Section 4.3.

1.3 Proof overview

In this section we give a brief overview of our proofs. We start by recalling the underlying proof strategy of [CPTS20]. The key idea in [CPTS20] is to expand the test function under consideration in the Fourier basis. The question of fooling general test functions then reduces to the study of to what extent a Fourier character (namely, a parity) is fooled by an expander random walk. [CPTS20] obtain their results by combining the above with known facts about the Fourier expansion of the test function at hand.

1.3.1 Upper bound for symmetric functions – Theorem 1.1

For extending the result of [CPTS20] to all $\mu \in (-1, 1)$ we expand the test function $f : \{\pm 1\}^t \to \{\pm 1\}$ in a suitable variant of the "standard" Fourier basis. The basis we choose consists of $\prod_{i\in S} \frac{x_i-\mu}{\sqrt{1-\mu^2}}$ for all $S \subseteq [t]$. This change allows us to generalize the framework to all $\mu \in (-1, 1)$. Recall, however that Theorem 1.1 also improves upon the bound obtained by [CPTS20] by removing the poly-logarithmic factor (even for the special case $\mu = 0$). To achieve this, we deviate from the original analysis. The proof strategy of [CPTS20] is to bound the weight indicator functions so to handle weights around the mean and invoke the expander Chernoff bound for bounding the remaining weights. Our approach does not go through analyzing weight indicator functions nor uses the expander Chernoff bound. Instead, we deviate from the original analysis of [CPTS20] at a certain point and use a very simple bound on the Fourier mass of symmetric functions (Claim 3.9). This allows us to save upon the poly-logarithmic factor. Moreover, our proof is significantly simpler.

1.3.2 Tightness results

For the tightness results we work with a cayley graph over the communtative boolean group $G = \mathbb{Z}_2^n$. Such Cayley graphs cannot give Ramanujan exapnders, but with logarithmic degree can have vanishing second eigenvalue. The advantage of using such a graph is that the eigenvectors of the graph correspond to the characterisitic functions of $\mathbb{Z}_2^n$ (regardless of the set of the generators used). The important property that we use is that all the eigenvectors of the fraph, and in particular the eigenve tor with the second largest eigenvalue, has ± 1 entries, up to normalization. We then choose the labelling to correspond to the eigenvector with the second largest eigenvalue.

For a labelling $\ell : V \to \{\pm 1\}$ let us denote by P the diagonal matrix with $\ell(u)$ in the i'th element on the diagonal. Also, let G denote the transition matrix of the graph. It is an easy excersie that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^{t} P^{\delta_i} G \right) \mathbf{1},$$

where δ_1 is 1 if $i \in S$ an 0 otherwise. When we choose val to be the second eigenvector, then $P\mathbf{1} = v_2$ and $Pv_2 = \mathbf{1}$. Also $G\mathbf{1} = \mathbf{1}$ and $Gv_2 = \lambda v_2$. It follows that $\left(\prod_{i=1}^{t} P^{\delta_i} G\right) \mathbf{1}$ belongs to the two dimensional subspace Span $\mathbb{1}, v_2$ and, furthermore, has a closed expression as a function of t, λ and S. We then choose a function f (a thereshold function for the symmetric case and an iterated Tribes function for the $\mathsf{AC}(d)$ case) to prove tightness.

1.3.3 Tightness for function in $\mathsf{AC}(d)$ – Theorem 1.2

To prove Theorem 1.2 we take the Boolean hypercube G and induce a balanced ± 1 labelling val of the vertices of G using an eigenvector corresponding to its second largest eigenvalue. The key observation of the construction appears in Claim 4.7. The idea is that two adjacent bits obtained by a random walk are λ correlated. Thus, valuating a function f on the parity of consecutive bits obtained by a random walk, is the same as measuring the noise operator $T_\lambda(f)$. Thus, we construct functions that are highly sensitive to small noise. We start by constructing two functions, $f, g \in \mathsf{AC}(2)$, and then define a family of functions recursively by composing f and g in an alternating fashion. This recursive definition yields a family of functions $h_d \in \mathsf{AC}(d+1)$ for every d. In each step we increase the noise sensitivity of h_d by a logarithmic factor, that yields the desired of dependence of $\mathcal{E}_{G,\mathsf{val}}(h_d)$ on d.

1.3.4 Tightness for symmetric functions – Theorem 1.3

To prove Theorem 1.3 we consider the threshold function $f(x_1, \ldots, x_t) = \mathsf{Sign}(x_1 + \cdots + x_t - \sqrt{t})$. We again use the Boolean hypercube G with the same labelling.. Our proof uses the structure of this specific graph (Corollary 4.3) to argue that not too many cancellations occur in the contribution of the second Fourier level of f to $\mathcal{E}_{G,\mathsf{val}}(f)$. Further, f was chosen specifically so to have high mass on its second Fourier level. On the other hand, we bound the absolute value of the contribution of the higher Fourier levels to $\mathcal{E}_{G,\mathsf{val}}(f)$. Combining these two bounds, we deduce the desired lower bound on $\mathcal{E}_{G,\mathsf{val}}(f)$.

1.4 Discussion and open problems

We conclude this section with several remarks and open problems that follow from our work.

1. Can one combine the distribution obtained by a random walk on an expander with another pseudorandom distribution to obtain stronger results for functions in AC^0. For example, does permuting the values of the random walk with a pairwise independent permutation yields a distribution that fools AC^0?

2. The lower bounds proved in Section 4 are based on the tightness of our analysis for Caylay Graphs over the Boolean cube. It is well-known that every Cayley Graph with constant

expansion gap, has degree that is at least logarithmic in the number of vertices. Thus, a natural question is whether we can give similar lower bounds for graphs with constant degree.

3. On the other hand, it is still possible that there is a family of graphs that fools all symmetric functions with vanishing bounds, and/or functions in AC^0? Finding such graphs is a compelling goal, that might require studying additional properties of expander graphs.

4. There is still a polynomial gap between the value of λ that fools functions in AC^0 and the lower bound obtained in Section 4.2. Any progress towards closing this gap will be interesting.

2 Preliminaries

We let $[n]$ denote the set $\{1, \ldots, n\}$. We let $\mathbb{1} \in \mathbb{R}^n$ denote the all 1s vector, i.e., $\mathbb{1} = (1, \ldots, 1)^\mathsf{T} \in \mathbb{R}^n$. We let $\mathbf{1} \in \mathbb{R}^n$ denote the normalize vector of $\mathbb{1}$, i.e $\mathbf{1} = \frac{1}{\sqrt{n}} \cdot \mathbb{1}$. We let $\mathbf{J} = \mathbf{1}\mathbf{1}^\mathsf{T}$.Throughout the paper, we make use of the following well known inequalities about binomial coefficients.

Claim 2.1. *Let $0 < \lambda < 1$, and integers $r \geqslant 0$, $a \geqslant b \geqslant 1$. Then, $(\frac{a}{b})^b \leqslant \binom{a}{b} \leqslant (\frac{ea}{b})^b$.*

2.1 Fourier analysis

Consider the space of functions $f : \{\pm 1\}^t \to \mathbb{R}$, along with the inner product

$$\langle f, g \rangle = 2^{-t} \sum_{x \in \{\pm 1\}^t} f(x)g(x).$$

It is a well-known fact that the set $\{\chi_S \mid S \subseteq [t]\}$, where $\chi_S = \prod_{i \in S} x_i$, forms an orthonormal basis with respect to this inner product, which is called the Fourier basis. Thus every function $f : \{\pm 1\}^t \to \mathbb{R}$ can be uniquely represented as $f(x) = \sum_{S \subseteq [t]} \widehat{f}(S)\chi_S(x)$, where $\widehat{f}(S) \in \mathbb{R}$.

In this work we consider other bases, with respect to a similar inner product. Let $\mu \in [0, 1]$, and by denote U_t^μ the distribution of the distribution over $\{\pm 1\}^t$ where each bit is chosen independently with expectation μ. Define $\langle f, g \rangle = \mathbf{E}_{x \sim U_t^\mu}[f(x)g(x)]$. Denote by $\sigma = \sqrt{1 - \mu^2}$, and let $\chi_S^\mu(x) = \prod_{i \in S} \frac{x_i - \mu}{\sigma}$. It is easy to see that the set $\left\{\chi_S^\mu \mid S \subseteq [t]\right\}$, forms an orthonormal basis with respect to this inner product, which is called the μ-biased Fourier basis. To see this, note that by design for $S \neq \emptyset$, $\mathbf{E}[\chi_S^\mu] = 0$ and $\mathbf{E}[(\chi_S^\mu)^2] = 1$. Similarly to the standard Fourier basis, every function $f : \{\pm 1\}^t \to \mathbb{R}$ can be uniquely represented as $f(x) = \sum_{S \subseteq [t]} \widehat{f_\mu}(S)\chi_S^\mu(x)$, where $\widehat{f_\mu}(S) \in \mathbb{R}$.

3 Positive results

In this section we give upper bounds on $\mathcal{E}_{\lambda,\mu}(f)$ for different families of functions. Hereby generalizing, the results if [CPTS20] for $\mu \in (-1, 1)$. We start with presenting the general framework of the proofs, giving bounds on parity test functions in Section 3.1. We prove Theorem 1.1 in Section 3.2. In Section 3.3 we improve the bounds obtained in Section 3.2 for certain families of functions.

Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a labeling of the vertices of G with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Let $t \geqslant 1$ an integer. We recall from the introduction that we want to compare two distributions on $\{\pm 1\}^t$.

- The distribution obtained by sampling t vertices $v_1, \ldots, v_t$ uniformly and independently at random, and outputting the ordered tuple $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. Note that this is the same distribution as sampling a sequence of t elements in $\{\pm 1\}$ independently such that the marginal distribution of each entry has expectation μ. We denote this distribution by U_t^μ.
- $\mathrm{RW}_{G,\mathsf{val}}$ is the distribution obtained by sampling a random length $t-1$ path $v_1, \ldots, v_t$ in G and outputting the ordered tuple $(\mathsf{val}(v_1), \ldots, \mathsf{val}(v_t))$. Equivalently, sample v_1 uniformly at random from V. Then, for $i = 2, 3, \ldots, t$, sample v_i uniformly at random from the neighbors of v_{i-1}.

Let $f : \{\pm 1\}^t \to \{\pm 1\}$ be a test function. Expand f in the μ-biased Fourier basis,

$$f(x) = \sum_{S \subseteq [t]} \widehat{f_\mu}(S) \chi_S^\mu(x).$$

Lemma 3.1. *Let $G = (V, E)$ be a regular λ-spectral expander, and let $\mathsf{val} : V \to \{\pm 1\}$ be a labelling of the vertices of G with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Then, for every function $f : \{\pm 1\}^t \to \mathbb{R}$,*

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{\substack{S \subseteq T \\ S \neq \emptyset}} |\widehat{f_\mu}(S)| \mathcal{E}_{G,\mathsf{val}}(\chi_S^\mu).$$

Proof. Since val has expectation μ, for $S \neq \emptyset$, $\mathbf{E}[\chi_S^\mu(U_t^\mu)] = 0$ and thus $\mathbf{E}[f(U_t^\mu)] = \widehat{f_\mu}(\emptyset)$.

$$\begin{aligned}\mathcal{E}_{G,\mathsf{val}}(f) &= |\mathbf{E}\, f(\mathrm{RW}_{G,\mathsf{val}}) - \mathbf{E}\, f(U_t^\mu)| \\ &= \Bigg| \sum_{\substack{S \subseteq T \\ S \neq \emptyset}} \widehat{f_\mu}(S)\, \mathbf{E}[\chi_S^\mu(\mathrm{RW}_{G,\mathsf{val}})] \Bigg|.\end{aligned}$$

Thus, for $S \neq \emptyset$, $\mathcal{E}_{G,\mathsf{val}}(\chi_S^\mu) = |\,\mathbf{E}[\chi_S^\mu(\mathrm{RW}_{G,\mathsf{val}})]|$. The proof follows by the triangle inequality. $\square$

Lemma 3.1 motivates us to consider parity test functions. This is the content of the following section.

3.1 Biased parity test functions

In this section we analyze to what extent expander random walks fool parity tests functions. We start by introducing some notation. For an integer $k \geqslant 2$, we define the family $\mathcal{F}_k$ of subsets of $[k-1]$ that, informally, consists of all subsets for which at least one of every two consecutive elements participate in the set. We also require the "end points" $1, k-1$ to participate in the set. Formally, we define

$$\mathcal{F}_k = \{I \subseteq [k-1] \mid \{1, k-1\} \subseteq I \text{ and } \forall j \in [k-2] \ \{j, j+1\} \cap I \neq \emptyset\}. \tag{3.1}$$

So, for example, $\mathcal{F}_6$ consists of the elements $\{1,3,5\}$, $\{1,2,4,5\}$ as well as of all subsets of $[5]$ that has as a subset any one of these two elements, namely, $\{1,2,3,5\}$, $\{1,3,4,5\}$ and $\{1,2,3,4,5\}$. We extend the definition in the natural way to $k = 0, 1$ by setting $\mathcal{F}_0 = \mathcal{F}_1 = \emptyset$.

Definition 3.2. *For an integer $t \geqslant 1$, and $2 \leqslant k \leqslant t$, and $j \in [k-1]$ define the map $\mathbf{\Delta}_j : \binom{[t]}{k} \to \mathbb{N}$ as follows. Let $S \subseteq [t]$, of size $k \geqslant 2$, and denote $S = \{s_1, \ldots, s_k\}$ where $s_1 < \cdots < s_k$. For $i \in [k-2]$ write $\Delta_i = s_{i+1} - s_i$. Define $\mathbf{\Delta}_j(S) = \min(\Delta_j, \Delta_{j+1})$.*

Definition 3.3. *For an integer $t \geqslant 1$ define the map $\mathbf{\Delta} : \binom{[t]}{\geqslant 2} \to \mathbb{N}$ as follows. Let $S \subseteq [t]$, of size $k \geqslant 2$. For $k = 2$ we define $\mathbf{\Delta}(S) = \mathbf{\Delta}_1(S)$, and for $k \geqslant 3$,*

$$\mathbf{\Delta}(S) = \sum_{i=1}^{k-2} \mathbf{\Delta}_i(S). \tag{3.2}$$

Proposition 3.4. *Let $G = (V, E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ be a labelling of the vertices of G with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Then, for every integers $1 \leqslant k \leqslant t$ and every subset $S \subseteq [t]$ of size k,*

$$\mathcal{E}_{G,\mathsf{val}}(\chi_S^{\mu}) \leqslant \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \cdot \sum_{I \in \mathcal{F}_k} \lambda^{\sum_{j \in I} \Delta_j(S)}.$$

Before proving Proposition 3.4, we remark that for sets of size $|S| = 1$, the sum is taken over the empty index set $\mathcal{F}_1$ and so, by the standard convention, the sum equals to 0. We also observe that Proposition 3.5 follows by Proposition 3.4. To see this, note that for every $I \in \mathcal{F}_k$,

$$2\sum_{i \in I} \Delta_i \geqslant \sum_{i=1}^{k-2} \min(\Delta_i, \Delta_{i+1}). \tag{3.3}$$

Indeed, if we define δ_i to be the corresponding indicator for $i \in I$, namely, $\delta_i = 1$ if $i \in I$ and $\delta_i = 0$ otherwise, we see that

$$2\sum_{i \in I} \Delta_i \geqslant \sum_{i=1}^{k-2} \delta_i \Delta_i + \delta_{i+1}\Delta_{i+1}.$$

Equation (3.3) follows since $\delta_i \Delta_i + \delta_{i+1}\Delta_{i+1} \geqslant \min(\Delta_i, \Delta_{i+1})$ as indeed, for every $i \in [k-2]$, at least one of $i, i+1$ is in I. Now, recall that in Equation (3.2), the right hand side of Equation (3.3) was denoted by $\mathbf{\Delta}(S)$. As $|\mathcal{F}_k| \leqslant 2^{k-1}$, Proposition 3.5 follows by Proposition 3.4. We turn to prove Proposition 3.4.

Proof of Proposition 3.4. Consider any nonempty set $S \subseteq [t]$ of size $|S| = k$. As $\mathbf{E}[\chi_S(U_t^{\mu})] = 0$, we have that

$$\mathcal{E}_{G,\mathsf{val}}(\chi_S^{\mu}) = \left|\mathbf{E}[\chi_S^{\mu}(\mathrm{RW}_{G,\mathsf{val}})]\right|.$$

We wish to express the right hand side algebraically. Let $n = |V|$ and identify V with $[n]$ in an arbitrary way. Let P be the $n \times n$ diagonal matrix with

$$P_{v,v} = \frac{\mathsf{val}(v) - \mu}{\sqrt{1-\mu^2}}$$

for every $v \in [n]$. We slightly abuse notation and denote the random walk matrix (that is, the normalized adjacency matrix) of G also by G. Define $\delta_i = 1$ if $i \in S$ and $\delta_i = 0$ otherwise and observe that

$$\mathbf{E}[\chi_S^\mu(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^{t} P^{\delta_i} G \right) \mathbf{1},$$

where recall $\mathbf{1}$ is the vector all of whose entries equal to $\frac{1}{\sqrt{n}}$. Indeed, informally, at the i'th step we take a random step using G and then, depending on i being an element of I or not, we multiply by P or by I, respectively. Thus, we can write

$$\mathbf{E}[\chi_S^\mu(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T G^{t-s_k} \left(\prod_{i=1}^{k-1} P G^{\Delta_i} \right) P G^{s_1} \mathbf{1} = \mathbf{1}^T \left(\prod_{i=1}^{k-1} P G^{\Delta_i} \right) P\mathbf{1}, \tag{3.4}$$

where we have used the regularity of G, namely, $G\mathbf{1} = \mathbf{1}$.

Next, we use the spectral decomposition of G. As G is a λ-spectral expander we know that $G = \mathbf{J} + \lambda E$ where $\| E \| \leqslant 1$. Similarly, As G^ℓ is a λ^ℓ-spectral expander we have that $G^\ell = \mathbf{J} + \lambda^\ell E_\ell$ for some operator E_ℓ with bounded norm $\| E_\ell \| \leqslant 1$. Thus,

$$\prod_{i=1}^{k-1} P G^{\Delta_i} = \sum_{I \subseteq [k-1]} \prod_{i=1}^{k-1} P B_i(I), \tag{3.5}$$

where

$$B_i(I) = \begin{cases} \lambda^{\Delta_i} E_{\Delta_i} & i \in I; \\ \mathbf{J} & \text{otherwise.} \end{cases}$$

For $I \subseteq [k-1]$ let

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{k-1} P B_i(I) \right) P\mathbf{1}.$$

Equations (3.4) and (3.5) imply that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \sum_{I \subseteq [k-1]} e_I. \tag{3.6}$$

Not all subsets $I \subseteq [k-1]$ contribute non-zero values e_I to the sum. Indeed, if $k-1 \notin I$ then

$B_{k-1}(I) = \mathbf{J}$ and so

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) (P\mathbf{J})P\mathbf{1}$$
$$= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) (P\mathbf{1}\mathbf{1}^T)P\mathbf{1}$$
$$= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I) \right) P\mathbf{1}(\mathbf{1}^T P\mathbf{1}).$$

As

$$\mathbf{1}^T P\mathbf{1} = \frac{1}{\sqrt{1-\mu^2}} \cdot \sum_{i\in[n]} \frac{\mathsf{val}(i) - \mu}{n} = \frac{\mathbf{E}[\mathsf{val}(V)] - \mu}{\sqrt{1-\mu^2}} = 0,$$

we have that $e_I = 0$. Similarly $e_I = 0$ for I not containing 1. Moreover, if $j, j+1$ are both not contained in I for some $j \in [k-2]$ then

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I) \right) (PB_j(I))(PB_{j+1}(I)) \left(\prod_{i=j+2}^{k-2} PB_i(I) \right) P\mathbf{1}$$
$$= \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I) \right) (P\mathbf{J})(P\mathbf{J}) \left(\prod_{i=j+2}^{k-2} PB_i(I) \right) P\mathbf{1}.$$

However,

$$(P\mathbf{J})(P\mathbf{J}) = (P\mathbf{1}\mathbf{1}^T)(P\mathbf{1}\mathbf{1}^T) = P\mathbf{1}(\mathbf{1}^T P\mathbf{1})\mathbf{1}^T = 0.$$

Thus, any subset $I \subseteq [k-1]$ that may contribute to the sum in Equation (3.6) is contained in $\mathcal{F}_k$ as defined in Equation (3.1). Let M be the $n \times n$ diagonal matrix defined by $M_{v,v} = \mathsf{val}(v)$ for all $v \in [n]$. Note that $P = \frac{1}{\sqrt{1-\mu^2}}(M - \mu I)$. As $\|M\| = 1$, using the triangle inequality we get

$$\|P\| \leqslant \frac{\|M\| + \|\mu I\|}{\sqrt{1-\mu^2}} \leqslant \frac{1+|\mu|}{\sqrt{1-\mu^2}} = \sqrt{\frac{1+|\mu|}{1-|\mu|}}.$$

By the Cauchy-Schwartz inequality and by the sub-multiplicativity of the Euclidean norm, for every $I \in \mathcal{F}_k$ we have that

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{k-1} PB_i(I) \right) P\mathbf{1}$$
$$\leqslant \prod_{i=1}^{k-1} \|PB_i(I)\|$$
$$\leqslant \|P\|^{k-1} \prod_{i\in I} \|B_i(I)\|.$$

Recall that for every $i \in I$, $B_i(I) = \lambda^{\Delta_i} E_{\Delta_i}$ and that $\|E_{\Delta_i}\| \leqslant 1$. Thus,

$$\prod_{i\in I} \|B_i(I)\| \leqslant \prod_{i\in I} \lambda^{\Delta_i},$$

which concludes that $e_I \leqslant \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \prod_{i\in I} \lambda^{\Delta_i}$. $\square$

In particular, we prove the following.

Proposition 3.5. *Let $G=(V,E)$ be a regular λ-spectral expander, and $\mathsf{val}: V \to \{\pm 1\}$ a labelling of the vertices of G with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Then, for every integers $1 \leqslant k \leqslant t$ and every subset $S \subseteq [t]$ of size k,*

$$\mathcal{E}_{G,\mathsf{val}}\left(\chi_S^{\mu}\right) \leqslant \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} 2^k \cdot \lambda^{\mathbf{\Delta}(S)/2}.$$

This follows immediately from Theorem 3.4

3.2 Symmetric test functions

In this section we prove Theorem 1.1. For convenience we restate it here,

Theorem 3.6. *For every symmetric function f, $\mu \in (-1,1)$ and $0 < \lambda < \frac{1-|\mu|}{128e}$ it holds that*

$$\mathcal{E}_{\lambda,\mu}(f) \leqslant \frac{124}{\sqrt{1-|\mu|}} \cdot \lambda.$$

Given a symmetric function $f: \{\pm 1\}^t \to \mathbb{R}$ and $k \in [t]$ we slightly abuse notation and denote by $\widehat{f_\mu}(k) = |\widehat{f_\mu}([k])|$. For analyzing the random walk with respect to symmetric test functions, we define for every integer $k \in \{0, 1, \ldots, t\}$,

$$\beta_k^{\mu} = \sum_{\substack{S\subseteq[t]\\|S|=k}} \mathbf{E}[\chi_S^{\mu}(\mathrm{RW}_{G,\mathsf{val}})]. \tag{3.7}$$

Note that β_k is independent of the choice of test function. However, for symmetric tests functions, these quantities will appear in the analysis, and so we begin by analyzing them. By Proposition 3.4,

$$\begin{aligned}\beta_k^{\mu} &\leqslant \sum_{\substack{S\subseteq[t]\\|S|=k}} \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \cdot \sum_{I\in\mathcal{F}_k} \lambda^{\sum_{j\in I}\Delta_j(S)} \\ &= \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \cdot \sum_{\substack{S\subseteq[t]\\|S|=k}} \sum_{I\in\mathcal{F}_k} \lambda^{\sum_{j\in I}\Delta_j(S)}.\end{aligned}$$

Denote

$$\beta_k = \sum_{\substack{S\subseteq[t]\\|S|=k}} \sum_{I\in\mathcal{F}_k} \lambda^{\sum_{j\in I}\Delta_j(S)}.$$

We have that $\beta_k^{\mu} \leqslant \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \beta_k$. A straightforward corollary of Lemma 3.1 is the following

Corollary 3.7. *Let $G = (V, E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a labelling of the vertices of G with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Then, for every symmetric function $f : \{\pm 1\}^t \to \mathbb{R}$,*

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=2}^{t} \widehat{f_\mu}(k)\beta_k^\mu \leqslant \sum_{k=2}^{t} \widehat{f_\mu}(k) \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \beta_k.$$

We invoke the following bound on β_k obtained in [CPTS20].

Lemma 3.8 ([CPTS20], Lemma 4.4)**.**

$$\beta_k \quad \leqslant \quad 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \tag{3.8}$$

Claim 3.9. *Let $f : \{\pm 1\}^t \to \{\pm 1\}$ be a symmetric function, then for every $\mu \in (0, 1)$ and $S \subset [t]$ it holds that*

$$\left|\widehat{f_\mu}(S)\right| \leqslant \frac{1}{\sqrt{\binom{t}{|S|}}}.$$

Proof. By Parseval's equality,

$$1 = \mathbf{E}[f^2] = \sum_{S \subset [t]} \widehat{f_\mu}(S)^2.$$

For a symmetric function, every $S_1, S_2 \subseteq [t]$ with $|S_1| = |S_2|$ satisfy $\widehat{f}^\mu(S_1) = \widehat{f}^\mu(S_2)$ and thus we can write

$$1 = \mathbf{E}[f^2] = \sum_{k \leqslant t} \widehat{f_\mu}(k)^2 \binom{t}{k}$$

which implies that for every $k \leqslant t$, $\left|\widehat{f_\mu}(k)\right| \leqslant \frac{1}{\sqrt{\binom{t}{k}}}$.

□

We are now ready to prove Theorem 3.6.

Proof of Theorem 3.6. Let $G = (V, E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a labelling of V with $\mathbf{E}(\mathsf{val}) = \mu$. Write

$$f(x_1, \ldots, x_t) = \sum_{S \subseteq [t]} \widehat{f}_\mu(|S|)\chi_S^\mu(x_1, \ldots, x_t).$$

Denote

$$\nu = \sqrt{\frac{1+|\mu|}{1-|\mu|}}.$$

By Corollary 3.7,

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=2}^{t} \widehat{f_\mu}(k)\nu^{k-1}\beta_k.$$

Claim 3.9 and Lemma 3.8 then imply that

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=2}^{t} \frac{1}{\sqrt{\binom{t}{k}}} \nu^{k-1} 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil}$$

A straightforward calculation implies that

$$\frac{1}{\sqrt{\binom{t}{k}}} \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \leqslant (2e)^{k/2}$$

and so

$$\begin{aligned}
\mathcal{E}_{G,\mathsf{val}}(f) &\leqslant \sum_{k=2}^{t} \nu^{k-1} (8e)^{k/2} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \\
&\leqslant \sum_{k=2}^{t} \nu^{k-1} (16e)^{k/2} \lambda^{\frac{k}{2}} \\
&= \nu^{-1} \sum_{k=2}^{t} \alpha^k \\
&\leqslant \frac{\nu^{-1}\alpha^2}{1-\alpha},
\end{aligned}$$

where $\alpha = \sqrt{16e\nu^2\lambda}$. Per our assumption that $\lambda < \frac{1-|\mu|}{128e}$ we have $\alpha \leqslant \frac{1}{2}$ and so

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant 2\nu^{-1}\alpha^2 = 32e\nu\lambda \leqslant \frac{124\lambda}{\sqrt{1-|\mu|}}.$$

□

3.3 Vanishing bounds

In this section we derive bounds on $\mathcal{E}_{\lambda,\mu}$ that the vanish with length of the walk t for specific functions. In particular, for the majority function (Section 3.3.1) and for weight indicator functions (Section 3.3.2).

3.3.1 Bounds for the majority function

In this section we use the results developed so far to prove that random walks fool the majority function. For $w \in [t]$ we define $\mathsf{Th}_w : \{\pm 1\}^t \to \{\pm 1, 0\}$ by $\mathsf{Th}_w(x) = 1$ if $|\{i \in [t] \mid x_i = 1\}| \geqslant w$ and $\mathsf{Th}_w(x) = -1$ otherwise. Put differently,

$$\mathsf{Th}_w(x_1, \dots, x_t) = \mathsf{Sign}(x_1 + \cdots + x_t - 2w + t),$$

with the understanding that $\mathsf{Sign}(0) = 0$. Note that for odd t, the function $\mathsf{Th}_{t/2}$ is the majority function.

Theorem 3.10. *For every $\mu \in (-1, 1)$, $0 < \lambda < \frac{1-|\mu|}{192e}$ and every $t \in \mathbb{N}$*

$$\mathcal{E}_{\lambda,\mu}(\mathsf{Th}_{\frac{t+\mu t}{2}}) \leqslant \frac{1}{\sqrt{t}} \cdot \frac{(96e)^2\lambda^2}{1-|\mu|}.$$

From here on, for ease of readability we denote $\mathsf{Th}_{\frac{t+\mu t}{2}}$ by f. Note that $f = \mathsf{Sign}(\sum_i x_i - \mu t)$.

Claim 3.11. *For $|S|$ even, $\widehat{f_\mu}(S) = 0$.*

Proof. Consider the transformation $y_i = 2\mu - x_i$. It holds that

$$\begin{aligned} f(y_1, \dots, y_t) &= \mathsf{Sign}\left(\sum_{i=1}^{t}(y_i - \mu)\right) \\ &= -\mathsf{Sign}\left(-\sum_{i=1}^{t}(\mu - x_i)\right) \\ &= -\mathsf{Sign}\left(\sum_{i=1}^{t}x_i - \mu t\right) \\ &= -f(x_1, \dots, x_t). \end{aligned}$$

Expanding f in the respective Fourier basis we get

$$f(x_1, \dots, x_t) = \sum_{S \in [t]} \widehat{f_\mu}(S)\chi_S^\mu(x_1, \dots, x_t).$$

Note that $\chi_S^\mu(y_1, \dots, y_t) = (-1)^{|S|}\chi_S^\mu(x_1, \dots, x_t)$, and so

$$\begin{aligned} \sum_{S \in [t]} -(\widehat{f_\mu}(S))\chi_S^\mu(x_1, \dots, x_t) &= -f(x_1, \dots, x_t) \\ &= f(y_1, \dots, y_t) \\ &= \sum_{S \in [t]} \widehat{f_\mu}(S)\chi_S^\mu(y_1, \dots, y_t) \\ &= \sum_{S \in [t]} (-1)^{|S|}\widehat{f_\mu}(S)\chi_S^\mu(x_1, \dots, x_t). \end{aligned}$$

By comparing both sides we get that $\widehat{f_\mu}(S) = 0$ for all S of even size. $\square$

Proof of Theorem 3.10. Let $G = (V, E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a labelling of V with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Expand

$$f(x_1, \dots, x_t) = \sum_{S \subseteq [t]} \widehat{f_\mu}(|S|)\chi_S^\mu(x_1, \dots, x_t).$$

By Corollary 3.7,

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=2}^{t} \widehat{f_\mu}(k)\nu^{k-1}\beta_k.$$

where $\nu = \sqrt{\frac{1+|\mu|}{1-|\mu|}}$. By Claim 3.11 we get that

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=1}^{\lfloor \frac{t-1}{2} \rfloor} \widehat{f_\mu}(2k+1)\nu^{2k}\beta_{2k+1}.$$

Applying Claim 3.9 and Lemma 3.8

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \sum_{k=1}^{\lfloor \frac{t-1}{2} \rfloor} \frac{1}{\sqrt{\binom{t}{2k+1}}}\nu^{2k}2^{2k+1}\binom{t-1}{k}\left(\frac{\lambda}{1-\lambda}\right)^{k+1}.$$

A straightforward calculation using Claim 2.1 implies that

$$\frac{1}{\sqrt{\binom{t}{2k+1}}}\binom{t-1}{k} \leqslant \frac{(6e)^k}{\sqrt{t}}.$$

Thus,

$$\begin{aligned}
\mathcal{E}_{G,\mathsf{val}}(f) &\leqslant \frac{1}{\sqrt{t}} \sum_{k=1}^{\lfloor \frac{t-1}{2} \rfloor} \nu^{2k}(6e)^k 2^{2k+1}\left(\frac{\lambda}{1-\lambda}\right)^{k+1} \\
&\leqslant \frac{\nu^{-2}}{\sqrt{t}} \sum_{k=1}^{\lfloor \frac{t-1}{2} \rfloor} \left(48e\nu^2\lambda\right)^{k+1} \\
&\leqslant \frac{\nu^{-2}}{\sqrt{t}} \sum_{k=2}^{\infty} \alpha^k \\
&\leqslant \frac{\alpha^2\nu^{-2}}{\sqrt{t}(1-\alpha)},
\end{aligned}$$

where $\alpha = 48e\nu^2\lambda$. Per our assumption that $\lambda < \frac{1-|\mu|}{192e}$ we have $\alpha \leqslant \frac{1}{2}$ and so

$$\mathcal{E}_{G,\mathsf{val}}(f) \leqslant \frac{2\alpha^2\nu^{-2}}{\sqrt{t}} \leqslant \frac{1}{\sqrt{t}} \cdot \frac{(96e)^2\lambda^2}{1-|\mu|}.$$

□

3.3.2 Bounds for weight indicators

For integers t and $w \in \{0, 1, \dots, t\}$ let $\mathbf{1}_w : \{\pm 1\}^t \to \{0, 1\}$ be the function indicating whether the weight of the input is w. That is, $\mathbf{1}_w(x_1, \dots, x_t) = 1$ if $|\{i \mid x_i = 1\}| = w$ and $\mathbf{1}_w(x_1, \dots, x_t) = 0$ otherwise. In this section we prove

Theorem 3.12. *For every $\mu \in (-1, 1)$, $0 < \lambda \leqslant \frac{1-|\mu|}{768e}$, every $t \in \mathbb{N}$ and $0 \leqslant w \leqslant t$, it holds that*

$$\mathcal{E}_\lambda(\mathbf{1}_w) \leqslant \frac{1}{\sqrt{t}} \cdot \frac{192e\lambda}{1-|\mu|}.$$

We analyze the weight indicator function in a similar way to the majority function, except that we need a new argument in order to bound its Fourier coefficients. The analysis is more delicate as the weight indicator function is not anti-symmetric and therefore has Fourier mass on even layers. For integers t and $w \in \{0, 1, \ldots, t\}$ let $\mathbf{1}_{>w} : \{\pm 1\}^{t+1} \to \{0, 1\}$ be the function indicating whether the weight of the input is greater w. That is, $\mathbf{1}_w(x_1, \ldots, x_t) = 1$ if $\sum_i x_i > w$ and $\mathbf{1}_w(x_1, \ldots, x_t) = 0$ otherwise. Note that,

$$\mathbf{1}_w(x_1, \ldots, x_t) = \mathbf{1}_{>w}(1, x_1, \ldots, x_t) - \mathbf{1}_{>w}(0, x_1, \ldots, x_t). \tag{3.9}$$

Claim 3.13. *For every $S \subseteq [t]$, it holds that*

$$\widehat{(\mathbf{1}_w)_\mu}(S) = \frac{\widehat{(\mathbf{1}_{>w})_\mu}(S \cup \{0\})}{\sqrt{1-\mu^2}}.$$

Proof.

$$\begin{aligned}
\mathbf{1}_w(x_1, \ldots, x_t) &= \mathbf{1}_{>w}(1, x_1, \ldots, x_t) - \mathbf{1}_{>w}(0, x_1, \ldots, x_t) \\
&= \sum_{S \subseteq \{0,\ldots,t\}} \widehat{(\mathbf{1}_{>w})_\mu}(S) \chi_S^\mu(1, x_1, \ldots, x_t) - \sum_{S \subseteq \{0,\ldots,t\}} \widehat{(\mathbf{1}_{>w})_\mu}(S) \chi_S^\mu(0, x_1, \ldots, x_t) \\
&= \sum_{S \subseteq \{0,\ldots,t\}} \widehat{(\mathbf{1}_{>w})_\mu}(S) (\chi_S^\mu(1, x_1, \ldots, x_t) - \chi_S^\mu(0, x_1, \ldots, x_t)) \\
& \sum_{\substack{S \subseteq \{0,\ldots,t\} \\ 0 \in S}} \widehat{(\mathbf{1}_{>w})_\mu}(S) \frac{1}{\sqrt{1-\mu^2}} \chi_{S \setminus \{0\}}^\mu(x_1, \ldots, x_t),
\end{aligned}$$

and the claim follows. $\square$

Claim 3.14. *For every $k \leqslant t$ it holds that*

$$\widehat{(\mathbf{1}_w)_\mu}(k)^2 \leqslant \frac{k+1}{(1-\mu^2)\binom{t}{k}(t+1)}.$$

Proof. As $\mathbf{1}_{>w}$ is symmetric with range $\{0, 1\}$,

$$\sum_{k=0}^{t+1} \binom{t+1}{k} \widehat{(\mathbf{1}_{>w})_\mu}^2(k) \leqslant 1.$$

In particular, for every $k \leqslant t+1$, $\widehat{(\mathbf{1}_{>w})_\mu}^2(k) \leqslant \binom{t+1}{k}^{-1}$. By Claim 3.13, for every $k \leqslant t$,

$$\widehat{(\mathbf{1}_w)_\mu}^2(k) = \frac{\widehat{(\mathbf{1}_{>w})_\mu}^2(k+1)}{1-\mu^2}.$$

Hence,

$$\widehat{(\mathbf{1}_w)_\mu}^2(k) \leqslant \frac{1}{(1-\mu^2)\binom{t+1}{k+1}},$$

which concludes the proof. $\square$

Proof of Theorem 3.12. Let $G = (V, E)$ be a regular λ-spectral expander, and $\mathsf{val} : V \to \{\pm 1\}$ a labelling of V with $\mathbf{E}[\mathsf{val}(V)] = \mu$. Expand

$$f(x_1, \ldots, x_t) = \sum_{S \subseteq [t]} \widehat{f_\mu}(|S|)\chi_S^\mu(x_1, \ldots, x_t).$$

By Corollary 3.7,

$$\mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) \leqslant \sum_{k=2}^{t} \widehat{(\mathbf{1}_w)_\mu}(k)\nu^{k-1}\beta_k.$$

where $\nu = \sqrt{\frac{1+|\mu|}{1-|\mu|}}$. Using Claim 3.14 to upper bound $\widehat{(\mathbf{1}_w)_\mu}$ and Lemma 3.8 to bound β_k we get,

$$\mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) \leqslant \sum_{k=2}^{t} \sqrt{\frac{k+1}{(1-\mu^2)(t+1)}} \frac{1}{\sqrt{\binom{t}{k}}} \cdot \nu^{k-1} 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil}.$$

As in the calculation in Theorem 3.10, it is not hard to verify that

$$\frac{1}{\sqrt{\binom{t}{k}}} \cdot \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \leqslant (3e)^{\frac{k}{2}}.$$

Thus,

$$\begin{aligned}
\mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) &\leqslant \sum_{k=2}^{t} \sqrt{\frac{k+1}{(1-\mu^2)(t+1)}} \cdot \nu^{k-1} (3e)^{\frac{k}{2}} 2^k \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil} \\
&\leqslant \frac{1}{\sqrt{t}} \cdot \frac{\nu^{-1}}{\sqrt{1-\mu^2}} \sum_{k=2}^{t} (\nu^2)^{\frac{k}{2}} (96e)^{\frac{k}{2}} \lambda^{\frac{k}{2}} \\
&= \frac{1}{\sqrt{t}} \cdot \frac{\nu^{-1}}{\sqrt{1-\mu^2}} \sum_{k=2}^{t} \alpha^k \\
&\leqslant \frac{1}{\sqrt{t}} \cdot \frac{\nu^{-1}}{\sqrt{1-\mu^2}} \cdot \frac{\alpha^2}{1-\alpha},
\end{aligned}$$

where $\alpha = \sqrt{96e\nu^2\lambda}$. Per our assumption that $\lambda < \frac{1-|\mu|}{768e}$ we have $\alpha \leqslant \frac{1}{2}$ and so

$$\mathcal{E}_{G,\mathsf{val}}(\mathbf{1}_w) \leqslant \frac{1}{\sqrt{t}} \cdot \frac{192e\lambda}{1-|\mu|}.$$

□

4 Lower Bounds

In this section we prove Theorems 1.2 and 1.3. In Section 4.1 we choose an expander graph for which we obtain a precise analytic formula for the expectation of characters under the input distribution given by the random walk. In Section 4.3 we use this formula to lower bound the bias of a certain symmetric function, thus proving Theorem 1.3. Furthermore, based on this formula, in Section 4.2 we prove Theorem 1.2.

4.1 Choosing the graph

Definition 4.1. *For $S \subseteq [t]$, we denote $\mathbf{\Delta}_{odd}(S) = \sum_{i=1}^{\lfloor (|S|-1)/2 \rfloor} \Delta_{2i+1}(S)$.*

Claim 4.2. *Let $G = ([n], E)$ be a Cayley graph over the boolean cube, with second largest eigenvalue λ_2 and corresponding eigenvector v_2. Define $\mathsf{val} : [n] \to \{\pm 1\}$ by $\mathsf{val}_2(i) = v_2(i)$. Let $S \subseteq [n]$, $|S| = k$. Let P be the diagonal matrix corresponding to val_2, that is, $P_{i,i} = \mathsf{val}_2(i) = v_2(i)$. Then,*

$$\left(\prod_{i=1}^{k-1} PG^{\Delta_i} \right) P\mathbb{1} = \begin{cases} \lambda^{\sum_{i=1}^{k-2/2} \Delta_{2i+1}} \mathbb{1} & k \in \mathbb{N}_{even}, \\ \lambda^{\sum_{i=1}^{k-1/2} \Delta_{2i+1}} v_2 & k \in \mathbb{N}_{odd}. \end{cases}$$

Proof. We will prove the claim using induction. For the base case $k = 1$ it holds that $\prod_{i=1}^{k-1} PG^{\Delta_i} = I$, and the statement follows as $IP\mathbb{1} = v_2 = \lambda^0 v_2$. For the induction step, note that

$$\left(\prod_{i=1}^{k} PG^{\Delta_i} \right) P\mathbb{1} = PG^{\Delta_k} \left(\prod_{i=1}^{k-1} PG^{\Delta_i} \right) P\mathbb{1}.$$

If $k \in \mathbb{N}_{even}$ than $k - 1 \in \mathbb{N}_{odd}$ and, using the induction hypothesis we get that

$$\begin{aligned} PG^{\Delta_k} \left(\prod_{i=1}^{k-1} PG^{\Delta_i} \right) P\mathbb{1} &= PG^{\Delta_k} \lambda^{\sum_{i=1}^{(k-2)/2} \Delta_{2i+1}} v_2 \\ &= \lambda^{\sum_{i=1}^{k/2} \Delta_{2i+1}} \mathbb{1}, \end{aligned}$$

which is what we wanted to prove. The proof in the case that $k \in \mathbb{N}_{odd}$ is similar. □

Corollary 4.3. *Let $G = ([n], E)$ be a Cayley graph over the boolean cube, with second largest eigenvalue λ_2 and corresponding eigenvector v_2. Define $\mathsf{val} : [n] \to \{\pm 1\}$ by $\mathsf{val}_2(i) = v_2(i)$. Then,*

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \begin{cases} \lambda^{\mathbf{\Delta}_{odd}(S)} & |S| \in \mathbb{N}_{even}, \\ 0 & |S| \in \mathbb{N}_{odd}. \end{cases}$$

Proof. Note that $Pv_2 = \mathbb{1}$ and $P\mathbb{1} = v_2$. As before, it holds that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^{k-1} PG^{\Delta_i} \right) P\mathbf{1} = \frac{1}{n} \mathbb{1}^\mathsf{T} \left(\prod_{i=1}^{k-1} PG^{\Delta_i} \right) P\mathbb{1}.$$

The fact that G is regular implies that $\mathbb{1}^\mathsf{T} v_2 = 0$, which finishes the case that k is odd; the case that k is even is handled similarly by noting that $\mathbb{1}^\mathsf{T}\mathbb{1} = n$. □

Cayley graphs over an Abelian groups commute and share an orthonormal basis of eigenvectors, which turns out to be the set of all characters of the group. The eigenvalues have a directe correspondence to the set of generators of the Caylely graph. Building on that [AR94] proved that for every $0 < \lambda < 1$, of the form $\frac{1}{m}$ for $m \in \mathbb{N}$, and $m \leqslant n \in \mathbb{N}$ there is a Cayley graph on the n dimensional boolean cube, with $\lambda_2 = \lambda$.

Throughout this section, we let G be a Cayley graph on the Boolean cube with $\lambda_2 = \lambda$ (which is given as parameter in the various statements of the results).

A technical tool that we use in Section 4.2 is the noise operator. The definitions and following claims appear in [O'D14].

Definition 4.4. *Let $\rho \in [-1, 1]$. For a fixed $x \in \{\pm 1\}^t$ we write $y \sim N_\rho(x)$ to denote the random string y that is drawn as follows: for each $i \in [t]$ independently,*

$$y_i = \begin{cases} x_i & \text{with probability } \frac{1+\rho}{2}, \\ -x_i & \text{with probability } \frac{1-\rho}{2}. \end{cases}$$

Definition 4.5. *Let $\rho \in [-1, 1]$. The noise operator T_ρ is the linear operator on functions $\{\pm 1\}^t \to \mathbb{R}$ defined by $T_\rho f(x) = \mathbf{E}_{y \sim N_\rho(x)} f(y)$ The fact that the operator is linear follows directly from the linearity of the expectation.*

Notice that $T_1(f) = f$ whereas $T_0(f)$ is the constant function $T_0(f) = \mathbf{E}\, f$. We make use of the following lemma.

Lemma 4.6. *For every function $f : \{\pm 1\}^t \to \mathbb{R}$ it holds that: $T_\rho f(x) = \sum_{S \subset [t]} \widehat{f}(S) \rho^{|S|} \chi_S(x)$.*

Claim 4.7. *For $f \colon \{\pm 1\}^t \to \mathbb{R}$, define $g \colon \{\pm 1\}^{2t} \to \mathbb{R}$ by*

$$g(x_1, x_2, \ldots, x_{2t-1}, x_{2t}) = f(x_1 \cdot x_2, \ldots, x_{2t-1} \cdot x_{2t}).$$

It holds that $\mathbf{E}[g(\mathrm{RW}_{G,\mathsf{val}_2})] = (T_\lambda f)(\mathbb{1})$.

Proof: For $\{s_1, \ldots, s_k\} = S \subseteq [t]$ denote $2S\colon = \{2s_1 - 1, 2s_1, \ldots, 2s_k - 1, 2s_k\} \subseteq [2 \cdot t]$. Note that $\mathbf{\Delta}_{odd}(2S) = |S|$. It is easy to verify that

$$g(x_1, x_2, \ldots, x_{2t-1}, x_{2t}) = \sum_{S \subseteq [t]} \widehat{f}(S) \chi_{2S}(x_1, x_2, \ldots, x_{2t-1}, x_{2t}).$$

Therefore,

$$\mathbf{E}[g(\mathrm{RW}_{G,\mathsf{val}_2})] = \sum_{S \subseteq [t]} \widehat{f}(S)\, \mathbf{E}[\chi_{2S}(\mathrm{RW}_{G,\mathsf{val}_2})] = \sum_{S \subseteq [t]} \widehat{f}(S) \lambda^{|S|} = \sum_{S \subseteq [t]} \widehat{f}(S) \lambda^{|S|} \chi_S(\mathbb{1}),$$

which is equal to $T_\lambda(f)(\mathbb{1})$ by Lemma 4.6. In the second equality we used Corollary 4.3. □

4.2 Lower bound for a function in $\mathsf{AC}(d)$

In this section, we prove Theorem 1.2, restated below. we continue with the choice of G and val as in the previous section.

Theorem 4.8 (Theorem 1.2; *restated*)**.** *There are universal constants $\varepsilon > 0$, $k \in \mathbb{N}$, satisfying the following. For every $3 \leqslant d \in \mathbb{N}$, there exists a constant t_d, and a family of functions $(h_t)_{t_d \leqslant t \in \mathbb{N}} \subset \mathsf{AC}(d)$, such that for every $\lambda \geqslant \frac{(40(d-2)k)^{d-2}}{\log^{d-2} t}$ there is a λ-spectral expander $G = (V, E)$, and a labelling $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = 0$ such that*

$$\mathcal{E}_{G,\mathsf{val}}(h_t) \geqslant \varepsilon.$$

As before, we will choose G to be a Cayley graph on the boolean cude with $\lambda_2 = \lambda$ and $\mathsf{val} = \mathsf{val}_2$. Our construction of the function f_d will be iterative, and we begin by presenting the basic building blocks used in it.

Fix t; we choose parameters r, h such that $r \cdot h \leqslant t$ by taking $h = \log(t) - \log\log(t)$ and $r = \lfloor \frac{t}{\log t} \ln(2) \rfloor$. Partition $[t]$ into disjoint sets $I_1, \ldots, I_r$, each of size h, and consider $f, g\colon \{-1,1\}^t \to \{0,1\}$ defined as

$$f(z_1, \ldots, z_t) = \bigvee_{i \in [r]} \bigwedge_{j \in I_i} z_j, \quad g(z_1, \ldots, z_t) = \bigwedge_{i \in [d]} \bigvee_{j \in I_i} z_j. \tag{4.1}$$

Here, -1 is interpreted as "true", 1 is interpreted as "false". Note that $f, g \in \mathsf{AC}(2)$.

Claim 4.9. *The functions f and g are almost balanced with respect to the uniform distribution. Quantitatively,* $\mathbf{E}[f], \mathbf{E}[g] \in \left[\frac{1}{2} - O\left(\frac{\log(t)}{t}\right), \frac{1}{2} + O\left(\frac{\log(t)}{t}\right)\right]$.

Proof. From De Morgan's identity we have $g(x_1, \ldots, x_t) = 1 - f(\overline{x_1}, \ldots, \overline{x_t})$, so $\mathbf{E}[g] = 1 - \mathbf{E}[f]$ and it is enough to prove the statement for f. To see that, we first write

$$\mathbf{E}[f] = \mathbf{Pr}[f = 1] = 1 - \prod_{i=1}^{r} \mathbf{Pr}[\bigwedge_{j \in I_i} z_j = 0] = 1 - \prod_{i=1}^{r} \left(1 - \mathbf{Pr}[\bigwedge_{j \in I_i} z_j = 1]\right) = 1 - \left(1 - \frac{1}{2^h}\right)^r.$$

Using the fact that $1 - \varepsilon = e^{-\varepsilon + O(\varepsilon^2)}$ we obtain that

$$1 - \left(1 - \frac{1}{2^h}\right)^r = 1 - e^{-2^{-h} r + O(r 2^{-2h})} = 1 - e^{-\ln 2 + O\left(\frac{\log t}{t}\right)} = \frac{1}{2} + \Theta\left(\frac{\log t}{t}\right),$$

as desired. $\square$

Denote by μ_p the product distribution over $\{\pm 1\}^t$, wherein for each i we have that $\mathbf{Pr}[z_i = -1] = p$. Abusing notations, we denote $\mu_(f) = \mathbf{E}_{x \sim \mu_p}[f(x)]$.

Claim 4.10. *There is sufficiently large constant $k > 0$ such that the following holds. Suppose $p = \frac{1-\varepsilon}{2}$ is such that $\frac{k^2}{t} \leqslant \varepsilon \leqslant \frac{k}{\log(t)}$. Then,*

$$\mu_p(f), \mu_p(g) \leqslant \frac{1 - \frac{h\varepsilon}{20k}}{2}.$$

Proof. First, we analyze $\mu_p(f)$. By definition, it is equal to

$$\mathop{\mathbf{Pr}}_{\mu_p}[f = 1] \leqslant 1 - (1 - p^h)^r = 1 - (1 - 2^{-h}(1 - \varepsilon)^h)^r = 1 - e^{-r \cdot 2^{-h}(1-\varepsilon)^h + O(2^{-2h} r (1-\varepsilon)^{2h})}$$

Using $(1 - \varepsilon)^h \leqslant 1 - \frac{\varepsilon h}{10k}$ that holds as $\varepsilon h \leqslant k$, we may upper bound the previous expression by

$$1 - e^{-r \cdot 2^{-h}\left(1 - \frac{h\varepsilon}{10k}\right) + O(2^{-2h} r (1-\varepsilon)^{2h})} \leqslant 1 - e^{-\ln 2 + \frac{h\varepsilon \ln 2}{10k} + O\left(\frac{\log t}{t}\right)} = 1 - \frac{1}{2} e^{\frac{h\varepsilon \ln 2}{10k} + O\left(\frac{\log t}{t}\right)}.$$

which is at most $\frac{1}{2} - \frac{h\varepsilon}{40k}$ provided that k is sufficiently large. Here, we used the bound $e^z \geqslant 1 + z$. Now we analyze $\mu_p(g)$. By definition, it is equal to

$$\mathop{\mathbf{Pr}}_{\mu_p}[g = 1] = (1 - (1 - p)^h)^r = (1 - 2^{-h}(1 + \varepsilon)^h)^r \leqslant e^{-r \cdot 2^{-h}(1+\varepsilon)^h + O(2^{-2h} r (1+\varepsilon)^{2h})}.$$

Using the fact that $(1+\varepsilon)^h \leqslant 1+h\varepsilon$, we may further upper bound this by

$$\leqslant e^{-r\cdot 2^{-h}(1+h\cdot\varepsilon)+O\left(\frac{\log t}{t}\right)} \leqslant e^{-\ln 2-h\varepsilon\ln 2+O\left(\frac{\log t}{t}\right)} \leqslant \frac{1}{2}e^{-h\varepsilon\ln 2+O(\frac{\log t}{t})}$$

Using $e^{-z} \leqslant 1-\frac{z}{10k}$ that holds for $0 \leqslant z \leqslant k$, we may upper bound this by

$$\frac{1}{2}\left(1-\frac{h\varepsilon\ln 2}{20k}+O\left(\frac{\log t}{t}\right)\right) \leqslant \frac{1}{2}-\frac{h\varepsilon}{40k},$$

where the last inequality holds provided that k is sufficiently large. ☐

Claim 4.11. *Let $p=\frac{1-\varepsilon}{2}$, then if $\varepsilon \geqslant \frac{k}{\log(t)}$. Then,*

$$\mu_p(f), \mu_p(g) \leqslant e^{-k/10}.$$

Proof. First, we analyze $\mu_p(f)$. By definition it is equal to

$$\begin{aligned}\Pr_{\mu_p}[f=1] = 1-(1-p^h)^r = 1-(1-2^{-h}(1-\varepsilon)^h)^r &= 1-\left(1-2^{-h}\left(1-\frac{k}{\log t}\right)^h\right)^r \\ &\leqslant 1-\left(1-2^{-h}e^{-k}\right)^r.\end{aligned}$$

Using $(1-\delta)^r \geqslant 1-r\delta$, we get that the above expression is at most $r2^{-h}e^{-k} \leqslant e^{-k}$. Next, we upper bound $\mu_p(g)$. By definition, it is equal to

$$\Pr_{\mu_p}[g=1] \leqslant (1-(1-p)^h)^r = (1-2^{-h}(1+\varepsilon)^h)^r = \left(1-2^{-h}\left(1+\frac{k}{\log t}\right)^h\right)^r.$$

Using $(1+\delta)^r \geqslant \delta r$ for $\delta > 0$, we get that this is at most

$$(1-2^{-h}k)^r \leqslant e^{-r2^{-h}k} \leqslant e^{-k/10}.$$

☐

Claim 4.12. *Let $p=\frac{1-\varepsilon}{2}$. There exists $k \in \mathbb{N}$ such that for $\frac{k^2}{t} \leqslant \varepsilon \leqslant \frac{1}{\sqrt{t}}$, then,*

$$\mu_p(f), \mu_p(g) \geqslant \frac{1-20h\cdot\varepsilon}{2}.$$

Proof. First, we analyze $\mu_p(f)$. By definition, it is equal to

$$\Pr_{\mu_p}[f=1] = 1-(1-p^h)^r = 1-(1-2^{-h}(1-\varepsilon)^h)^r = 1-e^{-r\cdot 2^{-h}(1-\varepsilon)^h+O(2^{-2h}r(1-\varepsilon)^{2h})}.$$

Using $(1-\varepsilon)^h \geqslant 1-\varepsilon h$, we may lower bound the previous expression by

$$1-e^{-r\cdot 2^{-h}(1-h\varepsilon)+O(2^{-2h}r(1-\varepsilon)^{2h})} \geqslant 1-e^{-\ln 2+h\varepsilon\ln 2+O\left(\frac{\log t}{t}\right)} = 1-\frac{1}{2}e^{h\varepsilon\ln 2+O\left(\frac{\log t}{t}\right)}.$$

which is at least $\frac{1}{2} - 20h\varepsilon$. Here, we used the bound $e^z \leqslant 1 + 2z$, provided $0 \leqslant z \leqslant 1$. Now we analyze $\mu_p(g)$. By definition, it is equal to

$$\Pr_{\mu_p}[g = 1] = (1 - (1-p)^h)^r = (1 - 2^{-h}(1+\varepsilon)^h)^r \geqslant e^{-r\cdot 2^{-h}(1+\varepsilon)^h + O(2^{-2h}r(1+\varepsilon)^{2h})}.$$

Using the fact that $(1+\varepsilon)^h \leqslant 1 + 10 \cdot h\varepsilon$ as $h\varepsilon \leqslant 1$, we may lower bound this by

$$\geqslant e^{-r\cdot 2^{-h}(1+10h\cdot\varepsilon)+O\left(\frac{\log t}{t}\right)} = e^{-\ln 2 - 10h\cdot\varepsilon\ln 2 + O\left(\frac{\log t}{t}\right)} \geqslant \frac{1}{2}e^{-h\cdot\varepsilon\cdot 10\ln 2 + O\left(\frac{\log t}{t}\right)}$$

Using $e^{-z} \geqslant 1 - z$, we may lower bound this by

$$\frac{1}{2}\left(1 - 10\ln 2h\cdot\varepsilon + O\left(\frac{\log t}{t}\right)\right) \geqslant \frac{1 - 20h\cdot\varepsilon}{2},$$

□

Remark 4.13. *Notice that when we take $h(x_1, y_1, \dots, x_t, y_t) = f(x_1 \oplus y_1, \dots, x_t \oplus y_t)$, we can use Claim 4.7 to obtain that $\mathbf{E}[h(\mathrm{RW}_{G,\mathsf{val}_2})] = T_\lambda(f)(\mathbb{1}) = \mu_{\frac{1-\lambda}{2}}(f)$ and thus, if $\lambda \geqslant \frac{k}{\log t}$ we get that $\mathbf{E}[h(\mathrm{RW}_{G,\mathsf{val}_2})] < e^{-C}$ while $\mathbf{E}[h] = \mathbf{E}[f] \geqslant \frac{1}{2} - O(\frac{\log t}{t})$. This implies that h distinguishes between the two distributions.*

We want to generalize this result for $\mathsf{AC}(d)$. In order to do so, we define a sequence of functions $\{h_d\}_{d\in\mathbb{N}}$ as follows. The function h_1 is defined to be g; the function h_2 operates on t^2 coordinates, denote them by $x^1, \dots, x^t \in \{\pm 1\}^t$, and is defined as $h_2 = g(f(x^1), \dots, f(x^t))$. Iteratively, once h_d has been defined, we view the input to $h_{d+1} \colon \{\pm 1\}^{t^{d+1}} \to \{0,1\}$ as $y^1, \dots, y^t \in \{\pm 1\}^{t^d}$ and define

$$h_{d+1}(y^1, \dots, y^t) = \begin{cases} h_d(g(y^1), \dots, g(y^t)) & 1 < d \in \mathbb{N}_{even} \\ h_d(f(y^1), \dots, f(y^t)) & 2 < d \in \mathbb{N}_{odd} \end{cases}$$

Observe that $h_d \in \mathsf{AC}(d+1)$.

Claim 4.14. *There is $c \in \mathbb{N}$. Let $p = \frac{1-\varepsilon}{2}$. such that for $\frac{c^2}{t} \leqslant \varepsilon \leqslant \frac{1}{20^d \log^d t\sqrt{t}}$, then,*

$$\mu_p(h_d) \geqslant \frac{1 - (20h)^d\varepsilon}{2}.$$

Proof. Using induction. When $d = 1$ the claim follows from Claim 4.12. Assume that the claim holds for $d' < d$, and that $d \in \mathbb{N}_{even}$.

$$\mu_p(h_d) = \mu_{\mu_p(f)}(h_{d-1}) \geqslant \mu_{\frac{1-20h\varepsilon}{2}}(h_{d-1}).$$

Where the last transition holds as h_{d-1} is monotone, and $\mu_p(f) \geqslant \frac{1-20h\varepsilon}{2}$ by Claim 4.12.

As $\varepsilon \leqslant \frac{1}{20^d \log^d t\sqrt{t}}$, using the induction hypothesis, it holds that $\mu_p(h_d) \geqslant \mu_{\frac{1-20h\varepsilon}{2}}(h_{d-1}) \geqslant \frac{1-(20h)^d\varepsilon}{2}$. When $d \in \mathbb{N}_{odd}$ the analysis is analogous. □

Corollary 4.15. *For every $\delta > 0$, $d \in \mathbb{N}$ there is $t_d \in \mathbb{N}$ such that for every $t_d \leqslant t \in \mathbb{N}$.*

$$\mathbf{E}[h_d] \geqslant \frac{1}{2} - \delta.$$

Proof. This follows immediately from Claim 4.14 and Claim 4.9. $\square$

Claim 4.16. *There is sufficiently large constant $k > 0$ such that the following holds. If $\frac{k^2}{t} \leqslant \lambda \leqslant \frac{k}{\log t}$ then, $\mu_{\frac{1-\lambda}{2}}(h_d) \leqslant \mu_{\frac{1-(h/20k)\cdot\lambda}{2}}(h_{d-1})$.*

Proof. Denote $p = \frac{1-\lambda}{2}$. If $2 < d \in \mathbb{N}_{odd}$, then $\mu_p(h_d) = \mu_p(h_{d-1}(g_1, \dots, g_t)) = \mu_q(h_{d-1}(x_1, \dots, x_t))$ Where $q = \mathbf{Pr}_{\mu_p}[g = 1]$. Using Claim 4.10, we get that $q \leqslant \frac{1-(h/20k)\cdot\lambda}{2}$ and from monotonicity we get that $\mu_{\frac{1-\lambda}{2}}(h_d) \leqslant \mu_{\frac{1-(h/20k)\cdot\lambda}{2}}(h_{d-1})$. When $2 < d \in \mathbb{N}_{even}$ the proof is identical. $\square$

Claim 4.17. *Let $p = \frac{1-\varepsilon}{2}$, then if $\varepsilon \geqslant \frac{k}{\log(t)}$. Then, for every $d \in \mathbb{N}$,*

$$\mu_p(h_d) \leqslant e^{-k/10}.$$

Proof. Using induction. $d = 1$ implies $h_r = g$ and the claim follows from Claim 4.11. Assuming the claim holds for $d' \leqslant d$, then, if $d + 1 \in \mathbb{N}_{odd}$

$$\mu_p(h_{d+1}) = \mu_p(h_d(g_1, \dots, g_t)) = \mu_{\mu_p(g)}[h_d(x_1, \dots, x_t)]$$

From Claim 4.11 we know that $m_p(g) \leqslant e^{-k/10}$, which means that $\varepsilon = 1-2p \geqslant 1-2e^{-k/10} \geqslant \frac{k}{\log t}$, and thus we can use the induction hypothesis and conclude that

$$\mu_p(h_{d+1}) \leqslant e^{-k}.$$

The induction step when $r \in \mathbb{N}_{even}$ is identical. $\square$

Claim 4.18. *For all $k \in \mathbb{N}$, setting $C_d = k(20k)^d$, if $\lambda \geqslant \frac{C_d}{\log^d t}$ then $\mu_{\frac{1-\lambda}{2}}(h_d) \leqslant e^{-k/10}$.*

Proof. Using induction, if $\lambda \geqslant \frac{k}{\log t}$, the claim follows from Claim 4.17.

For $0 \leqslant j < d$, denote $\lambda_j = (\frac{h}{20k})^j\lambda$. While $\lambda_j \leqslant \frac{k}{\log t}$, we can use Claim 4.16, to obtain that $\mu_{\frac{1-\lambda_j}{2}}(h_{d-j}) \leqslant \mu_{\frac{1-\lambda_{j+1}}{2}}(h_{d-j-1})$.

The assumption that $\lambda > \frac{C_d}{\log^d t}$ implies that $\lambda_{d-1} = (h/20k)^{d-1}\lambda \geqslant \frac{\log^{d-1} t}{(20k)^{d-1}}\frac{C_d}{\log^d t} \geqslant k/\log t$. Thus, we may consider the minimal j such that $\lambda_j \geqslant k/\log t$ and get that

$$\mu_{\frac{1-\lambda}{2}}(h_d) \leqslant \mu_{\frac{1-\lambda_j}{2}}(h_{d-j}) \leqslant e^{-k/10},$$

where in the last inequality we used Claim 4.17. $\square$

Proof of Theorem 4.8. Let $C'_d = (2d)^{d+1}C_d$. As before consider, $\tilde{h}_d(x_1, y_1, \dots, x_t, y_t) = h_d(x_1 \oplus y_1, \dots, x_t \oplus y_t)$. Note that $\tilde{h}_d\colon \{\pm 1\}^{t'_d} \to \{0, 1\}$ where $t'_d = 2t^d$. Thus $\log(t'_d)/2d \leqslant \log(t)$, and $\frac{C_d}{\log^d t} \leqslant \frac{C'_d}{\log^d t'_d}$ It holds that $\tilde{h}_d \in \mathsf{AC}(d+2)$. Using Claim 4.7 we obtain that $\mathbf{E}[\tilde{h}_d(\mathrm{RW}_{G,\mathsf{val}_2})] = T_\lambda(h_d)(\mathbb{1}) = \mu_{\frac{1-\lambda}{2}}(h_d)$. From Claim 4.18 we obtain that if $\lambda \geqslant \frac{C'_d}{\log^d t'_d} \geqslant \frac{C_d}{\log^d t}$ then

$$\mathbf{E}[\tilde{h}_r(\mathrm{RW}_{G,\mathsf{val}_2})] = \mu_{\frac{1-\lambda}{2}}(h) \leqslant e^{-k/10}.$$

Corollary 4.15 implies that

$$\mathbf{E}[\tilde{h}_d(\mathrm{Ind}_{2t})] = \mathbf{E}[h_d(\mathrm{Ind}_t)] = \mu_{\frac{1}{2}}(h_d) \geqslant \frac{1}{2} - o(1).$$

This implies that $\tilde{h}_d$ distinguish between the distribution obtained by a random walk on G with $\lambda \geqslant \frac{C'_d}{\log^d t'_d}$ and t'_d independent bits, as we wanted to show. $\square$

4.3 Lower bound for symmetric functions

The main goal of this section is to prove the following theorem.

Theorem (Theorem 1.3, restated)**.** *There exists a universal constant c, and a family of symmetric functions $(f_t)_{t\in\mathbb{N}}$ where $f_t : \{\pm 1\}^t \to \{\pm 1\}$ satisfying the following. For every λ there is a λ-spectral expander $G = (V, E)$, and a labelling $\mathsf{val} : V \to \{\pm 1\}$ with $\mathbf{E}[\mathsf{val}(V)] = 0$ such that for every t, $\mathcal{E}_{G,\mathsf{val}}(f_t) \geqslant c \cdot \lambda$.*

Theorem 4.19. *For all $0 < c_0 \leqslant 1$, if $0 < \lambda < \frac{c_0^2}{12800\cdot e}$ and $f\colon \{\pm 1\}^t \to \{\pm 1\}$ is a symmetric function with $\left|\widehat{f}(2)\right| \geqslant \frac{c_0}{\sqrt{\binom{t}{2}}}$, then*

$$\mathcal{E}_{G,\mathsf{val}_2}(f) \geqslant 0.001 c_0 \lambda.$$

Proof. Denote by $\mathcal{B}_2 = \{\{i, i+1\} \mid i \in [t-1]\}$. Note that $|\mathcal{B}_2| = t-1$ and that for every $S \in \mathcal{B}_2$ it holds that $\mathbf{\Delta}_{odd}(S) = 1$.

$$\begin{aligned}
\mathcal{E}_{G,\mathsf{val}_2}(f) &= \left| \sum_{\substack{S\subseteq[t]\\|S|\geqslant 2}} \widehat{f}(|S|)\,\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \right| \\
&\geqslant \left|\widehat{f}(2)\right| \left| \sum_{S\subseteq[t],|S|=2} \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \right| - \left| \sum_{S\subseteq[t],|S|>2} \widehat{f}(|S|)\mathcal{E}_{G,\mathsf{val}_2}(\chi_S) \right|
\end{aligned}$$

It holds that

$$\begin{aligned}
\left|\widehat{f}(2)\right| \left| \sum_{S\subseteq[t],|S|=2} \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \right| &\geqslant \left|\widehat{f}(2)\right| \sum_{S\subseteq\mathcal{B}_2} \lambda \\
&\geqslant c_0\sqrt{2}\sqrt{\frac{(t-1)}{t}}\lambda \\
&\geqslant \frac{c_0}{\sqrt{2}}\lambda,
\end{aligned}$$

and,

$$\left| \sum_{S\subseteq[t],|S|>2} \widehat{f}(S)\mathcal{E}_{G,\mathsf{val}_2}(\chi_S) \right| \leqslant \sum_{k\geqslant 3} \left|\widehat{f}(k)\right| \beta_k \leqslant \sum_{k\geqslant 3} \frac{1}{\sqrt{\binom{t}{k}}} 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil},$$

where in the last inequality we used Lemma 3.8 and Claim 3.9. Simplifying, this is bounded by

$$\sum_{k\geqslant 3} (16e)^{k/2}\lambda^{k/2} \leqslant 124\lambda^{1.5}.$$

We omit the calculations (a similar calculation appears in Theorem 3.6). Assume that $\lambda \leqslant \frac{c_0^2}{128e\cdot 100}$. Overall, we get

$$\mathcal{E}_{G,\mathsf{val}_2}(f) \geqslant \frac{c_0}{\sqrt{2}}\lambda - 124\lambda^{1.5} \geqslant 0.04 c_0 \lambda. \qquad \square$$

We are now ready to prove Theorem 4.19.

Proof of Theorem 4.19. We take $f = \mathbf{1}_{>w}$ for $w = \frac{t-\sqrt{t}}{2}$, and appeal to Theorem 4.19. Thus, it suffices to show that $\left|\widehat{f}(2)\right| > \frac{c_0}{\sqrt{\binom{t}{2}}}$, for some absolute constant $c_0 > 0$. Indeed, by Claim 3.13 we have $\widehat{f}(2) = \widehat{\mathbf{1}_w}(1)$ for $\mathbf{1}_w : \{\pm\}^{t-1} \to \{0,1\}$. To compute the last Fourier coefficient we appeal to [CPTS20, Claim 4.9] for $w = \frac{t-\sqrt{t}}{2}$ and get

$$\left|\widehat{\mathbf{1}_w}(1)\right| = \left|\frac{1}{2^{t-1}}\frac{\binom{t-1}{w}}{\binom{t-1}{1}}\sum_{\ell=0}^{\lfloor\frac{1}{2}\rfloor}(-1)^{1-\ell}\binom{w}{\ell}\binom{t-2w-1}{1-2\ell}\right| = \frac{1}{2^{t-1}}\frac{\binom{t-1}{w}}{t-1}(t-1-2w).$$

Substituting $w = \frac{t-\sqrt{t}}{2}$, together with the fact that $\binom{t-1}{\frac{t-\sqrt{t}}{2}} \geqslant \Omega\left(\frac{1}{\sqrt{t}}2^t\right)$, finishes the proof. $\square$

References

[AC88] Noga Alon and Fan RK Chung. Explicit construction of linear sized tolerant networks. *Discrete Mathematics*, 72(1-3):15–19, 1988.

[AEL95] Noga Alon, Jeff Edmonds, and Michael Luby. Linear time erasure codes with nearly optimal recovery. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 512–519. IEEE, 1995.

[AKS87] Miklós Ajtai, János Komlós, and Endre Szemerédi. Deterministic simulation in logspace. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, pages 132–140, 1987.

[AR94] Noga Alon and Yuval Roichman. Random cayley graphs and expanders. *Random Struct. Algorithms*, 5(2):271–285, 1994.

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

[BCG20] Mark Braverman, Gil Cohen, and Sumegha Garg. Pseudorandom pseudo-distributions with near-optimal error for read-once branching programs. *SIAM Journal on Computing*, 49(5):STOC18–242–STOC18–299, 2020.

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006.

[Bra10] Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5):Art. 28, 10, 2010.

[CPTS20] Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: A fourier-analytic approach. In *Electron. Colloquium Comput. Complex*, volume 27, page 6, 2020.

[CW89] Aviad Cohen and Avi Wigderson. Dispersers, deterministic amplification, and weak random sources. In *30th Annual Symposium on Foundations of Computer Science*, pages 14–19. IEEE Computer Society, 1989.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007.

[Gil98] David Gillman. A chernoff bound for random walks on expander graphs. *SIAM Journal on Computing*, 27(4):1203–1220, 1998.

[GK21] Venkatesan Guruswami and Vinayak M Kumar. Pseudobinomiality of the sticky random walk. In *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2021.

[Hea08] Alexander D Healy. Randomness-efficient sampling within nc. *Computational Complexity*, 17(1):3–37, 2008.

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006.

[INW94] Russell Impagliazzo, Noam Nisan, and Avi Wigderson. Pseudorandomness for network algorithms. In *Proceedings of the twenty-sixth annual ACM symposium on Theory of computing*, pages 356–364, 1994.

[IZ89] Russell Impagliazzo and David Zuckerman. How to recycle random bits. In *FOCS*, volume 89, pages 248–253, 1989.

[Klo17] Benoît Kloeckner. Effective limit theorems for Markov chains with a spectral gap. *arXiv preprint arXiv:1703.09623*, 2017.

[KMRZS17] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally correctable and locally testable codes with sub-polynomial query complexity. *J. ACM*, 64(2):Art. 11, 42, 2017.

[KV86] Claude Kipnis and SR Srinivasa Varadhan. Central limit theorem for additive functionals of reversible markov processes and applications to simple exclusions. *Communications in Mathematical Physics*, 104(1):1–19, 1986.

[Lez01] Pascal Lezaud. Chernoff and Berry-Esséen inequalities for markov processes. *ESAIM: Probability and Statistics*, 5:183–201, 2001.

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.

[Mar88] Grigorii Aleksandrovich Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problemy peredachi informatsii*, 24(1):51–60, 1988.

[MOP20] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-ramanujan graphs of every degree. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, pages 510–523, 2020.

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014.

[Rei05] Omer Reingold. Undirected ST-connectivity in log-space. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 376–385. ACM, New York, 2005.

[RRV99] Ran Raz, Omer Reingold, and Salil Vadhan. Error reduction for extractors. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 191–201. IEEE, 1999.

[RV05] Eyal Rozenman and Salil Vadhan. Derandomized squaring of graphs. In *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 436–447. Springer, 2005.

[RVW00] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders and extractors. In *Proceedings 41st Annual Symposium on Foundations of Computer Science*, pages 3–13. IEEE, 2000.

[SS96] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[Tal17] Avishay Tal. Tight bounds on the fourier spectrum of ac0. In *32nd Computational Complexity Conference (CCC 2017)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017.

[Tre17] Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley, https://people. eecs. berkeley. edu/˜ luca/books/expanders-2016. pdf*, 2017.

[TS17] Amnon Ta-Shma. Explicit, almost optimal, epsilon-balanced codes. In *STOC'17—Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 238–251. ACM, New York, 2017.

[Vad12] Salil P Vadhan. *Pseudorandomness*, volume 7. Now, 2012.

[Val76] Leslie G. Valiant. Graph-theoretic properties in computational complexity. *J. Comput. System Sci.*, 13(3):278–285, 1976.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il