Expander Random Walks: The General Case and Limitations

Gil Cohen ✉
Department of Computer Science, Tel Aviv University, Israel

Dor Minzer ✉
Department of Mathematics, Massachusetts Institute of Technology, Cmabridge, MA, USA

Shir Peleg ✉
Department of Computer Science, Tel Aviv University, Israel

Aaron Potechin ✉
Department of Computer Science, University of Chicago, IL, USA

Amnon Ta-Shma ✉
Department of Computer Science, Tel Aviv University, Israel

Abstract

Cohen, Peri and Ta-Shma [11] 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? [11] considered only balanced labellings, and proved that for all symmetric functions the distinguishability goes down to zero with the spectral gap λ of the expander G. In addition, [11] show that functions computable by AC^0 circuits are fooled by expanders with vanishing spectral expansion.

We continue the study of this question. We generalize the result to all labelling, not merely balanced ones. We also improve the upper bound on the error of symmetric functions. More importantly, we give a matching lower bound and show a symmetric function with distinguishability going down to zero with λ but not with t. Moreover, we prove a lower bound on the error of functions in AC^0 in particular, we prove that a random walk on expanders with constant spectral gap does not fool AC^0.

2012 ACM Subject Classification Theory of computation → Random walks and Markov chains

Keywords and phrases Expander Graphs, Random Walks, Lower Bounds

Digital Object Identifier 10.4230/LIPIcs.ICALP.2022.43

Category Track A: Algorithms, Complexity and Games

Related Version *Full Version*: https://eccc.weizmann.ac.il/report/2021/091/

Funding *Gil Cohen*: Supported by ERC starting grant 949499 and by the Israel Science Foundation grant 1569/18.
Shir Peleg: 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.
Amnon Ta-Shma: The research leading to these results has received funding from the Israel Science Foundation (grant number 952/18)

Acknowledgements The authors thank Amir Yehudayoff for pointing out that the analysis of the error of the parity function is tight if the second eigenvector of the graph is Boolean. This remark, several years later, matured to the current paper. We thank Venkat Guruswami and Vinayak Kumar for discussing their results with us. We thank Ron Peled for discussions on the CLT in TVD, and for pointing us to results on local convergence for independent processes. We thank Oded Goldreich for the vanishing with t and λ notation.

49th International Colloquium on Automata, Languages, and Programming (ICALP 2022).
Editors: Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff;
Article No. 43; pp. 43:1–43:18
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

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 where an undirected graph $G = (V, E)$ is a λ-*spectral expander* if the second largest eigenvalue of its normalized adjacency matrix M is bounded above by λ. For simplicity, we only consider regular graphs, in which case M is also the random walk matrix of G. Expander graphs are among the most useful combinatorial objects in theoretical computer science, pivotal in derandomization [18, 29], complexity theory [37, 1, 12] and coding theory [32, 22, 33, 13] to name a few. Many works in the literature have studied explicit constructions of expander graphs (see, e.g., [24, 25, 7, 30, 6, 26]) and utilized their pseudorandom properties. We refer the reader to the excellent expositions [17, 35] and to Chapter 4 of [36].

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, i.e., the $n \times n$ matrix with all entries equal to $\frac{1}{n}$. One can express the normalized adjacency matrix M of G as $M = (1-\lambda)\mathbf{J} + \lambda E$ 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 [2]. This idea also appears in many derandomization results, [18, 3, 28, 29, 31, 9].

A useful generalization of the above 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. Next, we elaborate on the hitting property of expanders [1, 10, 19, 5] as well as the expander Chernoff bound [1, 10, 19, 14, 16].

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 [33] where the result appears).

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 [15, 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 forefront a natural question that was recently raised by [11] (see also the work of Guruswami and Kumar [15] 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 $i = 2, 3, \ldots, t$, 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 [11] focuses on the case $\mu = 0$. One result shows that

$$\mathcal{E}_{\lambda,0}(\mathsf{MAJ}) \leqslant O\left(\frac{\lambda^2}{\sqrt{t}}\right) \tag{1.1}$$

Their main result states that for each balanced labelling, for every symmetric function $f : \{\pm 1\}^t \to \{\pm 1\}$,

$$\mathcal{E}_{\lambda,0}(f) = O(\lambda \cdot \log^{3/2}(1/\lambda)). \tag{1.2}$$

This readily implies, for the specific case of balanced labelling, a central limit theorem with respect to the total variation distance, that vanishes as $\lambda \to 0$, thus strengthens previous results that considered the Kolmogorov distance [20, 23, 21] instead of the total variation distance.

To summarize the state of knowledge so far:

- Every symmetric function is fooled with error probability going down to zero with the spectral gap λ (see Equation (1.2)), where $\mu = 0$.
- The MAJ function is fooled with error probability going down to zero with t even when λ is fixed (see Equation (1.1)); and,
- The PARITY, AND, OR functions are fooled with error probability going down to zero *exponentially* fast with t even when λ is fixed.

Accordingly, let us say an error function vanishes with λ, if the error function is vanishing as $\lambda \to 0$. Similarly, we say an error function vanishes with t, if for some fixed $\lambda \geqslant 0$, it is going down to zero together with t.

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

[11] 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

$$\mathcal{E}_{\lambda,0}(f) = O(\sqrt{\lambda} \cdot (\log s)^{2(d-1)}). \tag{1.3}$$

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 small. This result can be thought of as an analog of Braverman's celebrated result [8] (see also [34]) that studies the pseudorandomness of k-wise independent distributions with respect to AC^0 test functions. However, for it to be meaningful, the spectral gap λ should be small.

1.1 Our contribution

The work of [11] leaves several open problems. First, and foremost, while [11] show the error function of any symmetric function vanishes with λ, it leaves open the possibility that a better convegence exists and, perhaps, the error function of any symmetric function vanishes with t, i.e., for some fixed λ, the error function goes down to zero together with the walk length t. Indeed, this is the case with the AND, OR and PARITY functions, where the error vanishes exponentially fast with t, and the MAJ function where the error goes down polynomially in t (see Equation (1.1)). Similarly, one may ask whether the error of AC^0 functions decays faster than Equation (1.3) and allows for larger spectral gaps λ then dictated by the above bound.

Our first result is that there exists a symmetric function for which the error function does not vanish with t:

▶ **Theorem 1.** *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, $\mathcal{E}_{G,\mathsf{val}}(f_t) = \Omega(\lambda)$.*

To explain how we obtain such a lower bound on a function f, we first review how [11] obtained their upper bound. The key idea in [11] is to expand the test function f under consideration in the Fourier basis. The question of fooling general test functions then reduces to the study of test functions that are Fourier characters. Now, let G denote the adjacency matrix of the graph (i.e, $M = \frac{1}{d}G$). Also, for a labelling $\ell : V \to \{\pm 1\}$ let us denote by P the diagonal matrix with $\ell(i)$ in the i'th element on the diagonal. One can check that for the parity function $\chi_{[t]} : \{\pm 1\}^t \to \{\pm 1\}$, $\chi_{[t]}(x) = \prod_{i=1}^t x_i$, we get $\mathbf{E}[\chi_{[t]}(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^t PG\right) \mathbf{1}$,, where $\mathbf{1} = (\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}})$.

In general,

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T \left(\prod_{i=1}^t P^{\delta_S(i)} G\right) \mathbf{1}, \tag{1.4}$$

where $\delta_S(i)$ is 1 if $i \in S$ an 0 otherwise. In [11] it is shown how to upper bound this expression for any λ-expander G and μ-biased function val.

For the proof of Theorem 1 we choose a λ-expander G and a labelling function val such that we can *exactly* express Equation (1.4) in terms of λ, S. To do so, we first choose G to be a Cayley graph over an Abelian group, and we use the fact that the eigenvectors of such a graph correspond to the characteristic functions of the underlying group, regardless of the set of generators used. One disadvantage in choosing a Cayley graph over an Abelian group is that it cannot give constant degree exapnders, though this is not a concern to us because with logarithmic degree we can have vanishing second eigenvalue. Next, we choose the underlying group to be $\mathbb{Z}_2^n$. This guarantees that the characteristic functions of $\mathbb{Z}_2^n$, and therefore also all the entries in all eigenvectors, are Boolean, i.e., either 1 or -1. Finally, we choose the labelling function val to correspond to the entries of the eigenvalue with the second largest eigenvalue.

The above choices guarantee that $P\mathbf{1} = v_2$ and $Pv_2 = \mathbf{1}$ (because $P^2 = I$). Also $G\mathbf{1} = \mathbf{1}$ and $Gv_2 = \lambda v_2$. It follows that no matter what S is, $\left(\prod_{i=1}^{t} P^{\delta_S(i)} G\right) \mathbf{1}$ belongs to the two dimensional subspace $\mathrm{Span}\,(\mathbf{1}, v_2)$ and, furthermore, has a closed expression as a function of t, λ and S.

We finally choose a function f for which we can estimate the expression we get. We choose f to have high mass on its second Fourier level. It turns out that we can take f to be, e.g., the threshold function that returns one if the number of ones exceeds the mean by one standard deviation, and this function has error function that is of the order λ, and, therefore, in particular, vanishes with λ but not with t. It is interesting to note that, in contrast, the MAJ function, that has threshold at the mean, vanishes with t.

Next, using the same graph and labelling we also prove that constant spectral expansion does not suffice to fool AC^0 circuits. In fact, the bound obtained by [11] is tight up to a polynomial. Let $\mathsf{AC}(d)$ denote the class of all languages with polynomial size boolean circuit of depth at most d. Then:

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

The choice of function f here is more complicated. The key idea is that two adjacent bits obtained by such a random walk are λ correlated. Thus, evaluating a function f on the parity of consecutive bits obtained by a random walk is the same as applying the noise operator $T_\lambda(f)$ (see Claim 17 for an exact statement). Having this key fact, we construct small depth functions that are highly sensitive to small noise. We first start with the Tribes function composed with XOR on two adjacent bits. This gives a function in $\mathsf{AC}(3)$ with large distinguishability. We then give a recursive construction of a family of functions $h_d \in \mathsf{AC}(d+1)$ for every d, where in each step we increase the depth by one and the noise sensitivity of h_d by a logarithmic factor. This gives the desired dependence of $\mathcal{E}_{G,\mathsf{val}}(h_d)$ on d.

Finally, we also tighten and simplify the upper bounds given in [11]. We prove:

▶ **Theorem 3.** *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 3 improves upon the corresponding theorem in [11] in two ways:

1. First, the results in [11] are obtained only for balanced test functions f. In contrast, Theorem 3 holds for every test function f with arbitrary bias μ.
2. Second, the bound stated in Theorem 3 improves upon the bound in Equation (1.2) by removing the $\log^{3/2}(1/\lambda)$ factor.

The extension of the results of [11] to arbitrary bias μ is obtained by modifying the Fourier basis we work with. For a given bias μ we choose a basis that consists of $\prod_{i\in S}\frac{x_i-\mu}{\sqrt{1-\mu^2}}$ for all $S\subseteq[t]$. The improvement of the poly-logarithmic factor is achieved by using a more direct Fourier analysis argument. The proof strategy of [11] is to bound the error of weight indicator functions, and use it to handle weights around the mean. Then the argument invokes the expander Chernoff bound for bounding the remaining weights. Our approach does not go through analyzing weight indicator functions nor it uses the expander Chernoff bound. Instead, we use a very simple bound on the Fourier mass of symmetric functions, which gives a simpler and better analysis.

1.2 Open problems

We conclude the introduction with several 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 better fools AC^0?
2. As explained before, our lower bounds are obtained for a graph G that is a Cayley graph over an Abelian group. It is well-known that such a Cayley graph with constant expansion gap, has degree that depends on the number of vertices. Thus, a natural question is whether we can give similar lower bounds for constant degree graphs.
3. Continuing this line of thought, it is still possible that there is a family of graphs that fools all symmetric functions with error going down to zero with t. I.e., that while for some graphs (like Cayley graphs over $\mathbb{Z}_2^n$) there are bad labelling functions, for some other expander graphs, no such bad labellings exist. Similarly, it is possible that for some specific expanders better bounds exist for test functions in AC^0. Finding such graphs is a compelling goal that might require studying additional properties of graphs beyond expansion.
4. Finally, there is still a polynomial gap between the value of λ that fools functions in AC^0 and the corresponding lower bound we obtain. Any progress towards closing this gap will be interesting.

1.3 Paper organization

In Section 2 we give some background, mainly on Fourier Analysis. In Section 3 we recall the basic framework of [11], except that we do it for arbitrary bias μ rather than just bias $\mu=0$. In Section 4 we choose the graph and labelling function that we use for the lower bounds, and for which we can compute exactly the error induced by characters. In Section 5 we prove Theorem 1 and show that threshold function at one standard deviation away from the mean has error that goes down to zero with λ but not with t. In Section 6 we prove a special case of Theorem 2 for the case of $d=3$. The full proof of Theorem 2 appears in the full version of the paper. Then we turn to give a better upper bound on the error function and in show a better and tight upper bound with a simpler proof. Finally we show the threshold function about the mean (if $\mu=0$ it is MAJ) and weight indicator functions do vanish with t. The last two results appear in the full version of the paper.

2 Preliminaries

We let $[n] = \{1, \ldots, n\}$, $\mathbb{1} \in \mathbb{R}^n$ denote the all 1s vector, i.e., $\mathbb{1} = (1, \ldots, 1)^T \in \mathbb{R}^n$. We let $\mathbf{1} \in \mathbb{R}^n$ denote the normalized vector of $\mathbb{1}$, i.e $\mathbf{1} = \frac{1}{\sqrt{n}} \cdot \mathbb{1}$, we also use $\mathbf{J} := \mathbf{1}\mathbf{1}^{\mathsf{T}}$. When we write $\| \cdot \|$ we always refer to the L_2-norm. Unless stated otherwise, $\log x = \log_2 x$. Throughout the paper, we make use of the following well known inequalities about binomial coefficients. Let $a \geqslant b \geqslant 1$ be integers. 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 $\{\chi_S(x) \mid S \subseteq [t]\}$, where $\chi_S(x) = \prod_{i \in S} x_i$, 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 $\mu \in [-1, 1]$, and denote by U_t^μ the distribution over $\{\pm 1\}^t$ where each bit is chosen independently with expectation μ. Define $\langle f, g \rangle_\mu = \mathbf{E}_{x \sim U_t^\mu}[f(x) g(x)]$. 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 $\{\chi_S^\mu(x) \mid S \subseteq [t]\}$, forms an orthonormal basis with respect to this new 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 $\widehat{f_\mu}(S) \in \mathbb{R}$.

We say that a function $f : \{\pm 1\}^t \to \mathbb{R}$ is symmetric if for every permutation $\sigma \in S_t$, $f(x_1, \ldots, x_t) = f(x_{\sigma(1)}, \ldots, x_{\sigma(t)})$. It is not hard to show that if f is symmetric, then for every $S_1, S_2 \subseteq [t]$, with $|S_1| = |S_2|$, $\widehat{f_\mu}(S_1) = \widehat{f_\mu}(S_2)$. This allows us to use the following definition for symmetric functions: $\widehat{f_\mu}(k) = \left| \widehat{f_\mu}([k]) \right|$, which is the absolute value of the Fourier coefficients of any weight k character. For more details on biased Fourier analysis see Chapter 8 of [27].

3 The basic framework extended to arbitrary balanced tests

[11] reduced the analysis of the error function of a *balanced* test function f to the analysis of the error function of characters. In this section we restate this framework, but do it in a more general way that applies to any test function f, no matter how balanced it is.

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 $\mathbf{E}[\mathsf{val}(V)] = \mu$. Let $t \geqslant 1$ be an integer. 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 from a μ-biased distribution, that is a distribution with 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 $v_1, \dots, v_t$ in G and outputting the ordered tuple $(\mathsf{val}(v_1), \dots, \mathsf{val}(v_t))$. Equivalently, sample v_1 uniformly at random from V. Then, for $i = 2, 3, \dots, t$, sample v_i uniformly at random from the neighbours 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 4.** *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 $\mathbf{E}[\mathsf{val}] = \mu$, for $S \neq \emptyset$, $\mathbf{E}[\chi_S^\mu(U_t^\mu)] = 0$ and thus $\mathbf{E}[f(U_t^\mu)] = \widehat{f_\mu}(\emptyset)$. Hence,

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

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. ◀

Lemma 4 motivates us to consider parity test functions which we do next. 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 have 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 5.** *For integers $t \geqslant 1$, $2 \leqslant k \leqslant t$ and $j \in [k-2]$ define the map*

$$\Delta_j : \binom{[t]}{k} \to \mathbb{N}$$

as follows. Let $S \subseteq [t]$ of size $k \geqslant 2$ and denote $S = \{s_1, \dots, s_k\}$ where $s_1 < \cdots < s_k$. For $i \in [k-1]$ write $\delta_i = s_{i+1} - s_i$. Define

$$\Delta_j(S) = \min(\delta_j, \delta_{j+1}).$$

▶ **Definition 6.** *For an integer $t \geqslant 1$ define the map $\Delta : \binom{[t]}{\geqslant 2} \to \mathbb{N}$ as follows. Let $S \subseteq [t]$ of size $k \geqslant 2$. For $k = 2$ we define $\Delta(S) = \Delta_1(S)$, and for $k \geqslant 3$,*

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

We prove:

▶ **Proposition 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 $1 \leqslant k \leqslant t$ and $S \subseteq [t]$ of size k,*

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

We remark that for sets of size $|S| = 1$ the sum is taken over the empty index set $\mathcal{F}_1$ and so equals 0. We also note that when $|\mu| = 1$ the error is trivially zero, while our bound tends to infinity.

Proof. Consider any non-empty subset $S \subseteq [t]$ 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 $\delta_S(i) = 1$ if $i \in S$ and $\delta_S(i) = 0$ otherwise and observe that

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

where, recall, $\mathbf{1}$ is the all one vector normalized by $\frac{1}{\sqrt{n}}$. 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

$$\mathbf{E}[\chi_S^\mu(\mathrm{RW}_{G,\mathsf{val}})] = \mathbf{1}^T G^{t-s_k} \left(\prod_{i=1}^{k-1} P G^{\Delta_i}\right) P G^{s_1} \mathbf{1} = \mathbf{1}^T \left(\prod_{i=1}^{k-1} P G^{\Delta_i}\right) P \mathbf{1}, \tag{3.3}$$

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,

$$\prod_{i=1}^{k-1} P G^{\Delta_i} = \sum_{I \subseteq [k-1]} \prod_{i=1}^{k-1} P B_i(I), \tag{3.4}$$

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 (3.3) and (3.4) imply that

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] = \sum_{I\subseteq[k-1]} e_I. \tag{3.5}$$

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

$$\begin{aligned} e_I &= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I)\right)(P\mathbf{J})P\mathbf{1} \; = \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I)\right)(P\mathbf{1}\mathbf{1}^T)P\mathbf{1} \\ &= \mathbf{1}^T \left(\prod_{i=1}^{k-2} PB_i(I)\right) P\mathbf{1}(\mathbf{1}^T P\mathbf{1}) = 0, \end{aligned}$$

because

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

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

$$\begin{aligned} e_I &= \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I)\right)(PB_j(I))(PB_{j+1}(I))\left(\prod_{i=j+2}^{k-2} PB_i(I)\right) P\mathbf{1} \\ &= \mathbf{1}^T \left(\prod_{i=1}^{j-1} PB_i(I)\right)(P\mathbf{J})(P\mathbf{J})\left(\prod_{i=j+2}^{k-2} PB_i(I)\right) P\mathbf{1} = 0, \end{aligned}$$

Because

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

Thus, any subset $I \subseteq [k-1]$ that may contribute to the sum in Equation (3.5) is contained in $\mathcal{F}_k$ as defined in Equation (3.1).

Next, we look at $I \in \mathcal{F}_k$. We have that

$$e_I = \mathbf{1}^T \left(\prod_{i=1}^{k-1} PB_i(I)\right) P\mathbf{1} \; \leqslant \prod_{i=1}^{k-1} \|PB_i(I)\| \; \leqslant \|P\|^{k-1} \prod_{i\in I} \|B_i(I)\|. \tag{3.6}$$

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

$$\|P\| \leqslant \frac{\|M\| + \|\mu I\|}{\sqrt{1-\mu^2}} \leqslant \frac{1+|\mu|}{\sqrt{1-\mu^2}} = \sqrt{\frac{1+|\mu|}{1-|\mu|}}. \tag{3.7}$$

Equation (3.6) and Equation (3.7) together imply that $e_I \leqslant \left(\frac{1+|\mu|}{1-|\mu|}\right)^{\frac{k-1}{2}} \prod_{i\in I} \lambda^{\Delta_i}$. This proves the first inequality in the proposition.

To prove the second inequality consider $I \in \mathcal{F}_k$, and notice that

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

because for every $i \in [k-2]$, at least one of $i, i+1$ is in I. To complete the proof of the second inequality notice also that $|\mathcal{F}_k| \leqslant 2^{k-1}$. ◀

4 Choosing the graph

In this section 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.

▷ Claim 8. Let $G = ([n], E)$ be a regular graph with second largest eigenvalue λ_2 and corresponding eigenvector v_2. Further assume all coordinates in v_2 have ± 1 values. Define $\mathsf{val}_2 : [n] \to \{\pm 1\}$ 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 by 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

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

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

▶ **Definition 9.** *For $S \subseteq [t]$ denote $\Delta_{odd}(S) = \sum_{i=1}^{\lfloor (|S|-1)/2 \rfloor} \Delta_{2i+1}(S)$.*

▶ **Corollary 10.** *Let $G = ([n], E)$ and $\mathsf{val}_2 : [n] \to \{\pm 1\}$ be as above. Then,*

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}_2})] = \begin{cases} \lambda^{\Delta_{odd}(S)} & |S| \in \mathbb{N}_{even}, \\ 0 & |S| \in \mathbb{N}_{odd}. \end{cases}$$

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

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}_2})] = \mathbf{1}^T \left(\prod_{i=1}^{k-1} PG^{\Delta_i}\right) P\mathbf{1} = \frac{1}{n}\mathbb{1}^\mathsf{T} \left(\prod_{i=1}^{k-1} PG^{\Delta_i}\right) P\mathbb{1}.$$

Using Claim 8, we conclude that,

$$\mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}_2})] = \begin{cases} \lambda^{\Delta_{odd}(S)} \cdot \frac{1}{n}\mathbb{1}^\mathsf{T}\mathbb{1} & k \in \mathbb{N}_{even}, \\ \lambda^{\sum_{i=1}^{k-1/2} \Delta_{2i+1}} \cdot \frac{1}{n}\mathbb{1}^\mathsf{T} v_2 & k \in \mathbb{N}_{odd}. \end{cases}$$

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 $\mathbb{1}^\mathsf{T}\mathbb{1} = n$. ◀

We now give an example to such a graph G. Cayley graphs over an Abelian group commute and share an orthonormal basis of eigenvectors, which is known to be the set of all characters of the group. If the group is $\mathbb{Z}_2^n$, the eigenvectors have entries that are 2nd roots of unity, i.e., have ± 1 entries as desired. The eigenvalues have a direct correspondence to the set of generators of the Cayley graph. Building on that, [4] 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 $\lambda_2 = \lambda$. The degree of this graph depends both on n and λ.

From now on we let G be a regular expander with second largest eigenvalue λ and corresponding eigenvector with ± 1 entries, and we let val_2 reflect that eigenvector.

5 A lower bound for symmetric functions

In this section we prove the following theorem.

▶ **Theorem 11.** *Let $0 < c_0 \leqslant 1$, and let G, val_2 be as in the previous section, for $0 < \lambda < \frac{c_0^2}{12800 \cdot e}$. Let $f \colon \{\pm 1\}^t \to \{\pm 1\}$ be a symmetric function with $\left|\widehat{f}(2)\right| \geqslant \frac{c_0}{\sqrt{\binom{t}{2}}}$. Then,*

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

The idea behind the proof is to show that when choosing G, val_2 as in Section 4, the upper bound given by [11] is tight (up to the redundant poly logarithmic factor). We will use the following claim from [11].

▶ **Lemma 12** ([11], Lemma 4.4). *Denote*

$$\beta_k = \sum_{\substack{S \subseteq [t] \\ |S| = k}} \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})]. \quad \textit{Then,} \quad \beta_k \leqslant 2^k \binom{t-1}{\lfloor \frac{k}{2} \rfloor} \left(\frac{\lambda}{1-\lambda}\right)^{\lceil \frac{k}{2} \rceil}. \tag{5.1}$$

Using these notations we are now ready to prove Theorem 11.

Proof of Theorem 11. 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 $\Delta_{odd}(S) = 1$. Recall that $\mathcal{E}_{G,\mathsf{val}_2}(\chi_S) = 0$ if $|S| = 1$, therefore,

$$\mathcal{E}_{G,\mathsf{val}_2}(f) = \left| \sum_{S \subseteq [t], |S| \geqslant 2} \widehat{f}(|S|)\, \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \right| \tag{5.2}$$

$$\geqslant \left|\widehat{f}(2)\right| \left| \sum_{S \subseteq [t], |S| = 2} \mathbf{E}[\chi_S(\mathrm{RW}_{G,\mathsf{val}})] \right| - \left| \sum_{S \subseteq [t], |S| > 2} \widehat{f}(|S|) \mathcal{E}_{G,\mathsf{val}_2}(\chi_S) \right|. \tag{5.3}$$

However, by Corollary 10,

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

Furthermore,

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

where in the last inequality we used Lemma 12. The right hand side of the above equation is bounded above by

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

We omit the calculations. Assume that $\lambda \leqslant \frac{c_0^2}{128e \cdot 100}$. Then Equation (5.2) yields

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

In order to prove Theorem 1, we are left with providing a function f that satisfies the conditions of Theorem 11. Next, we show that the threshold function at one standard deviation distance from the mean has non-vanishing error in t.

We use the following definitions and claim. For integers t and $w \in \{0, 1, \ldots, t\}$ let $\mathbf{1}_w : \{\pm 1\}^t \to \{0, 1\}$ be the function indicating whether the weight of the input is w. That is, $\mathbf{1}_w(x_1, \ldots, x_t) = 1$ if $|\{i \in [t] \mid x_i = 1\}| = w$ and $\mathbf{1}_w(x_1, \ldots, x_t) = 0$ otherwise. We also define $\mathbf{1}_{>w} : \{\pm 1\}^{t+1} \to \{0, 1\}$ 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.

▷ Claim 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.

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

and the claim follows. ◁

Proof of Theorem 1. Take $f = \mathbf{1}_{>w}$ for $w = \frac{t - \sqrt{t}}{2}$. We claim that $\left|\widehat{f}(2)\right| > \frac{c_0}{\sqrt{\binom{t}{2}}}$, for some absolute constant $c_0 > 0$ and therefore by Theorem 11, $\mathcal{E}_{G,\mathsf{val}}(f_t) \geqslant c \cdot \lambda$ for some constant c. Indeed, by Claim 13 we have $\widehat{f}(2) = \widehat{\mathbf{1}_w}(1)$ for $\mathbf{1}_w : \{\pm 1\}^{t-1} \to \{0, 1\}$. To compute $\widehat{\mathbf{1}_w}(1)$ we apply [11, 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 $\binom{t-1}{\frac{t - \sqrt{t}}{2}} \geqslant \Omega\left(\frac{1}{\sqrt{t}} 2^t\right)$, concludes the proof. ◀

6 A lower bound for AC^0 tests

In this section we use the noise operator. The following definitions and claims appear in [27].

▶ **Definition 14.** *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,*

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

▶ **Definition 15.** *Let $\rho \in [-1,1]$. The noise operator T_ρ is the linear operator on functions $\{\pm1\}^t \to \mathbb{R}$ 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 16.** *For every function $f : \{\pm1\}^t \to \mathbb{R}$ it holds that: $\widehat{T_\rho f}(S) = \widehat{f}(S)\rho^{|S|}$.*

The starting point of this section is to connect the expectation of f under a random walk and the noise function $T_\lambda(f)$, we prove this claim in the full version of the paper.

▷ Claim 17. For $f \colon \{\pm1\}^t \to \mathbb{R}$ define $\tilde{f} \colon \{\pm1\}^{2t} \to \mathbb{R}$ by

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

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

Proof. For $\{s_1, \dots, s_k\} = S \subseteq [t]$ denote $2S \colon = \{2s_1 - 1, 2s_1, \dots, 2s_k - 1, 2s_k\} \subseteq [2t]$. Note that $\Delta_{odd}(2S) = |S|$.

$$\begin{aligned}\tilde{f}(x_1, x_2, \dots, x_{2t-1}, x_{2t}) &= f(x_1 \cdot x_2, \dots, x_{2t-1} \cdot x_{2t}) \\ &= \sum_{S \subseteq [t]} \widehat{f}(S)\chi_S(x_1 \cdot x_2, \dots, x_{2t-1} \cdot x_{2t}) \\ &= \sum_{S \subseteq [t]} \widehat{f}(S)\chi_{2S}(x_1, x_2, \dots, x_{2t-1}, x_{2t}).\end{aligned}$$

Therefore,

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

which is equal to $T_\lambda(f)(\mathbb{1})$ by Lemma 16. For the second equality we used Corollary 10. ◁

6.1 A lower bound for the Tribes function composed with IP

We now construct a function in $\mathsf{AC}(3)$, that satisfies Theorem 2. Later on we extend the construction inductively to obtain the general theorem. The idea behind the depth-3 construction is the following. We look for a function $f = f(x_1, \dots, x_t) \in \mathsf{AC}(2)$ such that

$$|\,\mathbf{E}[f(U_t)] - T_\lambda(f)(\mathbb{1})| \geqslant \lambda \cdot \log t. \tag{6.1}$$

We then look at $\tilde{f}(y_1, \dots, y_{2t}) = f(y_1 \cdot y_2, \dots, y_{2t-1} \cdot y_{2t}) \in \mathsf{AC}(3)$ and note that:

- $\mathbf{E}[\tilde{f}(U_t)] = \mathbf{E}[f(U_t)]$ as the product of two uniform ±1 bits is uniform; However,
- by Claim 17, $\mathbf{E}[\tilde{f}(\mathrm{RW}_{G,\mathsf{val}_2})] = T_\lambda(f)(\mathbb{1})$.

Together,

$$\mathcal{E}_\lambda(\tilde{f}) \geqslant \mathcal{E}_{G,\mathsf{val}_2}(\tilde{f}) = |\,\mathbf{E}[\tilde{f}(U_t)] - \mathbf{E}[\tilde{f}(\mathrm{RW}_{G,\mathsf{val}_2})]| = |\,\mathbf{E}[f(U_t)] - T_\lambda(f)(\mathbf{1})] \geqslant \lambda \cdot \log t,$$

which in turns implies Theorem 2, for $d = 3$.

We take f to be the Tribes function. Fix t; we choose parameters r, h such that $r \cdot h \leqslant t$ by taking $h = \log(t) - \log\log(t)^2$ and $r = \lfloor \frac{t}{\log t} \ln(2) \rfloor$. Partition $[t]$ into disjoint sets $I_1, \dots, I_r$, each of size h. We define $f : \{\pm 1\}^t \to \{0, 1\}$ to be the Tribes function on t bits and define g to be the related function

$$f(z_1, \dots, z_t) = \bigvee_{i \in [r]} \bigwedge_{j \in I_i} z_j, \qquad\qquad g(z_1, \dots, z_t) = \bigwedge_{i \in [r]} \bigvee_{j \in I_i} z_j.$$

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

As before, we choose G to be a Cayley graph on the boolean hypercube with $\lambda_2 = \lambda$ and $\mathsf{val} = \mathsf{val}_2$.

▷ Claim 18. 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 $g(x_1, \dots, x_t) = 1 - f(\overline{x_1}, \dots, \overline{x_t})$, so $\mathbf{E}[g] = 1 - \mathbf{E}[f]$ and so it is enough to prove the statement for f. To this end write

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

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

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

as desired. ◁

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

▷ Claim 19. Let $p = \frac{1-\varepsilon}{2}$ and assume $\varepsilon \geqslant \frac{k}{\log(t)}$. Then,

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

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

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

[2] Recall that $\log t = \log_2 t$

Using $(1-\delta)^r \geqslant 1 - r\delta$, we get that the above expression is bounded by $r2^{-h}e^{-k} \leqslant e^{-k}$. 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

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

We now prove Theorem 2 for $d = 3$. We take $h(x_1, y_1, \dots, x_t, y_t) = f(x_1 \cdot y_1, \dots, x_t \cdot y_t)$. $h \in \mathsf{AC}(3)$ because $f \in \mathsf{AC}(2)$.

- On the one hand, by Claim 17, $\mathbf{E}[h(\mathrm{RW}_{G,\mathsf{val}_2})] = T_\lambda(f)(\mathbb{1}) = \mu_{\frac{1-\lambda}{2}}(f)$. By Claim 19, and using $\lambda \geqslant \frac{k}{\log t}$, we get that $\mathbf{E}[h(\mathrm{RW}_{G,\mathsf{val}_2})] < e^{-k/10}$.
- On the other hand, By Claim 18, $\mathbf{E}[h] = \mathbf{E}[f] \geqslant \frac{1}{2} - O(\frac{\log t}{t})$.

Together, h is as desired.

References

1 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.

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

3 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.

4 Noga Alon and Yuval Roichman. Random cayley graphs and expanders. *Random Struct. Algorithms*, 5(2):271–285, 1994. doi:10.1002/rsa.3240050203.

5 Mihir Bellare, Oded Goldreich, and Shafi Goldwasser. Randomness in interactive proofs. *Computational Complexity*, 3(4):319–354, 1993.

6 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. doi:10.1137/080732651.

7 Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006. doi:10.1007/s00493-006-0029-7.

8 Mark Braverman. Polylogarithmic independence fools AC^0 circuits. *J. ACM*, 57(5):Art. 28, 10, 2010. doi:10.1145/1754399.1754401.

9 Mark Braverman, Gil Cohen, and Sumegha Garg. Pseudorandom pseudo-distributions with near-optimal error for read-once branching programs. *SIAM J. Comput.*, 49(5), 2020. doi:10.1137/18M1197734.

10 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.

11 Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: a fourier-analytic approach. In Samir Khuller and Virginia Vassilevska Williams, editors, *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing, Virtual Event, Italy, June 21-25, 2021*, pages 1643–1655. ACM, 2021. doi:10.1145/3406325.3451049.

12 Irit Dinur. The PCP theorem by gap amplification. *J. ACM*, 54(3):Art. 12, 44, 2007. doi:10.1145/1236457.1236459.

13 Irit Dinur, Shai Evra, Ron Livne, Alexander Lubotzky, and Shahar Mozes. Locally testable codes with constant rate, distance, and locality. *CoRR*, abs/2111.04808, 2021. arXiv:2111.04808.

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

15 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.

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

17 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc. (N.S.)*, 43(4):439–561, 2006. doi:10.1090/S0273-0979-06-01126-8.

18 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.

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

20 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.

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

22 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. doi:10.1145/3051093.

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

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

25 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.

26 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.

27 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CBO9781139814782.

28 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.

29 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. doi:10.1145/1060590.1060647.

30 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.

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

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

33 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. doi:10.1145/3055399.3055408.

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

35 Luca Trevisan. Lecture notes on graph partitioning, expanders and spectral methods. *University of California, Berkeley*, 2017. URL: https://lucatrevisan.github.io/books/expanders-2016.pdf.

36 Salil P Vadhan. *Pseudorandomness*, volume 7. Now, 2012.

37 Leslie G. Valiant. Graph-theoretic properties in computational complexity. *J. Comput. System Sci.*, 13(3):278–285, 1976. doi:10.1016/S0022-0000(76)80041-4.