A Lower Bound on the Trace Norm of Boolean Matrices and Its Applications

Tsun-Ming Cheung
School of Computer Science, McGill University, Montreal, Canada

Hamed Hatami
School of Computer Science, McGill University, Montreal, Canada

Kaave Hosseini
Department of Computer Science, University of Rochester, NY, USA

Aleksandar Nikolov
Department of Computer Science, University of Toronto, Canada

Toniann Pitassi
Department of Computer Science, Columbia University, New York, NY, USA

Morgan Shirley
Department of Computer Science, University of Victoria, Canada

Abstract

We present a simple method based on a variant of Hölder's inequality to lower-bound the trace norm of Boolean matrices. As the main result, we obtain an exponential separation between the randomized decision tree depth and the spectral norm (i.e. the Fourier L_1-norm) of a Boolean function. This answers an open question of Cheung, Hatami, Hosseini and Shirley (CCC 2023). As immediate consequences, we obtain the following results.

- We give an exponential separation between the logarithm of the randomized and the deterministic parity decision tree *size*. This is in sharp contrast with the standard binary decision tree setting where the logarithms of randomized and deterministic decision tree size are essentially polynomially related, as shown recently by Chattopadhyay, Dahiya, Mande, Radhakrishnan, and Sanyal (STOC 2023).
- We give an exponential separation between the approximate and the exact spectral norm for Boolean functions.
- We give an exponential separation for XOR functions between the deterministic communication complexity with oracle access to Equality function (D^{EQ}) and randomized communication complexity. Previously, such a separation was known for general Boolean matrices by Chattopadhyay, Lovett, and Vinyals (CCC 2019) using the Integer Inner Product (`IIP`) function.
- Finally, our method gives an elementary and short proof for the mentioned exponential D^{EQ} lower bound of Chattopadhyay, Lovett, and Vinyals for Integer Inner Product (`IIP`).

2012 ACM Subject Classification Theory of computation → Communication complexity; Theory of computation → Oracles and decision trees

Keywords and phrases Boolean function complexity, parity decision trees, randomized communication complexity

Digital Object Identifier 10.4230/LIPIcs.ITCS.2025.37

Funding *Hamed Hatami*: Supported by an NSERC grant.
Kaave Hosseini: Partially supported by the Goergen Institute for Data Science at the University of Rochester.
Morgan Shirley: Research was done while the author was a student at the University of Toronto. Supported by an NSERC grant.

1 Introduction

We start by recalling two fundamental notions related to the complexities of Boolean functions: the *randomized parity decision tree complexity* and the *spectral norms*.

A *parity decision tree* (PDT) for a Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ is similar to a standard decision tree with the following strengthened query power. Instead of a single variable, each internal node can query the parity (sum over $\mathbb{F}_2$) of any subset of variables, e.g., $x_1 \oplus x_5 \oplus x_6$. The *parity decision tree complexity* of f, denoted by $\mathsf{PDT}_{\mathrm{depth}}(f)$, is the smallest depth of a PDT that outputs the correct value of $f(x)$ on every input x. A *randomized parity decision tree* (RPDT) of depth at most d is a probability distribution over PDTs of depth at most d. It computes f with error ϵ if, for every input x, the RPDT outputs $f(x)$ with probability at least $1-\epsilon$. The *randomized parity decision tree complexity*, denoted by $\mathsf{RPDT}_{\mathrm{depth}}(f)$, is the smallest depth of an RPDT computing f with error $\epsilon = 1/3$.

Parity decision trees are closely related to the Fourier expansions of Boolean functions. The *spectral norm* (also known as *algebra norm*) of a function $f : \mathbb{F}_2^n \to \mathbb{R}$, denoted by $\|\widehat{f}\|_1$, is the sum of the absolute values of the Fourier coefficients.

It is easy to see from the Fourier expansion of a PDT that every Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ satisfies $\log \|\widehat{f}\|_1 \leq \mathsf{PDT}_{\mathrm{depth}}(f)$. The randomized counterpart of this inequality does not hold as illustrated by h, the indicator function of the standard basis vectors $\{e_1, \ldots, e_n\}$: as observed in [14], $\mathsf{RPDT}_{\mathrm{depth}}(h) = O(1)$, but $\log \|\widehat{h}\|_1 = \Theta(\log n)$. Nonetheless, this still leaves the possibility of $\log \|\widehat{f}\|_1 = \widetilde{O}(\mathsf{RPDT}_{\mathrm{depth}}(f))$ open, where $\widetilde{O}(\cdot)$ hides polylogarithmic dependencies on n. Cheung et al. [8] specifically asked whether such a relation could hold, which if confirmed, would have established the stronger lower bound on $\mathsf{RPDT}_{\mathrm{depth}}$.

The main result of this work gives an example in which $\log \|\widehat{f}\|_1$ is exponential in $\mathsf{RPDT}_{\mathrm{depth}}(f)$, answering the question by [8] in the negative.

▶ **Theorem 1** (Main theorem). *There exists a function $f : \mathbb{F}_2^n \to \{0,1\}$ with*

$$\mathsf{RPDT}_{\mathrm{depth}}(f) = O(\log n) \quad \textit{and} \quad \log \|\widehat{f}\|_1 = \Omega(n).$$

In the converse direction, the quadratic gap $\mathsf{RPDT}_{\mathrm{depth}}(f) = O(\|\widehat{f}\|_1^2)$ holds for every Boolean function (see e.g. [14, Lemma 2.7]). Chattopadhyay, Mande, and Sherif [5] proved that the sink function SINK satisfies $\mathsf{RPDT}_{\mathrm{depth}}(\mathsf{SINK}) = \Omega(\|\widehat{\mathsf{SINK}}\|_1)$. This result and Theorem 1 together demonstrate that the measures $\log \|\widehat{f}\|_1$ and $\mathsf{RPDT}_{\mathrm{depth}}(f)$ are not polynomially related in both ways.

The proof technique used in the main theorem involves an application of Hölder's inequality with carefully-chosen parameters (see Section 3) that allow for a combinatorial interpretation of the problem. This framework is useful not only for the question considered above but also in communication complexity, where we will use the Hölder's inequality technique to simplify the proofs of existing lower bounds. For the rest of this section, we will discuss applications of both the proof technique and of the main theorem itself.

1.1 Communication Complexity: The Power of Oracle Access to Equality

Arguably the most well-known communication problem is the Equality function EQ, in which the two parties compare if their inputs are identical. In the model of *public-coin randomized communication*, two parties are given a publicly accessible random string and are

allowed to make errors with probability bounded away from 1/2. It is an elementary result that $\text{EQ} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ has a randomized protocol requiring only $O(1)$ bits of communication, while any deterministic protocol of EQ requires $n+1$ bits of communication.

A natural question would be whether Equality fully captures the power of randomness in communication. A formal formulation of the question is to compare the relative power of randomized protocols and *deterministic protocols with oracle access to* Equality or, in short, D^{EQ} protocols. A D^{EQ} protocol performs communications as usual; in addition, the parties are given access to an oracle that computes Equality function, and each oracle call is charged at a cost of one bit. The seminal work by Chattopadhyay, Lovett and Vinyals [4] considered this question and proved an exponential separation between the D^{EQ} complexity and randomized communication complexity (denoted as R) of the *Integer Inner Product* function ($\texttt{IIP}$) in dimension at least 6.

For a chosen dimension k and parameter n, the Boolean matrix $\texttt{IIP}_k^{(n)} : \{-2^n, \ldots, 2^n\}^k \times \{-2^n, \ldots, 2^n\}^k \to \{0,1\}$ is defined by

$$\texttt{IIP}_k[(x_1, \ldots, x_k), (y_1, \ldots, y_k)] = \begin{cases} 1 & \text{if } \sum_{i=1}^k x_i y_i = 0 \\ 0 & \text{otherwise} \end{cases}.$$

For this communication problem, each player holds a $\Theta(kn)$-bit input. As we only consider the case when $k = O(1)$, we treat $\texttt{IIP}_k^{(n)}$ as a $\Theta(n)$-bit communication problem.

Chattopadhyay, Lovett, and Vinyals presented a (one-way) randomized protocol for $\texttt{IIP}_k^{(n)}$ of cost $O(k \log n)$. For the lower bound, they introduced a lower bound technique which we call *relative area method* (see Theorem 13), and showed that no D^{EQ} protocol could compute the function with fewer than $\Omega(n)$ bits of communication.

▶ **Theorem 2** ([4]). *For constant $k \geq 6$, $\mathsf{R}\left(\texttt{IIP}_k^{(n)}\right) = O(\log n)$ and $\mathsf{D}^{\mathsf{EQ}}\left(\texttt{IIP}_k^{(n)}\right) = \Omega(n)$.*

In a subsequent work, Cheung et al. [8] obtained a strengthening of Theorem 2 via a different approach of spectral methods. It was shown in [14] that

$$\mathsf{D}^{\mathsf{EQ}}(A) \geq \frac{1}{2} \log \|A\|_{\text{ntr}} \tag{1}$$

for any boolean matrix $A \in \{0,1\}^{m \times n}$, where $\|A\|_{\text{ntr}}$ is the *normalized trace norm* of A with the normalization factor from the matrix dimensions (see Definition 10 for the precise definition).

Cheung et al. [8] observed that for $k \geq 3$, the $\Theta(2^{kn}) \times \Theta(2^{kn})$ matrix $\texttt{IIP}_k^{(n)}$ contains a $2^{4n/3} \times 2^{4n/3}$ *point-line incidence* matrix $\texttt{PL}^{(4n/3)}$ as a submatrix. The matrix $\texttt{PL}^{(n)}$ is defined on the domains $\mathcal{P} = \mathcal{L} = [2^{n/4}] \times [2^{3n/4}] \subseteq \mathbb{Z}^2$, and the entries of $\texttt{PL}^{(n)} : \mathcal{P} \times \mathcal{L} \to \{0,1\}$ are given by

$$\texttt{PL}[(x,x'),(y,y')] = \begin{cases} 1 & \text{if } xy + x' = y' \\ 0 & \text{otherwise} \end{cases}.$$

They proved that the normalized trace norm of $\texttt{PL}^{(n)}$ is exponential in n, which shows that $\mathsf{D}^{\mathsf{EQ}}\left(\texttt{PL}^{(n)}\right) = \Omega(n)$. Since communication complexity does not increase under restriction, this subsequently implies Theorem 2.

▶ **Theorem 3** ([8]). *The $2^m \times 2^m$ Boolean matrix $\texttt{PL}^{(m)}$ satisfies $\|\texttt{PL}^{(m)}\|_{\text{ntr}} = \Omega(2^{m/32})$. Consequently, for $k \geq 3$,*

$$\mathsf{D}^{\mathsf{EQ}}\left(\texttt{IIP}_k^{(n)}\right) \geq \frac{n}{48} + O(1).$$

A common shortcoming of both proofs of Theorem 2 in [4, 8] is their highly technical nature. In Section 4, we give a considerably simplified proof – with improved linear factor on $\mathsf{D}^{\mathsf{EQ}}\left(\mathtt{IIP}_k^{(n)}\right)$ – based on the Hölder's inequality technique. Here, we consider a different submatrix of $\mathtt{IIP}_k$, which is the point-line incidence system $\mathtt{PL}_*$ that proves the optimality of Szemerédi–Trotter theorem. We will provide the precise definition of $\mathtt{PL}_*^{(m)}$ in Section 4.

▶ **Theorem 4** (Improved version of Theorem 3). *The $2^{m+1} \times 2^m$ Boolean matrix $\mathtt{PL}_*^{(m)}$ satisfies $\|\mathtt{PL}_*^{(m)}\|_{\mathrm{ntr}} = \Omega(2^{m/6})$. Consequently, for $k \geq 3$,*

$$\mathsf{D}^{\mathsf{EQ}}\left(\mathtt{IIP}_k^{(n)}\right) \geq \frac{n}{8} + O(1).$$

Nondeterministic communication

When analyzing the power of Equality in communication, another model of interest is *nondeterministic protocols with oracle access to* Equality, or N^{EQ} for short. For the precise definition of the N^{EQ} model, we refer readers to [19]. N^{EQ} and its relationships with related models have been studied previously, both implicitly in [12] and explicitly in [19]. In the latter work, Pitassi, Shirley and Shraibman analyzed the relative area method used by [4] to lower bound $\mathsf{D}^{\mathsf{EQ}}(\mathtt{IIP}_6)$ and showed that the same technique is applicable to N^{EQ} complexity as well. It is not clear that the lower bound technique of [8] on the D^{EQ} complexity of $\mathtt{IIP}_3$ works for N^{EQ}, so prior to this work it was open whether $\mathtt{IIP}_3$ is hard for N^{EQ}. We observe that the analysis in Theorem 4 can be adapted to the relative area method, and therefore yields an almost linear lower bound on $\mathsf{N}^{\mathsf{EQ}}(\mathtt{IIP}_k)$ for every $k \geq 3$.

▶ **Theorem 5.** *For constant $k \geq 3$,*

$$\mathsf{N}^{\mathsf{EQ}}\left(\mathtt{IIP}_k^{(n)}\right) = \Omega\left(\frac{n}{\log n}\right).$$

As far as we know, this method cannot be improved to give a linear lower bound on $\mathsf{N}^{\mathsf{EQ}}(\mathtt{IIP}_3)$. We leave resolving this logarithmic gap as an open question.

XOR-lifts

A related open problem posed in [8] is whether an exponential separation of D^{EQ} and $\mathbb{R}$ holds for an XOR-lift. For a Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$, the XOR-lift $f^{\oplus} : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \{0,1\}$ is defined by $f^{\oplus}(x,y) = f(x \oplus y)$ for each $x, y \in \mathbb{F}_2^n$. It can be verified that $\mathtt{IIP}_k$ is not an XOR-lift, so the D^{EQ}-vs-R separation for the XOR-lift case was open.

The matrix class of XOR-lifts is interesting in complexity theory since many complexity measures of the matrix $f^{\oplus}$ can be related (or even equated) to complexity measures of the Boolean function f. Indeed, the query-to-communication connections allow us to translate the separation question to a purely query-complexity setting. For any Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$, the inequality

$$\mathsf{R}(f^{\oplus}) \leq 2\,\mathsf{RPDT}_{\mathrm{depth}}(f) \tag{2}$$

is evident from the standard simulation of a randomized parity decision tree by a communication protocol. A result of [10] shows that

$$\|f^{\oplus}\|_{\mathrm{ntr}} = \|\widehat{f}\|_1. \tag{3}$$

Immediate from Equations (1)–(3), we obtain the exponential D^{EQ}-vs-R separation for XOR-lifts from Theorem 1.

▶ **Corollary 6.** *There exists a Boolean function* $f : \mathbb{F}_2^n \to \{0,1\}$ *such that its* XOR*-lift* $f^{\oplus}$ *satisfies* $\mathsf{R}(f^{\oplus}) = O(\log n)$ *and* $\mathsf{D}^{\mathsf{EQ}}(f^{\oplus}) = \Omega(n)$.

1.2 Query Complexity: The Power of Randomness

Understanding the power of randomized versus deterministic algorithms is a fundamental problem in complexity theory. Depending on the computational model, this problem varies from fully resolved to forbiddingly out of reach. In the standard decision tree model, it is well known that randomness does not significantly reduce the number of queries. One early result in complexity theory by Nisan [18] showed that a randomized decision tree of depth d computing a Boolean function can be simulated by a deterministic decision tree of depth at most $O(d^3)$. On the other end of the spectrum, the randomized-versus-deterministic problem remains overwhelmingly difficult for Turing Machines.

A generic framework to study the relative powers of two computation models is to study the relations of the complexity classes defined by suitable complexity measures. For the sake of comparisons between measures, we consider the complexity measures normalized to the range $[0, n]$. Given a complexity measure $C(\cdot)$ for a deterministic computation model, $\mathtt{P}$ is defined to be the class of functions $f : \mathbb{F}_2^n \to \{0,1\}$ (more accurately, sequences of functions f_n) such that $C(f) \leq \text{polylog}(n)$. We define the class $\mathtt{BPP}$ similarly for the randomized counterpart of $C(\cdot)$. The inclusion $\mathtt{P} \subseteq \mathtt{BPP}$ for every measure is obvious, so the randomized-versus-deterministic problem amounts to whether the classes $\mathtt{P}$ and $\mathtt{BPP}$ are equal.

We consider the two natural measures for trees, namely depth and logarithmic size (for normalization purposes) and four types of decision trees: (standard) decision tree (DT), parity decision tree (PDT), randomized decision tree (RDT), and randomized parity decision tree (RPDT). This branches into eight complexity measures in concern, and we define the other six measures $\mathsf{DT}_{\text{depth}}$, $\log \mathsf{DT}_{\text{size}}$, $\mathsf{RDT}_{\text{depth}}$, $\log \mathsf{RDT}_{\text{size}}$, $\mathsf{PDT}_{\text{depth}}$ and $\log \mathsf{PDT}_{\text{size}}$ in an analogous manner to $\mathsf{RPDT}_{\text{depth}}$ and $\log \mathsf{RPDT}_{\text{size}}$. Based on the type of queries equipped and the complexity measure, the P-vs-BPP question branches into four pairs for comparison.

Standard decision trees

Nisan [18] showed that $\mathsf{DT}_{\text{depth}}(f) \leq O(\mathsf{RDT}_{\text{depth}}(f)^3)$ in 1988 and settled that $\mathtt{P} = \mathtt{BPP}$ in the depth setting. The question in the size setting remained unsettled for decades until recently Chattopadhyay, Dahiya, Mande, Radhakrishnan, and Sanyal [3] showed that

$$\log \mathsf{DT}_{\text{size}}(f) = O\left(\log^4(\mathsf{RDT}_{\text{size}}(f)) \log^3 n\right),$$

concluding that $\mathtt{P} = \mathtt{BPP}$ in this setting as well.

Parity decision trees

In the depth setting, the two classes $\mathtt{P}$ and $\mathtt{BPP}$ are strongly separated by the simple function of OR. It is well-known that the OR function on n bits satisfies $\mathsf{RPDT}_{\text{depth}}(\mathsf{OR}) = O(1)$ but $\mathsf{PDT}_{\text{depth}}(\mathsf{OR}) = n$, which provides the optimal separation and hence $\mathtt{P} \neq \mathtt{BPP}$ in this setting.

Note that however $\mathsf{PDT}_{\text{size}}(\mathsf{OR}) = O(n)$, so OR does not attest a separation between $\mathtt{P}$ and $\mathtt{BPP}$ in the size setting. Indeed, previous to this work the P-vs-BPP question was unknown in the size setting. As mentioned that $\log \|\widehat{f}\|_1 \leq \mathsf{PDT}_{\text{depth}}(f)$ for every function f, an immediate corollary of Theorem 1 implies that $\mathtt{P} \neq \mathtt{BPP}$ in the size setting for parity decision trees.

▶ **Corollary 7.** *There is a function* $f : \mathbb{F}_2^n \to \{0,1\}$ *such that* $\log \mathsf{RPDT}_{\text{size}}(f) = O(\log n)$, *but* $\log \mathsf{PDT}_{\text{size}}(f) = \Omega(n)$.

In fact, the randomized parity decision tree in our construction is non-adaptive and has a one-sided error, hence we obtain the stronger separation $\mathtt{P} \neq \mathtt{RP}$[1] in the size setting parity decision tree size model. It remains an interesting open problem to determine whether a separation of $O(1)$-vs-$\Omega(n)$ for $\log \mathsf{RPDT}_{\text{size}}$ and $\log \mathsf{PDT}_{\text{size}}$ is possible, as in the case of depth illustrated by the OR function.

▶ **Question 8.** *Is there a boolean function* $f : \mathbb{F}_2^n \to \{0,1\}$ *such that* $\log \mathsf{RPDT}_{\text{size}}(f) = O(1)$, *but* $\log \mathsf{PDT}_{\text{size}}(f) = \Omega(n)$, *or even* $\log \mathsf{PDT}_{\text{size}}(f) = \omega(\log n)$*?*

1.3 Fourier Analysis: Approximate versus Exact Spectral Norms

The notion of *approximate norms* is customary in complexity theory as complexity measures for models with error tolerance. The *approximate spectral norm* of a function f with error ϵ, denoted as $\widehat{L}_{1,\epsilon}(f)$[2], is the minimum $\|\widehat{g}\|_1$ for some function g such that $\|f - g\|_\infty \leq \epsilon$. As in the case of other complexity measures, we adopt the canonical choice $\epsilon = 1/3$ when referring to constant error. Spectral norm and approximate spectral norm of a Boolean function are fundamental parameters that find applications in many areas such as learning theory [17], complexity theory [21, 22, 23, 1], communication complexity [23, 14, 8], Fourier analysis [23, 11] and additive combinatorics [13, 20, 9].

It is natural to ask whether the spectral norm of a Boolean function is upper bounded by its approximate spectral norm. Towards answering this question, Cheung et al. [9] observed that if f is the indicator function of n-bit strings of Hamming-weight 1, then $\widehat{L}_{1,1/3}(f) = O(1)$ but $\|\widehat{f}\|_1 = \Omega(\sqrt{n})$. Nevertheless, this separation does not rule out the possibility of polynomial relations with dependency on n such as $\|\widehat{f}\|_1 = \text{poly}\left(\widehat{L}_{1,1/3}(f), n\right)$. Using the result by [14] that $\log \|\widehat{f}\|_1 \leq \mathsf{RPDT}_{\text{depth}}(f)$ for any Boolean function f, Theorem 1 immediately rules out the possibility of polynomial dependency.

▶ **Corollary 9.** *There exists a function* $f : \mathbb{F}_2^n \to \{0,1\}$ *such that* $\widehat{L}_{1,1/3}(f) = n^{O(1)}$ *but* $\|\widehat{f}\|_1 = 2^{\Omega(n)}$.

2 Preliminaries

For a positive integer k, we denote $[k] \coloneqq \{1, \ldots, k\}$. All logarithms in this paper are base 2.

We denote the indicator function of a predicate P as $\mathbf{1}[P]$ and the indicator function of a set S as $\mathbf{1}_S$. For a random variable r and a set S, we write $r \sim S$ to indicate that r is uniformly sampled from S. Throughout this work, we adopt the standard big-O notations in computer science.

2.1 Schatten norms

For a vector $v \in \mathbb{R}^k$ and $p \in [1, \infty]$, we denote the ℓ_p-norm of v as $\|v\|_p$. For a matrix $A \in \mathbb{R}^{m \times n}$, the singular values of A are the square roots of the eigenvalues of $AA^\top$, which we denote by $\sigma_1(A) \geq \sigma_2(A) \geq \ldots \geq \sigma_{\min\{m,n\}}(A) \geq 0$. The central matrix norms in this paper are *Schatten p-norm*, which is defined as

[1] RP denotes the class of functions computable by RPDT with constant one-sided error and poly-logarithmic costs

[2] We adopt this function-like notation to emphasize that approximate spectral norm is not a norm despite what the name suggests

$$\|A\|_{S_p} = \left(\sum_i \sigma_i(A)^p\right)^{1/p}$$

for $p \in [1,\infty)$, and $\|A\|_{S_\infty} = \sigma_1(A)$. Put differently, Schatten p-norm of a matrix is the ℓ_p norm of the vector $\big[\sigma_1(A), \sigma_2(A), \ldots, \sigma_{\min\{m,n\}}(A)\big]$. One property of Schatten norms inherited from ℓ_p norms is the monotonicity of Schatten p-norm in p: $\|A\|_{S_p} \geq \|A\|_{S_q}$ for $1 \leq p < q \leq \infty$.

The Schatten 1-norm, 2-norm and ∞-norm are frequently used and are commonly known as *trace norm*, *Frobenius norm*, and *spectral norm* respectively:

$$\|A\|_{\mathrm{tr}} = \|A\|_{S_1} = \sum_i \sigma_i(A);$$

$$\|A\|_{\mathrm{F}} = \|A\|_{S_2} = \sqrt{\sum_i \sigma_i(A)^2} = \sqrt{\sum_{ij} A_{ij}^2};$$

$$\|A\| = \|A\|_{S_\infty} = \sigma_1(A) = \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_2}{\|x\|}.$$

For applications in theoretical computer science, it is more convenient to work with the *normalized trace norm* as a complexity measure.

▶ **Definition 10** (Normalized trace norm). *For $A \in \mathbb{R}^{m\times n}$, the normalized trace norm of A is*

$$\|A\|_{\mathrm{ntr}} \coloneqq \frac{\|A\|_{\mathrm{tr}}}{\sqrt{mn}} = \frac{1}{\sqrt{mn}} \sum_i \sigma_i(A).$$

We remind readers that unlike essentially all other complexity measures and matrix norms used in this work, the normalized trace norm may increase upon restriction to a submatrix.

2.2 Fourier analysis of Boolean functions

This section gives a basic overview of Fourier analysis on the Boolean cube. For every $\eta \in \mathbb{F}_2^n$, define the Fourier character $\chi_\eta : \mathbb{F}_2^n \to \{\pm 1\}$ as $\chi_\eta(x) = (-1)^{\langle x,\eta\rangle}$, where $\langle x, \eta\rangle$ is the standard inner product over $\mathbb{F}_2^n$. For a function $f : \mathbb{F}_2^n \to \mathbb{R}$, its Fourier expansion is given by

$$f = \sum_{\eta \in \mathbb{F}_2^n} \widehat{f}(\eta)\chi_\eta, \tag{4}$$

where the Fourier coefficient $\widehat{f}(\eta)$ is defined as $\widehat{f}(\eta) \coloneqq \mathbb{E}_{x\sim\mathbb{F}_2^n}[f(x)\chi_\eta(x)]$.

For a function $f : \mathbb{F}_2^n \to \mathbb{R}$ and $p \in [1,\infty)$, the Fourier p-norm is defined as

$$\|\widehat{f}\|_p = \left(\sum_{x\in\mathbb{F}_2^n} |\widehat{f}(\eta)|^p\right)^{1/p},$$

and $\|\widehat{f}\|_\infty = \max_\eta |\widehat{f}(\eta)|$. In other words, the Fourier p-norm is the ℓ_p norm of the vector of Fourier coefficients. Fourier 1-norm is better known as the *spectral norm* of f, i.e. $\|\widehat{f}\|_1 \coloneqq \sum_{\eta\in\mathbb{F}_2^n} |\widehat{f}(\eta)|$.

For an Abelian group, the convolution of two *real-valued* functions $f, g : G \to \mathbb{R}$ is a function defined to be

$$f * g(x) \coloneqq \mathop{\mathbb{E}}_{y\sim G}[f(y)g(x-y)].$$

It is a standard fact in Fourier analysis that the Fourier transformation of convolution of two functions is the pointwise product of their respective Fourier transforms: $\widehat{f * g} = \widehat{f} \cdot \widehat{g}$.

Lastly, *Parseval's identity* states that the squared Fourier 2-norm of a function f is equal to its second moment (under uniform probability measure):

$$\|\widehat{f}\|_2^2 = \sum_{\eta \in \mathbb{F}_2^n} \widehat{f}(\eta)^2 = \mathop{\mathbb{E}}_{x \sim \mathbb{F}_2^n}[f(x)^2].$$

3 The Proof Framework

The key proof technique throughout this work is Hölder's inequality with a nuanced choice of p-norms. Hölder's inequality states that for any $p, q \in [1, \infty]$ satisfying $\frac{1}{p} + \frac{1}{q} = 1$,

$$|\langle u, v \rangle| \leq \|u\|_p \|v\|_q \tag{5}$$

for any real vectors $u, v \in \mathbb{R}^n$. A generalization of Hölder's inequality known as *Littlewood's inequality* [15], which can be deduced from the standard Hölder's inequality, relates the p-norms of a function for some "interpolated" tuples of p. The inequality states that for $1 \leq p_0 < p_* < p_1 \leq \infty$,

$$\|v\|_{p_*} \leq \|v\|_{p_0}^{\theta} \|v\|_{p_1}^{1-\theta} \tag{6}$$

for any vector $v \in \mathbb{R}^n$, where $\theta \in (0, 1)$ satisfies $\frac{1}{p_*} = \frac{\theta}{p_0} + \frac{1-\theta}{p_1}$. As Schatten norms and Fourier norms are defined based on ℓ_p norms, the above inequality also holds with replacing the ℓ_p norms with these families of norms.

Setting the parameters $(p_*, p_0, p_1) = (2, 1, 4)$ (so that $\theta = 1/3$) for Equation (6) yields the following lower bound for 1-norm (after some algebraic manipulations):

$$\|f\|_1 \geq \frac{\|f\|_2^3}{\|f\|_4^2}. \tag{7}$$

As this is the sole parametrization of Hölder's inequality that we employ, we colloquially refer to Equation (7) as "the Hölder's inequality" in the rest of this work.

Among all the possible parametrizations for Littlewood's inequality, the choice made in Equation (7) of $(p_*, p_0, p_1) = (2, 1, 4)$ is appealing because of the nice physical interpretations of 2-norm and 4-norm for Boolean matrices (the Schatten norms) and Boolean functions (the Fourier norms). One main inspiration of Equation (7) as a practical bound is the work of Chazelle and Lvov on hereditary discrepancy [7]. Chazelle and Lvov proved a hereditary discrepancy lower bound of a matrix in terms of its Schatten 2-norm and 4-norm, highlighting the combinatorial interpretations of these norms for Boolean matrices. In the next section, we illustrate the combinatorial perspective of Equation (7) with a simpler proof of Theorem 3.

4 Simplified Proof of Communication Complexity Lower Bounds

It is customary to associate a Boolean matrix A with the biadjacency matrix of a bipartite graph $G = (U \cup V, E)$. In view of this, the square of Schatten 2-norm i.e. Frobenius norm of A is simply the number of edges of G. The Schatten 4-norm of A also admits a graph-theoretic interpretation:

$$\|A\|_{S_4}^4 = \mathrm{Tr}((A^\top A)^2) = \sum_{i,j \in U} \sum_{k,\ell \in V} A_{ik} A_{i\ell} A_{jk} A_{j\ell} = \sum_{i,j \in U} \sum_{k,\ell \in V} \mathbf{1}[\{ik, i\ell, jk, j\ell\} \subseteq E].$$

In other words, $\|A\|_{S_4}^4$ is precisely the count of possibly degenerate 4-cycles in G. This quantity becomes especially easy to evaluate when the underlying graph G does not contain non-degenerate 4-cycles (C_4-free), or equivalently, A does not contain a 2×2 all-one submatrix.

For a C_4-free bipartite graph, all contributions to $\|A\|_{S_4}^4$ come from the counts of edges and paths of length 2. Furthermore, if G is almost balanced in the sense that the average degree of G is close to the maximum degree, Equation (7) indeed yields a non-trivial lower bound for $\|A\|_{\mathrm{ntr}}$:

▶ **Proposition 11.** *Let $G = (U \cup V, E)$ be a C_4-free bipartite graph with maximum degree $\Delta(G)$. For $x \in U \cup V$, let d_x be the degree of x. If $A \in \{0,1\}^{U\times V}$ is the biadjacency matrix of G, Then*

$$\|A\|_{\mathrm{tr}} \geq \frac{|E|^{3/2}}{\sqrt{\sum_{x\in U\cup V} d_x^2}} \geq \frac{|E|}{\sqrt{2\Delta(G)}}.$$

Proof. By Hölder's inequality (Equation (7)),

$$\|A\|_{\mathrm{tr}} \geq \frac{\|A\|_{\mathrm{F}}^3}{\|A\|_{S_4}^2} = \frac{|E|^{3/2}}{\sqrt{\mathrm{Tr}((A^\top A)^2)}}.$$

Since G is C_4-free, the expression of $\mathrm{Tr}((A^\top A)^2)$ is simplified to

$$\begin{aligned}\mathrm{Tr}((A^\top A)^2) &= \sum_{i\in U}\sum_{k\in V} A_{ik} + \sum_{i\in U}\sum_{\substack{k,\ell\in V\\ k\neq\ell}} A_{ik}A_{i\ell} + \sum_{\substack{i,j\in U\\ i\neq j}}\sum_{k\in V} A_{ik}A_{jk}\\ &= |E| + \sum_{i\in U} d_i(d_i-1) + \sum_{k\in V} d_k(d_k-1)\\ &= \left(\sum_{x\in U\cup V} d_x^2\right) - |E|\end{aligned}$$

and the required bounds follow. ◀

Since every two distinct points define a unique line, the bipartite graph G associated with any point-line incidence system (with no duplicated lines) is C_4-free. From Proposition 11, one can readily derive an improved exponential normalized trace norm lower bound for $\mathtt{PL}$ using the same point-incidence submatrix in [8].

We prove a better bound by considering a point-line incidence system $\mathtt{PL}_*$ attributed to Paul Erdős (see [6, Lemma 6.25]). This point-line incidence system is constructed to show the tightness of the Szemerédi–Trotter incidence bound. Concretely, the $2^{n+1}\times 2^n$ matrix $\mathtt{PL}_*^{(n)} : \mathcal{P}'\times\mathcal{L}' \to \{0,1\}$ is defined as follows: the input domains are $\mathcal{P}' = [2^{n/3}]\times[2^{2n/3+1}]$ and $\mathcal{L}' = [2^{n/3}]\times[2^{2n/3}]$, and the entries are given by

$$\mathtt{PL}_*^{(n)}[(x,x'),(y,y')] = \begin{cases} 1 & \text{if } xy + y' = x'\\ 0 & \text{otherwise}\end{cases}.$$

We first show that $\mathtt{IIP}_3$ contains $\mathtt{PL}_*$ as a submatrix.

▷ Claim 12. The matrix $\mathtt{IIP}_3^{(n)}$ contains the matrix $\mathtt{PL}_*^{(3n/2-3/2)}$ as a submatrix.

Proof. The condition of incidence of $\mathtt{PL}_*$ can be rewritten as $xy + (1)y' + x'(-1) = 0$, which is an instance of integer inner product. For $\mathtt{PL}_*^{(3n/2-3/2)}$, the magnitude of each entry is bounded by $2^{\frac{2}{3}\left(\frac{3}{2}n-\frac{3}{2}\right)+1} = 2^n$, therefore $\mathtt{PL}_*^{(3n/2-3/2)}$ is a submatrix of $\mathtt{IIP}_3^{(n)}$. ◁

Proof of Theorem 4. As noted above, $\mathtt{PL}_*^{(m)}$ is C_4-free because no two distinct lines intersect at the same pair of points. For the matrix $\mathtt{PL}_*^{(m)}$, each line is incident to exactly one point of the form $(i, x') \in \mathcal{P}$ for each $i \in [2^{m/3}]$. Therefore the bipartite graph defined by $\mathtt{PL}_*$ contains $|\mathcal{L}| \cdot 2^{m/3} = 2^{4m/3}$ edges. Also, each point is incident to at most one line of the form $(j, y') \in \mathcal{L}$ for each $j \in [2^{m/3}]$, hence the maximum degree of the graph is $2^{m/3}$. By Proposition 11,

$$\left\|\mathtt{PL}_*^{(m)}\right\|_{\mathrm{ntr}} \geq \Omega\left(\frac{1}{2^m} \cdot \frac{2^{4m/3}}{\sqrt{2 \cdot 2^{m/3}}}\right) = \Omega\left(2^{m/6}\right).$$

By Claim 12 and Equation (1), we obtain

$$\mathsf{D}^{\mathsf{EQ}}\left(\mathtt{IIP}_k^{(n)}\right) \geq \mathsf{D}^{\mathsf{EQ}}\left(\mathtt{IIP}_3^{(n)}\right) \geq \mathsf{D}^{\mathsf{EQ}}\left(\mathtt{PL}_*^{(3n/2-3/2)}\right) \geq \frac{1}{2}\left(\frac{1}{6} \cdot \frac{3n}{2}\right) + O(1) = \frac{n}{8} + O(1). \quad \blacktriangleleft$$

As mentioned earlier, Chattopadhyay, Lovett, and Vinyals [4] proved a linear lower bound on $\mathsf{D}^{\mathsf{EQ}}(\mathtt{IIP}_k)$ for $k \geq 6$ by a method different from the approach in [8] and this work. The method used in [4], the relative area method, considered a parametrized weighted sum of monochromatic rectangle partition of a Boolean matrix. The following D^{EQ} lower bound is deduced from the communication complexity lower bound with more general oracle access.

▶ **Theorem 13** ([4, Lemmas 3.5 and 3.7])**.** *Suppose M is a Boolean matrix with α 1-entries, and the maximum area of any 1-monochromatic rectangle in M is β. For any constant $\eta \in (1/2, 1)$,*

$$\mathsf{D}^{\mathsf{EQ}}(M) = \Omega\left(\log\left(\frac{\alpha}{\beta^{1-\eta}|M|^{\eta}}\right)\right).$$

For the matrix $\mathtt{PL}_*^{(m)}$, as observed in the proof of Theorem 4, $\alpha = |E(G)| = 2^{4m/3}$ and $\beta = \Delta(G) = 2^{m/3}$ since $\mathtt{PL}_*^{(m)}$ is C_4-free. Applying Theorem 13 with $\eta = 1/2 + \epsilon'$ for some very small constant $\epsilon' > 0$, we obtain a linear D^{EQ} lower bound with a slightly inferior constant due to the slackness in η. We can also use this technique to lower bound N^{EQ}, which proves Theorem 5.

▶ **Theorem 14** ([19])**.** *Let M be a $2^m \times 2^m$ Boolean matrix, α be the number of 1-entries in M and let β be the size of the largest 1-monochromatic rectangle in M. Then*

$$\mathsf{N}^{\mathsf{EQ}}(M) = \Omega\left(\log\left(\frac{\alpha}{\sqrt{\beta} 2^m}\right) \cdot \frac{1}{\log m}\right).$$

Proof of Theorem 5. As noted before, $\alpha = |E(G)| = 2^{4m/3}$ and $\beta = \Delta(G) = 2^{m/3}$. For $m = 3n/2 - 3/2$, applying Theorem 14 gives

$$\mathsf{N}^{\mathsf{EQ}}\left(\mathtt{PL}_*^{(m)}\right) = \Omega\left(\frac{n}{\log n}\right)$$

and therefore the same lower bound holds for $\mathsf{N}^{\mathsf{EQ}}\left(\mathtt{IIP}_k^{(n)}\right)$ for $k \geq 3$. ◀

5 Proof of the Main Theorem

The Hölder's inequality supplies a large lower bound on 1-norm in the scenario of large 2-norm and small 4-norm. The proof of Theorem 3 utilizes the C_4-free property and sufficient edge density of $\mathtt{PL}$ to show the desired trace norm lower bound. We adopt the same strategy in exhibiting a function that the separation of randomized parity decision tree depth and spectral norm.

As in the matrix case, we begin by examining the physical interpretations of p-norms in the Hölder's inequality. A fundamental quantity in additive combinatorics known as the *additive energy*, coincides with the fourth power of Fourier 4-norm in Boolean function setting.

▶ **Definition 15** (Additive energy). *For $f : G \to \{0,1\}$ over an Abelian group G, the additive energy of f is defined as*

$$E(f) \coloneqq \mathop{\mathbb{E}}_{x,y,z\sim G}[f(x)f(y)f(z)f(x+y-z)].$$

For a Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$, tt is straightforward to show that $E(f) = \|\widehat{f}\|_4^4$:

$$\begin{aligned} E(f) &= \mathop{\mathbb{E}}_{t\sim\mathbb{F}_2^n}\left[\mathop{\mathbb{E}}_{x\sim\mathbb{F}_2^n}[f(x)f(t-x)] \cdot \mathop{\mathbb{E}}_{z\sim\mathbb{F}_2^n}[f(z)f(t-z)]\right] \\ &= \mathop{\mathbb{E}}_{t\sim\mathbb{F}_2^n}[(f*f(t))^2] \\ &= \sum_{\eta\in\mathbb{F}_2^n} \widehat{f*f}(\eta)^2 = \sum_{\eta\in G} \widehat{f}(\eta)^4 = \|\widehat{f}\|_4^4. \end{aligned}$$

It is also direct from Parseval's identity that $\|\widehat{f}\|_2^2 = \mathbb{E}_x[f(x)]$. Therefore in the Boolean function setting, Equation (7) states that

$$\|\widehat{f}\|_1 \geq \frac{\|\widehat{f}\|_2^3}{\|\widehat{f}\|_4^2} = \sqrt{\frac{\mathbb{E}_x[f(x)]^3}{E(f)}}. \tag{8}$$

Emulating the approach in Section 4 to guarantee a small 4-norm, we consider a Boolean function $f : \mathbb{F}_2^n \to \{0,1\}$ that satisfies $f(x)f(y)f(z)f(x+y+z) = 0$ for any distinct $x, y, z \in \mathbb{F}_2^n$ as an analogue of C_4-free bipartite graphs. For such a function, the additive energy of f is lower than typical as only $O(|\mathbb{F}_2^n|^2)$ tuples of $(x,y,z) \in (\mathbb{F}_2^n)^3$ could contribute to $E(f)$. A function with this property is precisely the indicator function of a Sidon set.

▶ **Definition 16** (Sidon set). *For an Abelian group G, a set $S \subseteq G$ is called a* Sidon *set if for every $a, b, c, d \in S$ such that $a + b = c + d$, then $\{a, b\} = \{c, d\}$.*

The indicator of a Sidon set emerges as a natural candidate function that attains the exponential spectral norm lower bound stated in Theorem 1. Indeed, one can show that the spectral norm of an indicator of a sufficiently big Sidon set is large.

▶ **Proposition 17.** *Every Sidon set S of a finite Abelian group G satisfies $\|\widehat{\mathbf{1}_S}\|_1 \geq \sqrt{|S|/2}$.*

Proof. Since S is a Sidon set, $\mathbf{1}_S(x)\mathbf{1}_S(y)\mathbf{1}_S(z)\mathbf{1}_S(x+y-z) = 1$ implies that $z = x$ or $z = y$. Thus

$$E(\mathbf{1}_S) \leq \frac{2}{|G|} \mathop{\mathbb{E}}_{x,y\sim G}[\mathbf{1}_S(x)\mathbf{1}_S(y)] = \frac{2}{|G|} \times \left(\frac{|S|}{|G|}\right)^2 = \frac{2|S|^2}{|G|^3}.$$

Applying Hölder's inequality (Equation (8)) gives

$$\|\widehat{\mathbf{1}_S}\|_1 \geq \sqrt{\frac{\mathbb{E}_x[\mathbf{1}_S(x)]^3}{E(\mathbf{1}_S)}} = \sqrt{\frac{|S|^3/|G|^3}{2|S|^2/|G|^3}} = \sqrt{\frac{|S|}{2}}. \qquad \blacktriangleleft$$

By Cauchy-Schwarz inequality and Parseval identity, it is direct to check that for any set $T \subseteq G$, the spectral norm of $\mathbf{1}_T$ is at most $\sqrt{|T|}$. The above proposition shows that this upper bound is tight up to constant factor for a Sidon set.

By considering the possible sums of a pair of elements, it is easy to see that for a Sidon set S, we have $\binom{|S|}{2} \leq |G|$ and hence $|S| = O(\sqrt{|G|})$. In the case of $G = \mathbb{Z}_2^n$ [3], a well-known construction of a Sidon set matching the size upper bound is the Bose–Chaudhuri–Hocquenghem (BCH) code [2, 16]. This code is constructed from polynomials over a finite field of characteristic 2.

Denote $\mathbb{F}_{2^m}$ the characteristic-2 finite field of size 2^m. The elements of $\mathbb{F}_{2^m}$ can be isomorphically identified with univariate $\mathbb{F}_2$-polynomials of degree at most $m-1$, where multiplication is defined modulus a fixed irreducible polynomial $P(x) \in \mathbb{F}_2[x]$ of degree m. Viewing as additive groups, $\mathbb{F}_{2^m} \cong \mathbb{F}_2[x]/\langle P(x)\rangle$ is isomorphic to $\mathbb{Z}_2^m$.

For $\mathbb{F}_p$ where p is a prime, the Sidon set construction in [2] is given by the tuples (a, a^k) over all a in a suitable subfield and a suitable exponent k. For $p = 2$, the construction takes the form of $\mathtt{BCH}(S)$, where for a set S we define

$$\mathtt{BCH}(S) \coloneqq \{(a, a^3) : a \in S\}.$$

For the sake of completeness, we include the proof for this specific construction.

▶ **Theorem 18** ([2]). *For every even number m, the set $\mathtt{BCH}(\mathbb{F}_{2^{m/2}}) \subseteq \mathbb{F}_{2^{m/2}} \times \mathbb{F}_{2^{m/2}} \equiv \mathbb{Z}_2^m$ is a Sidon set.*

Proof. Let (a, a^3) and (b, b^3) be two pairs with a prescribed sum $(u, v) \in \mathbb{F}_{2^{m/2}} \times \mathbb{F}_{2^{m/2}}$. This means $a + b = u$ and $a^3 + b^3 = v$, which implies that

$$u^3 = (a+b)^3 = a^3 + b^3 + (a+b)ab = v + uab.$$

Suppose $a \neq b$ i.e. $u \neq 0$, then $a(u+a) = ab = u^2 + vu^{-1}$. The same calculation concludes that a and b are the roots of the quadratic equation $x(x+u) = u^2 + vu^{-1}$. Since a quadratic equation has at most two roots, $\{a, b\}$ is uniquely determined by u and v. ◀

Theorem 18 allows us to construct a large Sidon set in $\mathbb{Z}_2^n$. Since a subset of a Sidon set remains a Sidon set, it remains to select a suitably structured subset whose indicator function is computable with a short randomized parity decision tree. As mentioned earlier, every element in $\mathbb{F}_{2^n}$ can be uniquely identified with a polynomial in $\mathbb{F}_2[x]$ of degree at most $n-1$. For $d \in \mathbb{N}$, denote $\mathcal{P}_d$ the set of $\mathbb{F}_2$-polynomials of degree at most $d-1$. We show that $\mathtt{BCH}(\mathcal{P}_{n/4})$ is the desired Sidon set.

▶ **Theorem 19.** *Let $n = 4d$ for some $d \in \mathbb{N}$, and $S = \mathtt{BCH}(\mathcal{P}_d) \subseteq \mathcal{P}_d \times \mathcal{P}_{3d} \subseteq \mathbb{F}_{2^d} \times \mathbb{F}_{2^{3d}} \cong \mathbb{Z}_2^n$. The Boolean function $\mathbf{1}_S : \mathbb{F}_2^n \to \{0,1\}$ satisfies that $\mathsf{RPDT}_{\mathrm{depth}}(\mathbf{1}_S) = O(\log n)$ and $\|\widehat{\mathbf{1}_S}\|_1 = \Omega(2^{n/8})$.*

Proof. Note that S is a Sidon set of size $|\mathcal{P}_d| = 2^d = 2^{n/4}$. The spectral norm lower bound follows from Proposition 17. It remains to upper-bound the randomized parity decision tree complexity.

We first describe a randomized algorithm that computes $\mathbf{1}_S$. Suppose a pair $(u, v) \in \mathbb{F}_2[x] \times \mathbb{F}_2[x]$ with $\deg(u) < d$ and $\deg(v) < 3d$ is a given. To determine whether $(u, v) \in S$, we interpret $v - u^3$ as a polynomial in $\mathbb{F}_{2^m}[x]$, where $m = \lceil \log(3d) \rceil + 2$, and check whether $v - u^3 \equiv 0$. This can be accomplished using the standard randomized polynomial identity

[3] For clarity, in the rest of this section, we use the notation $\mathbb{Z}_2$ to refer to the additive group of size 2.

testing algorithm: pick a random $t \in \mathbb{F}_{2^m}$ and evaluate $v(t) - u(t)^3 \in \mathbb{F}_{2^m}$. The algorithm declares that "$v \neq u^3$" if any of the evaluated values is non-zero, and declares "$v = u^3$" otherwise.

Notice that this randomized algorithm has a one-sided error, as it always makes the correct declaration when $v = u^3$ (i.e. $\mathbf{1}_S(u, v) = 1$). In the case of $v \neq u^3$, since the polynomial $v - u^3$ has at most $3d$ roots, the probability of error is at most $3d/2^m \leq 1/4$.

Next, we convert the randomized algorithm into a randomized parity decision tree. For every $t \in \mathbb{F}_{2^m}$, the set

$$H_t \coloneqq \{u \in \mathcal{P}_d : u(t) = 0\} \subseteq \mathcal{P}_d \cong \mathbb{Z}_2^d$$

is a linear subspace of co-dimension m in $\mathbb{Z}_2^d$, and each coset of H_t takes a fixed value when evaluated at t. Therefore, the value of $u(t)$ is determined by m deterministic linear queries to u. Similarly, the value of $v(t)$ is determined by m deterministic linear queries to v. Hence, we construct the randomized parity decision tree as follows: pick a random $t \in \mathbb{F}_{2^m}$ and make the $2m$ parity queries that determine the values of $u(t), v(t) \in \mathbb{F}_{2^m}$, and the decision tree outputs whether $v(t)$ and $u(t)^3$ are equal. ◀

5.1 Payley-Zygmud inequality and typical Fourier coefficients

Here instead of Hölder's inequality, we use the Payley-Zygmund inequality to deduce a stronger conclusion than Proposition 17 which could be of independent interest. We show that for the indicator of a Sidon set, a constant fraction of the Fourier coefficients is large. We have shown that for a Sidon set S in an Abelian group G,

$$\sum_{\chi \in \widehat{G}} |\widehat{\mathbf{1}_S}(\chi)|^2 = \frac{|S|}{|G|} \quad \text{and} \quad \sum_{\chi \in \widehat{G}} |\widehat{\mathbf{1}_S}(\chi)|^4 \leq \frac{2|S|^2}{|G|^3}.$$

Consider the random variable $X = |\widehat{\mathbf{1}_S}(\chi)|^2$ where χ is a character picked uniformly at random. The above bounds translate to $\mathbb{E}[X] = |S|/|G|^2$ and $\mathbb{E}[X^2] = 2|S|^2/|G|^4 = 2\,\mathbb{E}[X]^2$. By Paley-Zygmund inequality, we have for any $\delta \in (0, 1)$,

$$\Pr[X > \delta\,\mathbb{E}[X]] \geq (1 - \delta)^2 \frac{(\mathbb{E}[X])^2}{\mathbb{E}[X^2]} \geq \frac{(1 - \delta)^2}{2}.$$

Taking $\delta = 1/2$ for example, this gives

$$\Pr_{\chi}\left[|\widehat{\mathbf{1}_S}(\chi)| > \frac{\sqrt{|S|}}{\sqrt{2}|G|}\right] \geq \frac{1}{8}.$$

References

1 Nikhil Bansal and Makrand Sinha. k-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, pages 1303–1316, New York, NY, USA, 2021. Association for Computing Machinery. doi:10.1145/3406325.3451040.

2 R.C. Bose and D.K. Ray-Chaudhuri. On a class of error correcting binary group codes. *Information and Control*, 3(1):68–79, March 1960. doi:10.1016/s0019-9958(60)90287-4.

3 Arkadev Chattopadhyay, Yogesh Dahiya, Nikhil S. Mande, Jaikumar Radhakrishnan, and Swagato Sanyal. Randomized versus deterministic decision tree size. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC '23. ACM, June 2023. doi:10.1145/3564246.3585199.

4 Arkadev Chattopadhyay, Shachar Lovett, and Marc Vinyals. Equality alone does not simulate randomness. In *34th Computational Complexity Conference (CCC 2019)*, 2019. doi:10.4230/LIPICS.CCC.2019.14.

5 Arkadev Chattopadhyay, Nikhil S. Mande, and Suhail Sherif. The log-approximate-rank conjecture is false. *J. ACM*, 67(4):23:1–23:28, 2020. doi:10.1145/3396695.

6 Bernard Chazelle. *The discrepancy method: randomness and complexity*. Cambridge University Press, 2001.

7 Bernard Chazelle and Alexey Lvov. A trace bound for the hereditary discrepancy. In *Proceedings of the sixteenth annual symposium on Computational geometry*, pages 64–69, 2000. doi:10.1145/336154.336179.

8 Tsun-Ming Cheung, Hamed Hatami, Kaave Hosseini, and Morgan Shirley. Separation of the factorization norm and randomized communication complexity. In *38th Computational Complexity Conference (CCC 2023)*. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPICS.CCC.2023.1.

9 Tsun-Ming Cheung, Hamed Hatami, Rosie Zhao, and Itai Zilberstein. Boolean functions with small approximate spectral norm. *Electron. Colloquium Comput. Complex.*, TR22-041, 2022. URL: https://eccc.weizmann.ac.il/report/2022/041, arXiv:TR22-041.

10 Kenneth R. Davidson and Allan P. Donsig. Norms of schur multipliers. *Illinois Journal of Mathematics*, 51(3), July 2007. doi:10.1215/ijm/1258131101.

11 Uma Girish, Avishay Tal, and Kewen Wu. Fourier Growth of Parity Decision Trees. In *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 39:1–39:36, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2021.39.

12 Mika Göös, Toniann Pitassi, and Thomas Watson. The landscape of communication complexity classes. *computational complexity*, 27(2):245–304, June 2018. doi:10.1007/s00037-018-0166-6.

13 Ben Green and Tom Sanders. Boolean functions with small spectral norm. *Geometric and Functional Analysis*, 18(1):144–162, March 2008. doi:10.1007/s00039-008-0654-y.

14 Lianna Hambardzumyan, Hamed Hatami, and Pooya Hatami. Dimension-free bounds and structural results in communication complexity. *Israel Journal of Mathematics*, 253(2):555–616, 2023.

15 G H Hardy, J E Littlewood, and Georg Polya. *Cambridge mathematical library: Inequalities*. Cambridge University Press, Cambridge, England, February 1988.

16 Alexis Hocquenghem. Codes correcteurs d'erreurs. *Chiffres*, 2:147–156, 1959.

17 Eyal Kushilevitz and Yishay Mansour. Learning decision trees using the fourier spectrum. In *Proceedings of the Twenty-Third Annual ACM Symposium on Theory of Computing*, STOC '91, pages 455–464, New York, NY, USA, 1991. Association for Computing Machinery. doi:10.1145/103418.103466.

18 Noam Nisan. Crew prams and decision trees. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 327–335, 1989. doi:10.1145/73007.73038.

19 Toniann Pitassi, Morgan Shirley, and Adi Shraibman. The strength of equality oracles in communication. In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, volume 251 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 89:1–89:19, Dagstuhl, Germany, 2023. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.ITCS.2023.89.

20 Tom Sanders. Boolean functions with small spectral norm, revisited. *Math. Proc. Camb. Philos. Soc.*, 167(02):335–344, September 2019.

21 Amir Shpilka, Avishay Tal, and Ben lee Volk. On the structure of boolean functions with small spectral norm. *Computational Complexity*, 26(1):229–273, September 2017. doi:10.1007/s00037-015-0110-y.

22 Avishay Tal. Tight Bounds on the Fourier Spectrum of AC0. In *32nd Computational Complexity Conference (CCC 2017)*, volume 79 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 15:1–15:31, Dagstuhl, Germany, 2017. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2017.15.

23 Hing Yin Tsang, Chung Hoi Wong, Ning Xie, and Shengyu Zhang. Fourier sparsity, spectral norm, and the log-rank conjecture. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 658–667, 2013. doi:10.1109/FOCS.2013.76.