Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 26 (2019)

LOWER BOUNDS ON BALANCING SETS AND DEPTH-2 THRESHOLD CIRCUITS

PAVEL HRUBEŠ,
SIVARAMAKRISHNAN NATARAJAN RAMAMOORTHY,
ANUP RAO, AND AMIR YEHUDAYOFF

Abstract. There are various notions of balancing set families that appear in combinatorics and computer science. For example, a family of proper non-empty subsets $S_1, \ldots, S_k \subset [n]$ is balancing if for every subset $X \subset \{1, 2, \ldots, n\}$ of size $n/2$, there is $i \in [k]$ so that $|S_i \cap X| = |S_i|/2$. We extend and simplify the framework developed by Hegedűs for proving lower bounds on the size of balancing set families. We prove that if $n = 2p$ for a prime p, then $k \geq p$. For arbitrary values of n, we show that $k \geq n/2 - o(n)$.

We then exploit the connection between balancing families and depth-2 threshold circuits. This connection helps resolve a question raised by Kulikov and Podolskii on the fan-in of depth-2 majority circuits computing the majority function on n bits. We show that any depth-2 threshold circuit that computes the majority on n bits has at least one gate with fan-in at least $n/2 - o(n)$. We also prove a sharp lower bound on the fan-in of depth-2 threshold circuits computing a specific weighted threshold function.

1. Introduction

1.1. **Balancing Families.** *Balancing set families* are families of proper non-empty subsets of a finite universe that satisfy a *discrepancy* type property. They are well studied objects in combinatorics [12, 10, 4, 15, 5, 14], and they have found many applications in computer science [4, 20, 16, 5, 14]. In this work we prove new lower bounds on the size of such families, and then use them to prove lower bounds on depth-2 *majority* and *threshold circuits* that compute the majority and *weighted threshold* functions. We establish new sharp lower bounds on the *fan-in* of the gates in such circuits.

This work was done while the authors were visiting the Simons Institute for the Theory of Computing. P. H. is supported by ERC grant FEALORA 339691 and the GACR grant 19-27871X. S. N. R. and A. R. are supported by the National Science Foundation under agreement CCF- 1420268. A.Y. is partially supported by ISF grant 1162/15.

ISSN 1433-8092

A central contribution of this work is the following lemma that shows a lower bound on the degree of a special class of polynomials.

Lemma 1. *Let p be prime, and let $f(x_1, \ldots, x_{2p})$ be a polynomial over $\mathbb{F}_p$, where $\mathbb{F}_p$ is the field with p elements. Let f be such that for every input $x \in \{0,1\}^{2p}$ with exactly p ones, we have $f(x) = 0$, and $f(x)$ is non-zero when $x_1 = x_2 = \ldots = x_{2p} = 0$. Then, the degree of f is at least p.*

Hegedűs [15] used a similar lemma to prove lower bounds for balancing sets (in his lemma there are $4p$ variables, and the focus is on inputs with $3p$ ones). Hegedűs's proof uses Gröbner basis methods and linear algebra. Srinivasan found a simpler proof of Hegedűs's lemma that is based on Fermat's little theorem and linear algebra. Alon [3] gave an alternate proof of Hegedűs's lemma using the Combinatorial Nullstellensatz. The above lemma is inspired by Srinivasan's proof of Hegedűs's lemma [21, 5]. Our simple proof is presented in Section 3.

For a positive integer n, let $[n]$ denote the set $\{1, 2, \ldots, n\}$. Various notions of balancing set families have been considered in the past [12, 10, 4, 15, 5] with various terminologies. We use the following definition in this work.

Definition 1. *Let k be a positive integer and n be a positive even integer. We say that proper non-empty subsets $S_1, \ldots, S_k \subset [n]$ are a* balancing set family *if for every $X \subset [n]$ of size $n/2$ there is an $i \in [k]$ such that $|S_i \cap X| = |S_i|/2$.*

Given any even n, let $\mathsf{B}(n)$ denote the minimum k for which a balancing set family of size k exists. Our first result gives tight bounds on $\mathsf{B}(n)$:

Theorem 2. *If $n = 2p$ for a prime p, then $\mathsf{B}(n) = n/2 = p$.*

Moreover, if n is divisible by 4, we give an example of a balancing set family establishing that $\mathsf{B}(n) \leq n/2 - 1$. If n is divisible by 2, we show that $\mathsf{B}(n) \leq n/2$ by constructing a balancing set family of size $n/2$, in which each set is of size 2. We also show that this is tight when each set in the family is of size 2 (see Theorem 13 in Section 4). Previously, for arbitrary values of n, Alon, Kumar and Volk [5] showed that $\mathsf{B}(n) \geq \Omega(n)$. We show

Theorem 3. *If n is an even integer, then $\mathsf{B}(n) \geq n/2 - O\left(n^{0.98}\right)$.*

Our lower bounds on $\mathsf{B}(n)$ are the most interesting and they are proved using Lemma 1. See Section 4 and Section 5 for a full exposition of the proofs. We also

apply our techniques to other questions about balancing sets in the literature and improve some of the previous bounds. We now briefly discuss two such notions from the literature.

(a) Galvin's question [12, 10, 15] asks for the smallest balancing family, denoted by $\mathsf{G}(n)$, where each set in the family is of size $n/2$, and n is a positive integer that is a multiple of 4.

(b) Jansen [16] and Alon, Kumar, and Volk [5] studied a variant where the size of each set in the family must satisfy $2\tau \leq |S_i| \leq n - 2\tau$ for a positive integer τ, and for every $X \subset [n]$ of size $n/2$, there is a set in the family such that $|S_i|/2 - \tau < |S_i \cap X| < |S_i|/2 + \tau$. Denote by $\mathsf{J}(n, \tau)$ to be the family of smallest size satisfying the above conditions.

We defer the discussion of previous known bounds on the quantities $\mathsf{G}(n)$ and $\mathsf{J}(n, \tau)$ to Section 2. We prove the following lower bounds on $\mathsf{G}(n)$ and $\mathsf{J}(n, \tau)$.

Theorem 4. *If n is divisible by 4, then $\mathsf{G}(n) \geq n/2 - O\left(n^{0.53}\right)$.*

Theorem 5.

(1) If $n = 2p$ for a prime p then $\mathsf{J}(n, \tau) \geq \frac{n}{4\tau - 2}$.
(2) $\mathsf{J}(n, \tau) \geq \frac{n - O\left(n^{0.98}\right)}{7\tau}$.

We proceed to define the notion of unbalancing set families used in this work.

Definition 2. *Let n be a positive even integer, and $k \geq 0, 0 \leq t \leq n/2$ be integers. We say that subsets $S_1, \ldots, S_k \subset [n]$ are an* unbalancing set family *if for every $X \subset [n]$ of size $n/2 - t$, there is an $i \in [k]$ such that $|S_i \cap X| > |S_i|/2$.*

Given any even n, let $\mathsf{U}(n, t)$ denote the minimum k for which an unbalancing set family of size k exists. For unbalancing set families, we determine $\mathsf{U}(n, t)$ exactly:

Theorem 6. $\mathsf{U}(n, t) = 2t + 2$.

Again, the lower bound here is more interesting than the upper bound. It is proved by showing a connection between $\mathsf{U}(n, t)$ and the chromatic number of an appropriately defined Kneser graph [18].

1.2. **Threshold Circuits.** We now discuss our results on depth-2 majority and threshold circuits. The majority function, $\mathsf{MAJ}(x)$ for $x \in \{0,1\}^n$, is defined as

$$\mathsf{MAJ}(x_1, \ldots, x_n) = \begin{cases} 1 & \sum_{i=1}^n x_i \geq n/2, \\ 0 & \text{otherwise.} \end{cases}$$

The unweighted threshold function, $T_t(x)$ for $x \in \{0,1\}^n$, is defined as

$$T_t(x_1, \ldots, x_n) = \begin{cases} 1 & \sum_{i=1}^n x_i \geq t, \\ 0 & \text{otherwise,} \end{cases}$$

for some non-negative integer t. In the rest of the paper, unless stated otherwise, we refer to threshold functions when we mean unweighted threshold functions.

A depth-2 circuit is defined by boolean functions $h, g_1, \ldots, g_k$, for some integer k, and the depth-2 circuit is said to compute a function f on input $x \in \{0,1\}^n$ if

$$f(x) = h(g_1(x), \ldots, g_k(x)).$$

Here $h, g_1, \ldots, g_k$ are called the *gates* of the circuit. h is referred to as the *top* gate, and $g_1, \ldots, g_k$ are referred to as the *bottom* gates of the circuit. Our lower bounds often hold even when h is allowed to be an arbitrary boolean function. The *fan-in* of a gate in the circuit measures the number of variables that need to be read for the gate to carry out its computation. The fan-in of the top gate in the circuit is defined to be k. The fan-in of each of the gates g_i is r_i if g_i depends on r_i of the input variables. We sometimes refer to the top fan-in when we mean k and the bottom fan-in when we mean the maximum of $r_1, \ldots, r_k$. We say that the fan-in of the circuit is r, if r is the maximum of the top fan-in and bottom fan-in.

When functions $g_1, \ldots, g_k, h$ each compute majority, the circuit is called a majority circuit. Similarly, if all gates compute thresholds, then the circuit is called a threshold circuit. Kulikov and Podolskii [17] asked the following question: What is the minimum fan-in required to compute majority using a depth-2 majority circuit? Balancing set families are closely related to depth-2 majority circuits computing majority. One can prove that there is a depth-2 majority circuit computing majority of n bits with top fan-in at most $2 \cdot \mathsf{B}(n) + 2$, when n is even. Indeed, let $S_1, \ldots, S_k$ be the balancing set family. Define k majority gates, each on variables indexed by S_i, and another k majority gates, each on variables indexed by $[n] \setminus S_i$. The top majority gate, with fan-in $2k + 2$, reads

these $2k$ gates along with two 0 inputs. It is easy to see that this circuit correctly computes the majority.

To obtain a lower bound on the fan-in of such circuits, a potential approach is to show that every depth-2 majority or threshold circuit corresponds to a balancing set family. We are able to leverage the ideas that are used to prove Theorem 2 to obtain lower bounds on the fan-in of these circuits. Moreover, our lower bounds are sharp up to a constant factor.

Let $n = 2p$ for a prime p. Note that the threshold function defined by the inequality $\sum_{i=1}^{n} x_i \geq p$ is the majority function on n bits, and yields a circuit with top fan-in 1. We prove a lower bound on the top fan-in of a depth-2 threshold circuit when the bottom gates do not have the threshold p:

Theorem 7. *Suppose that $n = 2p$ for a prime p. Then in any depth-2 circuit computing the majority of n bits, if the bottom gates compute unweighted thresholds and read no constants, either the top fan-in is at least $n/2 = p$, or some gate at the bottom computes a threshold T_t with $t = p$.*

In fact, Theorem 7 implies a similar lower bound on the top fan-in when the bottom threshold gates read constants - see Section 6. Observe that in Theorem 7 we do not assume that the top gate h computes a threshold function. The lower bound holds with no restrictions on h.

Theorem 7 also gives tight lower bounds for the fan-in of threshold circuits computing majority. Firstly, any non-constant threshold function T_t reading at most r inputs must have $t \leq r$. Secondly, any bottom gate that computes a threshold function T_t by reading constants is equivalent to computing a threshold function $T_{t'}$ on the same input variables, for some $t' \leq t$, and $T_{t'}$ reads no constants. Here, $t' = t - \alpha$ where α is the number of ones read by T_t. Consequently, we get:

Corollary 8. *Suppose that $n = 2p$ for a prime p. Then in any depth-2 circuit computing the majority of n bits, if the bottom gates compute unweighted thresholds, the fan-in of the circuit must be at least $n/2 = p$.*

Since majority is a special case of the threshold function, the above corollary implies the same lower bound on the fan-in of majority circuits that compute the majority. However, by directly invoking Theorem 7, we obtain a slightly stronger lower bound for majority circuits computing the majority:

Corollary 9. *Suppose that $n = 2p$ for a prime p. Then in any depth-2 majority circuit computing the majority of n bits, either the bottom fan-in is more than $2p - 2 = n - 2$ or the top-fan in is at least $p = n/2$.*

This is because when the bottom fan-in of the majority circuit is at most $2p-2$, the threshold of bottom gates are at most $p - 1$ and Theorem 7 applies.

Theorem 7, Corollary 8 and Corollary 9 discuss the case when $n = 2p$ for a prime p. For arbitrary values of n, we can generalize Theorem 7 to show that either the top fan-in is at least $n/2 - o(n)$ or some gate at the bottom computes a threshold T_t with $t \geq p$, where p is the largest prime such that $p \leq n/2$ (see Section 6 for the proof). Naturally, this lower bound translates to Corollary 8 and Corollary 9. In particular, we get that any depth-2 majority circuit computing the majority of n bits must have that either the bottom fan-in at least $n - o(n)$ or the top fan-in at least $n/2 - o(n)$. This nearly matches Amano's [6] construction of a depth-2 majority circuit with bottom fan-in $n - 2$ and top fan-in $n/2 + 2$.

Another kind of result that we investigate is whether *weighted* threshold functions can be computed using unweighted thresholds of low fan-in. To that end, let $n = (3p - 1)/2$ for an odd prime p, and consider the weighted threshold function

$$T(x) = \begin{cases} 1 & \text{if } \sum_{i \leq p-1} x_i + 2 \sum_{i > p-1} x_i \geq p, \\ 0 & \text{otherwise.} \end{cases}$$

$T(x)$ is a weighted threshold function with weights 1 and 2.

Theorem 10. *Any depth-2 circuit computing $T(x)$ where the bottom gates compute unweighted thresholds must have top fan-in at least $(p - 1)/2 = (n - 1)/3$.*

Observe that in Theorem 10 we do not assume an upper bound on the fan-in of the bottom gates. Our bounds are much stronger and significantly simpler than past lower bounds ([17, 9]) on such circuits. Our proofs of Theorem 2 and Theorem 7 are based on proving lower bounds on the degree of specific polynomials, using Lemma 1, that are constructed using the balancing set families and depth-2 threshold circuits, respectively.

Table 1 and Table 2 summarize all our results discussed in the introduction.

Outline. The rest of the paper is organized as follows. We discuss related work in Section 2. We prove Lemma 1 in Section 3. Theorem 2 is proved in Section 4, and

Balancing Sets	$\mathsf{B}(n) = n/2$ when $n = 2p$	Theorem 2
	$\mathsf{B}(n) \geq n/2 - o(n)$	Theorem 3
	$\mathsf{G}(n) \geq n/2 - o(n)$	Theorem 4
	$\mathsf{J}(n, \tau) \geq n/(4\tau - 2)$ when $n = 2p$	Theorem 5
	$\mathsf{J}(n, \tau) \geq n(1 - o(1))/7\tau$	Theorem 5
Unbalancing Sets	$\mathsf{U}(n, t) = 2t + 2$	Theorem 6

TABLE 1. Summary of results on balancing and unbalancing families. p is a prime.

Function	**Bottom Gates**	**Result**	
Majority	thresholds and reads no constants	$k \geq n/2$ or threshold $= p$ when $n = 2p$	Theorem 7
Majority	thresholds	$\max\{k, r\} \geq n/2$ when $n = 2p$	Corollary 8
Majority	majority	$k \geq n/2$ or $r > n - 2$ when $n = 2p$	Corollary 9
Majority	thresholds	$k \geq n/2 - o(n)$ or threshold $\geq \mu(n/2)$	Theorem 18
Majority	thresholds	$\max\{k, r\} \geq n/2 - o(n)$	Corollary 19
$T(x)$	unbounded fan-in thresholds	$k \geq (n-1)/3$	Theorem 10

TABLE 2. Summary of results on depth-2 circuits. n is the number of input bits and p is a prime. k is the top fan-in and r is the maximum fan-in of the bottom gates. $\mu(n)$ denotes the largest prime that is no more than n.

the application of our techniques to generalizations of balancing set families are discussed in Section 5. In particular, Section 5 contains the proofs of Theorem 3, 4 and 5. Theorems 7 and 10 are proved in Sections 6 and 7 respectively. Theorem 6 is proved in Section 8.

Notation. $\mathbb{F}_p$ denotes the field with p elements, where p is a prime. For a positive integer n, $\mu(n)$ denotes the largest prime p so that $p \leq n$. For a natural number n, $[n]$ denotes the set $\{1, 2, \ldots, n\}$. For every $x \in \{0,1\}^n$ and $i \in [n]$, x_i denotes the i'th coordinate of x. For $x \in \{0,1\}^n$, when $x_1 = x_2 = \ldots = x_n = 0$, we refer to x as the all-zeros vector or the all-zeros input. The all-ones vector or all-ones input is defined similarly.

Bounds on $\mu(n)$. Generalizations of Theorems 2 and 7 to the case when $n \neq 2p$ for a prime p are obtained by using a known lower bound on $\mu(n)$. Baker, Harman and Pintz [8] showed that the largest gap between consecutive primes is bounded by $O(n^{0.53})$. As a consequence, we can conclude that

Theorem 11. [8] $\mu(n) \geq n - O(n^{0.53})$.

2. Related Work

2.1. **Balancing Families.** Various notions of balancing set families have been studied. We first describe the question posed by Galvin [12, 10, 15].

Definition 3. *Let n be a positive integer that is divisible by 4. A family of proper subsets $S_1, \ldots, S_k \subset [n]$ is* exactly balancing *if each S_i is of size $n/2$ and for every $X \subset [n]$ of size $n/2$ there is an $i \in [k]$ such that $|X \cap S_i| = |S_i|/2$.*

When n is divisible by 4, let $\mathsf{G}(n)$ denote the minimum k for which an exactly balancing set family of size k exists. Clearly, the family of all subsets of $[n]$ of size $n/2$ is exactly balancing, and any family with only one set is not exactly balancing. Therefore finding the minimum number of sets in any exactly balancing set family is interesting.

Galvin [12] observed that $\mathsf{G}(n) \leq n/2$; take $n/2$ consecutive intervals of length $n/2$. Frankl and Rödl [12] proved that $\mathsf{G}(n) \geq \Omega(n)$ if $n/4$ is odd, and later Enomote, Frankl, Ito and Nomura [10] proved that if $n/4$ is odd, then $\mathsf{G}(n) \geq n/2$. Proofs in [12, 10] are based on techniques from linear algebra and extremal set theory. Recently, Hegedűs [15] used algebraic techniques to proved that if $n/4$ is prime, then $\mathsf{G}(n) \geq n/4$. For arbitrary values of n, Alon, Kumar and Volk [5] proved that $\mathsf{G}(n) \geq \Omega(n)$. Theorem 4 improves the bound of Alon, Kumar and Volk.

Several natural variants of Galvin's problem have been studied. One such variant was studied by Jansen [16], and Alon, Kumar and Volk [5]:

Definition 4. *Let n be an even integer, and let τ be a positive integer. Let $S_1, \ldots, S_k \subset [n]$ with $2\tau \leq |S_i| \leq n - 2\tau$. We say that $S_1, \ldots, S_k$ is a τ-*balancing set family *if for every $X \subset [n]$ of size $n/2$ there is an $i \in [k]$ such that*

$$|S_i|/2 - \tau < |X \cap S_i| < |S_i|/2 + \tau.$$

When n is even and τ is positive, let $\mathsf{J}(n,\tau)$ denote the minimum k for which such a family of size k exists. This variant allows the family to have sets with different sizes and the intersection sizes to take more than just one value. Alon, Kumar and Volk proved that $\mathsf{J}(n,\tau) \geq \frac{1}{10^5} \cdot (n/\tau)$. This lower bound is sharp up to a constant factor. Theorem 5 improves their bound to $\frac{n-o(n)}{7\tau}$.

Our techniques yield a quantitatively stronger lower bound on balancing set families. The improvement stems from the fact that the ratio of the degree of the polynomial to the number of variables of the polynomial increases from 1/4 to 1/2. Moreover, the application of Lemma 1 eliminates an additional argument using the probabilistic method employed in the work of Alon, Kumar and Volk.

There are many applications of balancing set families. Alon, Bergmann, Coppersmith and Odlyzko [4] studied a different version of balancing sets that has applications to optical data communication. Jansen [16] and Alon, Kumar, and Volk [5] showed applications to proving lower bounds for syntactic multilinear algebraic circuits (also see [20]).

2.2. **Threshold Circuits.** A depth-d majority circuit can be defined in analogy to depth-2 majority circuits. Let $\mathsf{M}_d(n)$ denote the minimum fan-in of a depth-d majority circuit that computes the majority of n bits. A long line of work has addressed the question of computing the majority function using majority circuits. Ajtai, Komlós and Szemerédi [1] showed that $\mathsf{M}_{c\cdot\log n}(n) = O(1)$, for some constant c. Using probabilistic arguments, Valiant [22] showed the existence of depth $O(\log n)$ majority circuit that computes the majority, where each gate has constant fan-in. Allender and Koucky [2] showed that $\mathsf{M}_c(n) = O(n^{\epsilon(c)})$, where c is a constant and $\epsilon(c)$ is a function of c. Kulikov and Podolskii proved that $\mathsf{M}_3(n) \leq \tilde{O}\left(n^{2/3}\right)$[1]. See [17, 9, 11] and references within for a detailed treatment.

We now discuss previous bounds on $\mathsf{M}_2(n)$. Kulikov and Podolskii [17] used probabilistic arguments to show that $\mathsf{M}_2(n) \geq \tilde{\Omega}\left(n^{7/10}\right)$. They also proved that $\mathsf{M}_2(n) \geq \tilde{\Omega}\left(n^{13/19}\right)$ when the gates are not required to read distinct variables. Amano and Yoshida [7] showed that for every odd $n \geq 7$, $\mathsf{M}_2(n) \leq n-2$, where they allowed some of the gates to read variables multiple times. Later, Engles, Garg, Makino and Rao [9] used ideas from discrepancy theory to prove that $\mathsf{M}_2(n) \geq \Omega(n^{4/5})$ when the gates do not read constants. Posobin [19] showed

[1]In the rest of the paper, $\tilde{O}(a)$ and $\tilde{\Omega}(a)$ mean that $\mathsf{polylog}(a)$ factors are ignored.

that majority can be computed by a depth-2 majority circuit of fan-in at most $2n/3 + 4$ (this was also proved independently by Bauwens [19]). Very recently, Amano [6] gave a construction of a depth-2 majority circuit computing majority with bottom fan-in $n - 2$ and top fan-in $n/2 + 2$.

Kulikov and Podolskii [17] studied and proved lower bounds on other variants of depth-2 majority circuits. In particular, they consider circuits in which each majority gate can read a variable multiple times. Let W be the maximum over the number of times a variable is read. They prove that $\mathsf{M}_2(n) \geq \min\left\{\tilde{\Omega}\left(n^{13/19}\right), \tilde{\Omega}\left(\frac{n^{7/10}}{W^{3/10}}\right)\right\}$. In this case, our techniques yield a lower bound of $\mathsf{M}_2(n) \geq \Omega\left(\frac{n}{W}\right)$. Essentially, their lower bound is stronger when $W \geq n^{6/19}$ and our bound is stronger when $W \leq n^{6/19}$.

The question of computing weighted thresholds using a depth-2 threshold function is connected to the study of exact threshold circuits initiated by Hansen and Podolskii [13]. It may also be useful in studying the expressibility of general functions using threshold or *ReLu* gates; see the work of Williams [23].

We would like to emphasize that the lower bounds in Theorems 7 and 10 are tight and only off by constant factors. In addition, most functions considered in past work on majority and threshold circuit lower bounds do not admit depth-2 majority or threshold circuits with linear fan-in on the gates. In fact, one can prove exponential lower bounds on the size of circuits computing these functions (see [13]).

3. Proof of Lemma 1

Let f be as in the assumption of Lemma 1. Consider the polynomial

$$g(x_1, \ldots, x_{2p}) = (1 - x_1) \cdot \prod_{i=1}^{p-1} \left(i - \sum_{i=1}^{2p} x_i \right),$$

which has degree p. For $x \in \{0,1\}^{2p}$, observe that $g(x) = 0$ if the number of ones in x is not a multiple of p or x is the all-ones input, and $g(x) \neq 0$ if x is the all-zeros input. Therefore, $f \cdot g$ is non-zero on the all-zeros input and 0 elsewhere in $\{0,1\}^{2p}$.

We will now show that the degree of $f \cdot g$ is at least $2p$. Consider the polynomial h that is obtained by multilinearizing $f \cdot g$. In other words, replace every power x_i^k with x_i in $f \cdot g$, for $k \geq 1$, Observe that the degree of h is at most the degree of f.

Define $\alpha = h(0, \ldots, 0)$. Recall that there is a one-to-one correspondence between multilinear polynomials over $\mathbb{F}_p$ on $2p$ variables and the set of all functions from $\{0,1\}^{2p} \to \mathbb{F}_p$. Since h is the same as the function that is α on the all-zeros input and 0 elsewhere in $\{0,1\}^{2p}$, we can use this correspondence to conclude that

$$h(x_1, \ldots, x_{2p}) = \alpha \cdot \prod_{i=1}^{2p} (1 - x_i).$$

Therefore the degree of h is $2p$.

Hence the degree of $f \cdot g$ is at least $2p$, implying that the degree of f is at least p.

4. Upper and Lower Bounds on $\mathsf{B}(n)$

In this section, we describe some explicit balancing set families.

Lemma 12.

(1) If n is divisible by 4 and $n \neq 4$, then $\mathsf{B}(n) \leq n/2 - 1$.
(2) If n is divisible by 2 and $n \neq 2$, then $\mathsf{B}(n) \leq n/2$.

Proof. When 4 divides n, there is a family of $k = \frac{n}{2} - 1$ sets that are balancing: take any k sets, each of size 4, satisfying $S_i \cap S_j = \{1, 2\}$ for all $i \neq j$. This family has the property that for any subset $X \subset [n]$ of size $n/2$, there is an $i \in [k]$ such that $|X \cap S_i| = 2$.

When 2 divides n, there is a family of $k = n/2$ sets that are balancing: take any k sets, each of size 2, satisfying $S_i \cap S_j = \{1\}$ for all $i \neq j$. This family has the property that for any subset $X \subset [n]$ of size $n/2$, there is an $i \in [k]$ such that $|X \cap S_i| = 1$. □

As implied by Theorem 2, when $n = 2p$ for a prime p, there is no construction with $k = \frac{n}{2} - 1$ sets; the minimum possible k in this case is $\frac{n}{2}$. We now prove Theorem 2.

Proof of Theorem 2. Lemma 12 implies that $\mathsf{B}(n) \leq p = n/2$. We now proceed to show that $\mathsf{B}(n) \geq p = n/2$. Let $S_1, \ldots, S_k$ be the balancing set family. Without loss of generality each $|S_i|$ is even, and therefore $1 \leq |S_i|/2 \leq p - 1$ for all $i \in [k]$. We will now construct a polynomial that is non-zero on the all-zeros input and

vanishes on all $x \in \{0,1\}^{2p}$ with p ones. Define the polynomial

$$f(x_1, \ldots, x_{2p}) = \prod_{i=1}^{k} \left(|S_i|/2 - \sum_{j \in S_i} x_j \right),$$

over $\mathbb{F}_p$ that has degree k. Since $1 \leq |S_i|/2 \leq p-1$ for all $i \in [k]$, $f(0) \neq 0$. We will show that $f(x) = 0$, for $x \in \{0,1\}^{2p}$, when x exactly has p ones. This is because the input x to f with exactly p ones corresponds to a set $X \subset [2p]$ of size p. The fact that there is an $i \in [k]$ such that $|S_i \cap X| = |S_i|/2$, implies that $|S_i|/2 - \sum_{j \in S_i} x_j = 0$. By applying Lemma 1, we can conclude that $k \geq p$. □

Remark. In Definition 1, since $|S_i \cap X| = |S_i|/2$, it is no loss of generality to assume that each S_i is even sized. The definition can be relaxed by having $|S_i \cap X| = \lceil |S_i|/2 \rceil$. In this relaxed definition, the family $\{1\}, \{2, \ldots, 2p\}$ is balancing and the size of the family is 2. However, if we impose an extra condition that each $|S_i| \geq 2$, then we can prove that the size of any such family is at least p.

When n is even and all sets in the balancing set family are of size 2, we use a graph theoretic argument to prove that the size of the family is at least $n/2$. This shows that the construction in the proof of part 2 of Lemma 12 is tight.

Theorem 13. *Let n be a positive even integer. Let $S_1, \ldots, S_k \subset [n]$ be a balancing set family. If $|S_i| = 2$ for all $i \in [k]$, then, $k \geq n/2$.*

We need the following lemma about integers to prove the above theorem.

Lemma 14. *Let $a_1, \ldots, a_k$ be positive integers such that $\sum_{i=1}^{k} a_i = n$. If $k > n/2$, then for every non-negative integer $s \leq n$, there is a $S \subseteq [k]$ such that $\sum_{i \in S} a_i = s$.*

Proof. We prove the lemma by induction on n. The base case is when $n = 1$. In this case $k = 1$, and the possible values for s is 0 and 1. Hence the statement is true for $n = 1$. Induction hypothesis assumes that the statement is true for all positive integers that are at most $n-1$, and we prove it for n. Without loss of generality, we can assume that a_1 is the largest among $a_1, \ldots, a_k$. If $a_1 = 1$, then the statement is true because this implies that $k = n$, and for every $s \leq n$, $\sum_{i=1}^{s} a_i = s$. For the rest of the proof, we assume that $a_1 \geq 2$. We will first show that $a_1 \leq \lceil n/2 \rceil$. Assume for contradiction that $a_1 > \lceil n/2 \rceil$. Since each $a_i \geq 1$, we can conclude that $\sum_{i=2}^{k} a_i \geq k - 1 \geq \lfloor n/2 \rfloor$. Therefore, we get that $\sum_{i=1}^{k} a_i > \lfloor n/2 \rfloor + \lceil n/2 \rceil = n$, which is a contradiction.

We claim that $k-1 > \frac{n-a_1}{2}$. Indeed, since $k > n/2$, we have

$$k-1 > \frac{n-2}{2} \geq \frac{n-a_1}{2},$$

where the last inequality follows from the fact that $a_1 \geq 2$. As $k-1 > \frac{n-a_1}{2}$, we can apply the induction hypothesis on $a_2, \ldots, a_k$. By the induction hypothesis, for every $s \leq n-a_1$, there is a $S \subseteq \{2,3,\ldots,k\}$ such that $\sum_{i\in S} a_i = s$. For every $s > n-a_1$, let s' be such that $s = a_1 + s'$. We have

$$n - a_1 \geq s' = s - a_1 > n - 2a_1 \geq -1$$

where the last inequality follows from $a_1 \leq \lceil n/2 \rceil$. This implies that $0 \leq s' \leq n-a_1$, and the induction hypothesis gives a set $S \subseteq \{2,3,\ldots,k\}$ such that $\sum_{i\in S} a_i = s'$. Hence $\sum_{i\in S\cup\{1\}} a_i = s$. This completes the proof. □

Proof of Theorem 13. Consider a graph defined on the vertex set $[n]$. $(u,v) \in [n]\times[n]$, for $u \neq v$, is an edge in the graph iff $u, v \in S_i$ for some $i \in [k]$. The number of edges in this graph is k. Let k' be the number of connected components, and let $a_1, \ldots, a_{k'}$ be such that a_i is the size of the i'th connected component. Observe that $\sum_{i=1}^{k'} a_i = n$. Since the number of edges in the i'th connected component is at least $a_i - 1$, we get that $k' \geq n-k$. We will proceed to show that $k \geq n/2$. Assume for contradiction that $k < n/2$. We then know that $k' > n/2$. Lemma 14 implies there is a $S \subset [k']$ such that $\sum_{i\in S} a_i = n/2$. Define $X \subset [n]$ to be the set of all vertices in the connected components indexed by S. $|X| = n/2$ and there is no edge from X to $[n]\setminus X$. Therefore, for every $i \in [n]$, $|S_i \cap X| \neq 1$ and this is a contradiction. Hence, $k' \leq n/2$ and $k \geq n/2$. □

5. Balancing Families: Generalizations and Improvements

In this section we prove Theorems 3, 4 and 5. The following lemma is crucial in the proofs these theorems.

Lemma 15. *Let n be an even integer. Let $S_1, \ldots, S_k \subset [n]$ and $T_1, \ldots, T_k \subseteq [\mu(n/2)-1]$. Suppose that there is a set $R \subseteq [n]$ of size $n - 2\mu(n/2)$ such that for every $i \in [k]$ and $t \in T_i$, $|S_i \cap R| < t$, and for every $X \subset [n]$ of size $n/2$ there is an $i \in [k]$ such that $|X \cap S_i| \in T_i$. Then $\sum_{i=1}^{k} |T_i| \geq \mu(n/2)$.*

Proof. Define the polynomial

$$F(x_1, \ldots, x_n) = \prod_{i=1}^{k} \prod_{t \in T_i} \left(t - \sum_{j \in S_i} x_j \right).$$

Let $p = \mu(n/2)$. Define the polynomial $f(x_1, x_2, \ldots, x_{2p})$ over $\mathbb{F}_p$ by setting in F half of the variables indexed by R to 0 and the other half to 1. The degree of f is at most $\sum_{i=1}^{k} |T_i|$. We claim that f takes the value 0 on all inputs with exactly p ones and f is non-zero on the all-zeros input. This is sufficient to prove the theorem as Lemma 1 implies that $\sum_{i=1}^{k} |T_i| \geq p$.

The former part of the claim is true because the input x to f with exactly p ones along with the variables in R that are set to 1 correspond to a set $X \subset [n]$ of size $n/2$. The fact that there is an $i \in [k]$ and $t \in T_i$ with $|S_i \cap X| = t$, implies $t - \sum_{j \in S_i} x_j = 0$.

We now proceed to show that f is non-zero on the all-zeros input. On the all-zeros input for f, we know that all variables indexed by $[n] \setminus R$ are set to 0 and we do not have any control on the assignment to the variables in R. However, since for every $i \in [k]$ and $t \in T_i$, $0 < t < p$ and $|S_i \cap R| < t$, f is non-zero on the all-zeros input. □

Implications of Lemma 15. We now discuss the implications of Lemma 15 to the questions about balancing set families discussed in Section 1 and Section 2. The choice of R in Lemma 15 depends on the context. We obtain an asymptotically sharp lower bound for Galvin's problem and an improvement over the lower bound of Alon, Kumar and Volk.

$\mathsf{B}(n)$. We prove Theorem 3 using the following claim, which is proved in Appendix A.

Claim 16. *Let n be a positive integer and $S_1, \ldots, S_k \subset [n]$ be a balancing set family. If n is large enough and $k < n/2 - 2n^{0.98}$, then there exists a $R \subset [n]$ of size $n - 2\mu(n/2)$ such that for every $i \in [k]$, $|S_i \cap R| < |S_i|/2$.*

Proof of Theorem 3. Assume for contradiction that $\mathsf{B}(n) < n/2 - 2n^{0.98}$. Let R be the set given by Claim 16. By invoking Lemma 15 with R and each $T_i = \{|S_i|/2\}$, we get $\mathsf{B}(n) \geq \mu(n/2) \geq n/2 - O\left(n^{0.53}\right)$, where the last inequality follows from Theorem 11. This contradicts the assumption for large values of n. □

$\mathsf{J}(n,\tau)$. We prove Theorem 5. We have that each

$$T_i = \{|S_i|/2 - \tau + 1, \ldots, |S_i|/2, \ldots, |S_i|/2 + \tau - 1\}.$$

When $n = 2p$ for a prime p, $R = \emptyset$. Observing that each T_i is of size $2\tau - 1$, Lemma 15 implies Part 1 of Theorem 5.

We now proceed to prove Part 2 of Theorem 5. We need the following claim, and this claim is proved in Appendix A.

Claim 17. *Let n be a positive integer, τ be a positive integer, and $S_1, \ldots, S_k \subset [n]$ be τ-balancing set family. If n is large enough and $k < n/(7\tau) - n^{0.98}/(7\tau)$, then there exists a $R \subset [n]$ of size $n - 2\mu(n/2)$ such that for every $i \in [k]$, $|S_i \cap R| \leq |S_i|/2 - \tau$.*

Proof of Part 2 of Theorem 5. Assume for contradiction that

$$\mathsf{J}(n,\tau) < n/(7\tau) - n^{0.98}/(7\tau).$$

Let R be the set given by Claim 17. By invoking Lemma 15 with R and each

$$T_i = \{|S_i|/2 - \tau + 1, \ldots, |S_i|/2, \ldots, |S_i|/2 + \tau - 1\},$$

we get $\mathsf{J}(n,\tau) \geq \frac{\mu(n/2)}{2\tau-1} \geq \frac{n - O\left(n^{0.53}\right)}{4\tau-2}$, where the last inequality follows from Theorem 11. This contradicts the assumption for large values of n. □

$\mathsf{G}(n)$. We prove Theorem 4. For Galvin's problem, n is divisible by 4, each S_i is of size $n/2$ and each $T_i = \{n/4\}$. R can be chosen to be any arbitrary set of size $n - 2\mu(n/2)$. For Lemma 15 to apply, we need that for each $i \in [k]$ and $t \in T_i$, $T_i \subseteq [\mu(n/2) - 1]$ and $|S_i \cap R| < t$. This translates in to the condition that $\mu(n/2) > 3n/8$. Lemma 15 in conjunction with Theorem 11 implies Theorem 4.

Specifically Theorem 4 shows that our lower bound is sharp up to an additive $o(n)$ term as $\mathsf{G}(n) \leq n/2$. It is worth noting that $\mathsf{G}(n) < n/2$ for $n \in \{8, 16\}$, so a general $n/2$ lower bound is false (see [5]).

6. Computing Majority using Depth-2 Threshold Circuits

We first prove Theorem 7.

Proof of Theorem 7. Let k be the top fan-in of the circuit, and let $g_1, \ldots, g_k$ be the threshold functions given by the bottom gates of the circuit. We know that g_i

is defined by an inequality of the form $L_i(x) \geq t_i$ for a linear function L_i. Assume towards a contradiction that $k < p$ and each $t_i \neq p$.

Define the polynomial

$$f(x) = \prod_{i \in \{j \mid 0 < t_j < 2p\}} (L_i(x) - t_i)$$

over $\mathbb{F}_p$ that has degree at most k. By definition, $f(0)$ is non-zero. We claim that $f(x) = 0$ on every $x \in \{0,1\}^{2p}$ with p ones. Indeed, for such a x we have that $\mathsf{MAJ}(x) = 1$, but for x' that is obtained from x by flipping a coordinate with value 1 to 0, we have that $\mathsf{MAJ}(x') = 0$. Observe that each L_i is a linear function with coefficients in $\{0,1\}$. Since x and x' only differ in one coordinate, we have $L_i(x) - L_i(x') \in \{0,1\}$ for every $i \in [k]$. $\mathsf{MAJ}(x) = 1$ and $\mathsf{MAJ}(x') = 0$ implies that there is an $i \in [k]$ such that $g_i(x) = 1$ and $g_i(x') = 0$. This means that $L_i(x) = t_i$, but $L_i(x') = t_i - 1$. Moreover, this implies that $0 < t_i < 2p$. Hence, for every $x \in \{0,1\}^{2p}$ with p ones, there is an $i \in [k]$ such that $L_i(x) = t_i$ and $0 < t_i < 2p$, which makes $f(x) = 0$. Therefore Lemma 1 implies that the degree of f is at least p, which is a contradiction. □

We obtain the following theorem for arbitrary values of n, which is proved using Theorem 7.

Theorem 18. *In any depth-2 circuit computing the majority of n bits, if the bottom gates compute unweighted thresholds, either the top fan-in is at least $\mu(n/2)$, or some gate at the bottom computes a threshold T_t with $t \geq \mu(n/2)$.*

Proof. Let k be the top fan-in of the circuit, and let $p = \mu(n/2)$. If there exists a bottom gate with threshold at least p, then we are done. So assume that all bottom gates have threshold less than p. Set half the variables in $x_{2p+1}, \ldots, x_n$ to 0 and the other half to 1. We get a new depth-2 circuit computing the majority of $x_1, \ldots, x_{2p}$. Any bottom threshold gate computing T_t that reads constants is equivalent to a threshold gate computing $T_{t'}$ on the same input variables with $t' \leq t < p$, and $T_{t'}$ reads no constants. Here, $t' = t - \alpha$, where α is the number of ones read by T_t. Replacing each bottom gate that reads constants with its equivalent gate that reads no constants, we obtain a depth-2 circuit in which each bottom gate computes a threshold function with threshold less than p and does not read constants. By applying Theorem 7, we can conclude that $k \geq p$. □

Using Theorem 11 we get a corollary to Theorem 18.

Corollary 19. *In any depth-2 circuit computing the majority of n bits, if the bottom gates compute unweighted thresholds, then the fan-in is at least $n/2 - O(n^{0.53})$.*

7. Proof of Theorem 10

Let $g_1, \ldots, g_k$ be the threshold functions given by the bottom gates of the circuit. Let

$$L(x) = \sum_{i \leq p-1} x_i + 2 \sum_{i > p-1} x_i.$$

Note that L is a polynomial on $\frac{3p-1}{2}$ variables. For $i \in [k]$, we know that g_i is defined by an inequality of the form $L_i(x) \geq t_i$ for a linear function L_i with coefficients in $\{0,1\}$.

Consider the polynomial

$$f(x) = \prod_{i \in \{j | 0 < t_j < p\}} (L_i(x) - t_i)$$

over $\mathbb{F}_p$ that has degree at most k. By definition, f is non-zero on the all-zeros input. We will show that $f(x) = 0$ on $x \in \{0,1\}^{\frac{3p-1}{2}}$ such that $L(x) = p$.

Let $x \in \{0,1\}^{\frac{3p-1}{2}}$ be such that $L(x) = p$. Note that for every such x, the number of ones in it is at most $p-1$ and at least 1. For every $x' \in \{0,1\}^{\frac{3p-1}{2}}$ that is obtained by flipping one of the coordinates of x with value 1 to 0, we have $T(x') = 0$. For such x, x', there must be an $i \in [k]$ such that $g_i(x) = 1$ and $g_i(x') = 0$. Moreover, L_i being a linear function with coefficients in $\{0,1\}$ implies that $L_i(x) - L_i(x') \in \{0,1\}$. Since $g_i(x) \neq g_i(x')$, we have $L_i(x) = t_i$. In addition, since the number ones in x is at most $p-1$ and at least 1, we get that $0 < t_i < p$. Hence we can conclude that $f(x) = 0$.

We now find a polynomial g that is 0 everywhere in $\{0,1\}^{\frac{3p-1}{2}}$, except on the all-zeros input and x such that $L(x) = p$. Define

$$g(x) = (1 - x_1) \cdot \prod_{i=1}^{p-1} (i - L(x)).$$

The degree of g is p, and $f \cdot g$ is non-zero on the all-zeros input and 0 elsewhere in $\{0,1\}^{\frac{3p-1}{2}}$. We will show that the degree of $f \cdot g$ is at least $(3p-1)/2$. As in the proof of Lemma 1, let h be the multilinearization of $f \cdot g$. Then h is non-zero

on the all-zeros input and 0 elsewhere in $\{0,1\}^{\frac{3p-1}{2}}$. Therefore the degree of h is at least $(3p-1)/2$. Since the degree of h is at most the degree of $f \cdot g$, the degree of f is at least $(p-1)/2$.

8. Upper and Lower Bounds on $\mathsf{U}(n,t)$

Theorem 6 is proved in this section. We first recall the definition of a Kneser graph. The Kneser graph $K_{n,\alpha}$ is a graph whose vertices are identified with the subsets of $[n]$ of size α, and there is an edge between two vertices if and only if the corresponding subsets are disjoint. We need the following theorem bounding the chromatic number of Kneser graphs.

Theorem 20 ([18])**.** *Consider the Kneser graphs in which the vertex set is given by subsets of $[n]$ of size α. Then the chromatic number of this graph is* $\max\{1, n-2\alpha+2\}$.

Proof of Theorem 6. We first prove the upper bound. The following $2t+2$ sets form an unbalancing family:

$$\{1\},\{2\},\ldots,\{2t+1\},\{2t+2,2t+3,\ldots,n\}.$$

The above family has the property that for a given $X \subseteq [n]$ of size $n/2-t$, either $X \subseteq \{2t+2,2t+3,\ldots,n\}$ or not. In the former case,

$$|X \cap \{2t+2,2t+3,\ldots,n\}| = n/2-t > \frac{n-2t-1}{2}.$$

In the latter case, there will be an $i \in [2t+1]$ such that $i \in X$. Therefore, $|X \cap \{i\}| = 1 > \frac{1}{2}$.

We now prove the lower bound. Consider the Kneser graph in which the vertex set is given by subsets of $[n]$ of size $n/2-t$. We claim that the chromatic number of this graph is at most k. The coloring is as follows: For every $X \subseteq [n]$ of size $n/2-t$, we know that there is an $i \in [k]$ such that $|S_i \cap X| > |S_i|/2$. The vertex associated with X is given the color i. This is a proper coloring because for every $X, Y \subseteq [n]$, each of size $n/2-t$ that are disjoint, it cannot be the case that $|X \cap S_i| > |S_i|/2$ and $|Y \cap S_i| > |S_i|/2$. Therefore by Theorem 20, we can conclude that $k \geq 2t+2$. $\square$

References

[1] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in c logn parallel steps. *Combinatorica*, 3(1):1–19, Mar 1983.

[2] E. Allender and M. Koucký. Amplifying lower bounds by means of self-reducibility. *Journal of the ACM*, 57(3):1–36, Mar 2010.

[3] N. Alon. Personal communication. 2019.

[4] N. Alon, E. E. Bergmann, D. Coppersmith, and A. M. Odlyzko. Balancing sets of vectors. *IEEE Transactions on Information Theory*, 34(1):128–130, 1988.

[5] N. Alon, M. Kumar, and B. L. Volk. Unbalancing sets and an almost quadratic lower bound for syntactically multilinear arithmetic circuits. *arXiv:1708.02037*, 2017.

[6] K. Amano. Depth Two Majority Circuits for Majority and List Expanders. In I. Potapov, P. Spirakis, and J. Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, volume 117, pages 81:1–81:13, 2018.

[7] K. Amano and M. Yoshida. Depth two (n-2)-majority circuits for n-majority. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, E101.A(9):1543–1545, 2018.

[8] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes, ii. *Proceedings of the London Mathematical Society*, 83(3):532–562, 2001.

[9] C. Engels, M. Garg, K. Makino, and A. Rao. On expressing majority as a majority of majorities. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 24, page 174, 2017.

[10] H. Enomoto, P. Frankl, N. Ito, and K. Nomura. Codes with given distances. *Graphs and Combinatorics*, 3(1):25–38, 1987.

[11] D. Eppstein and D. S. Hirschberg. From discrepancy to majority. *Algorithmica*, 80(4):1278–1297, 2018.

[12] P. Frankl and V. Rödl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987.

[13] K. A. Hansen and V. V. Podolskii. Exact threshold circuits. In *2010 25th Annual IEEE Conference on Computational Complexity*, pages 270–279. IEEE, 2010.

[14] J. Håstad, G. Lagarde, and J. Swernofsky. d-galvin families. 01 2019.

[15] G. Hegedűs. Balancing sets of vectors. *Studia Scientiarum Mathematicarum Hungarica*, 47(3):333–349, 2009.

[16] M. J. Jansen. Lower bounds for syntactically multilinear algebraic branching programs. In *International Symposium on Mathematical Foundations of Computer Science*, pages 407–418, 2008.

[17] A. S. Kulikov and V. V. Podolskii. Computing majority by constant depth majority circuits with low fan-in gates. *arXiv:1610.02686*, 2016.

[18] L. Lovász. Kneser's conjecture, chromatic number, and homotopy. *Journal of Combinatorial Theory, Series A*, 25(3):319 – 324, 1978.

[19] G. Posobin. Computing majority with low-fan-in majority queries. *arXiv:1711.10176*, 2017.
[20] R. Raz, A. Shpilka, and A. Yehudayoff. A lower bound for the size of syntactically multilinear arithmetic circuits. *SIAM Journal on Computing*, 38(4):1624–1647, 2008.
[21] S. Srinivasan. Personal communication. 2018.
[22] L. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5(3):363 – 366, 1984.
[23] R. R. Williams. Limits on representing boolean functions by linear combinations of simple functions: thresholds, relus, and low-degree polynomials. *arXiv:1802.09121*, 2018.

Appendix A. Proofs of Claim 16 and Claim 17

We need the following claim to prove Claim 16 and Claim 17.

Claim 21. *Let n be a positive integer, $\emptyset \subset S \subset [n]$, and $0 \leq t \leq n$ be an integer. Let T be a random subset of $[n]$ of size r. Then*

$$\Pr[|S \cap T| \geq t] \leq 2^{|S|} \cdot \left(\frac{r}{n}\right)^t.$$

Proof. We have

$$\Pr[|S \cap T| \geq t] \leq \binom{|S|}{t} \cdot \frac{\binom{n-t}{r-t}}{\binom{n}{r}} \leq 2^{|S|} \cdot \left(\frac{r}{n}\right)^t.$$

□

Proof of Claim 16. Let $p = \mu(n/2)$ and r be such that $r = n - 2p$. By Theorem 11, for large enough n, $r < n^{0.54}$. In addition, we can assume without loss of generality that for all $i \in [k]$, $|S_i|$ is even.

Define

$$A = \{j | j \in S_i \text{ for some } i \in [k] \text{ and } |S_i| = 2\}$$

and

$$B = \{j | j \in S_i \text{ and } j \in S_{i'} \text{ for } i, i' \in [k], i \neq i', \text{ and } |S_i| = |S_{i'}| = 4\}.$$

In words, A is the set of elements which are in sets of size 2 and B is the set of elements that belong to at least 2 sets of size 4. We claim that $|A \cup B| \leq n - 4n^{0.98}$. Indeed, we have that

$$|A| \leq 2 \cdot (\# \text{ sets of size } 2)$$

and

$$|B| \leq (4 \cdot (\# \text{ sets of size } 4))/2 = 2 \cdot (\# \text{ sets of size } 4).$$

Since $k \leq n/2 - 2n^{0.98}$, we get that $|A \cup B| \leq n - 4n^{0.98}$.

By the definition of B, we have the property that for every $j \in [n] \setminus B$, it is the case that j appears in at most one set S_i of size 4. Let $X \subseteq [n] \setminus (A \cup B)$ be the largest set such that for every $i \in [k]$ if $|S_i| = 4$ and $|S_i \cap (A \cup B)| < 4$, then $|S_i \cap X| = 1$. The definition of X implies that $|X| \geq |[n] \setminus (A \cup B)|/4 \geq n^{0.98}$ for large enough n.

We will pick a random subset R of size r from X, and show that

$$\Pr[\exists i \in [k], |S_i \cap R| \geq |S_i|/2] < 1/n^{0.2}. \tag{1}$$

This is sufficient as this implies that there exists a set $R \subset [n]$ of size r such that for every $i \in [k]$, $|S_i \cap R| < |S_i|/2$.

We now proceed to prove Inequality 1. For every $i \in [k]$, by the definition of X, if $|S_i| = 2$, then $|S_i \cap R| = 0$. Similarly, if $|S_i| = 4$, then $|S_i \cap R| \leq 1$. We now consider the case when $|S_i| \geq 6$ for $i \in [k]$. Using $r < n^{0.54}$ and $|X| \geq n^{0.98}$, we get from Claim 21 that for every $i \in [k]$,

$$\begin{aligned}\Pr[|S_i \cap R| \geq |S_i|/2] &\leq 2^{|S_i|} \cdot \left(\frac{1}{n^{0.44}}\right)^{|S_i|/2} \\ &\leq 2^{|S_i| - 0.22 \cdot |S_i| \cdot \log n}.\end{aligned}$$

For large enough n we have,

$$\Pr[|S_i \cap R| \geq |S_i|/2] \leq 2^{-0.21 \cdot |S_i| \cdot \log n} \leq n^{-1.2},$$

where the last inequality follows from the fact that $|S_i| \geq 6$. Therefore by an union bound over $i \in [k]$, we can conclude that

$$\Pr[\exists i \in [k], |S_i \cap R| \geq |S_i|/2] < 1/n^{0.2}.$$

□

Proof of Claim 17. Note that $\tau \geq 1$. Let $p = \mu(n/2)$ and r be such that $r = n - 2p$. By Theorem 11, for large enough n, $r < n^{0.54}$.

Consider the following iterative process: If there is a set S_i with at most 7τ elements, remove the set and its elements from other sets to which they belong. Repeat this process until all remaining sets have size more than 7τ. Let A denote the set of elements from $[n]$ that did not belong a S_i that was removed,

and let $T_1, \ldots, T_{k'} \subseteq A$ ($k' \leq k$) be the sets that remain. Observe that since $k < n/(7\tau) - n^{0.98}/(7\tau)$, $|A| \geq n^{0.98}$.

If $k' = 0$, then let R be any subset of A of size r. We now consider the case when $k' \geq 1$. We will pick a random subset R of size r from A, and show that

$$\Pr[\exists i \in [k'], |T_i \cap R| \geq |T_i|/2 - \tau] < 1/n^{0.01}. \tag{2}$$

This is sufficient because it shows the existence of a R of size r such that $|S_i \cap R| < |S_i|/2 - \tau$ for every $i \in [k]$. Indeed for each S_i that was removed, we have that $|S_i \cap R| = 0$. For each S_i that was not removed, we have that $|S_i \cap R| < |T_i|/2 - \tau \leq |S_i|/2 - \tau$, where T_i is the set corresponding to S_i that remained.

We now proceed to show Inequality 2. Using $r < n^{0.54}$ and $|A| \geq n^{0.98}$, we get from Claim 21 that for every $i \in [k']$,

$$\begin{aligned}\Pr[|T_i \cap R| \geq |T_i|/2 - \tau] &\leq 2^{|T_i|} \cdot \left(\frac{1}{n^{0.44}}\right)^{|T_i|/2-\tau} \\ &= 2^{|T_i|(1-0.22\log n)+0.44\tau\log n} \\ &\leq 2^{(0.44\tau - 0.21|T_i|)\cdot\log n},\end{aligned}$$

where the last inequality is true for large enough n. Since, $|T_i| \geq 7\tau$, we have that $0.44\tau - 0.21|T_i| < -1.01\tau$. Therefore, $\Pr[|T_i \cap R| \geq |T_i|/2 - \tau] \leq n^{-1.01}$, as $\tau \geq 1$. By an union bound over $i \in [k']$, we can conclude that

$$\Pr[\exists i \in [k'], |T_i \cap R| \geq |T_i|/2 - \tau] < 1/n^{0.01}.$$

□

Institute of Mathematics of ASCR, Prague
Email address: pahrubes@gmail.com

Paul G. Allen School of Computer Science & Engineering, University of Washington
Email address: sivanr@cs.washington.edu

Paul G. Allen School of Computer Science & Engineering, University of Washington
Email address: anuprao@cs.washington.edu

Department of Mathematics, Techion-IIT
Email address: amir.yehudayoff@gmail.com

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il