Electronic Colloquium on Computational Complexity, Report No. 174 (2017)

On Expressing Majority as a Majority of Majorities

Christian Engels[*] Mohit Garg[†] Kazuhisa Makino[‡] Anup Rao[§]

November 13, 2017

Abstract

If $k < n$, can one express the majority of n bits as the majority of at most k majorities, each of at most k bits? We prove that such an expression is possible only if $k \gtrsim n^{4/5} = n^{0.8}$. This improves on a bound proved by Kulikov and Podolskii [KP17], who showed that $k \gtrsim n^{0.7+o(1)}$. Our proof is based on ideas originating in discrepancy theory, as well as a strong concentration bound for sums of independent Bernoulli random variables and a strong anticoncentration bound for the hypergeometric distribution.

1 Introduction

Define the majority function $\mathsf{Maj}_n : \{+1, -1\}^n \to \{+1, -1\}$ by

$$\mathsf{Maj}_n(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

Is it possible to express the majority function on n coordinates as a composition of majority functions of at most k coordinates, when $k < n$? This is a question with a long tradition in complexity theory, and many interesting answers.

In their famous paper, Ajtai, Komlós and Szemerédi [AKS83] constructed a *sorting network* of depth[1] $\lesssim \log n$. This is a network that sorts n numbers by comparing pairs of numbers in each step. Each such sorting operation can be simulated by the majority of 2 inputs and a constant, and the majority of n bits is simply the middle number in the sorted order of all the bits. So, their construction shows that the majority of n bits can be expressed as a tree of majorities of depth $\lesssim \log n$, each taking only 3 bits as input. This was followed by a simple non-explicit construction

[*] Department of Computer Science and Engineering, Indian Institute of Technology Bombay, Mumbai, christian@cse.iitb.ac.in. Part of this work was done while the author was at the Research Institute for Mathematical Sciences, Kyoto University.

[†] Department of Mathematics and Computer Science, The Open University of Israel, Ra'anana, mohitga@openu.ac.il. Supported by ISF grant 1357/16. Part of this work was done while the author was at the Research Institute for Mathematical Sciences, Kyoto University and while visiting the Tata Institute of Fundamental Research, Mumbai.

[‡] Research Institute for Mathematical Sciences, Kyoto University, Japan, makino@kurims.kyoto-u.ac.jp.

[§] Allen School of Computer Science and Engineering, University of Washington, anuprao@cs.washington.edu. Supported by the National Science Foundation under agreements CCF-1420268 and CCF-1524251. This work was done while the author was visiting the Research Institute for Mathematical Sciences, Kyoto University.

[1] Here and below, we write $k \gtrsim n$ to denote $k = \Omega(n)$, and we write $k \lesssim n$ to denote $k = O(n)$.

ISSN 1433-8092

of such a tree by Valiant [Val84]. He showed that a random tree of majority gates, each reading a constant number of inputs, computes the majority function with positive probability. This is essentially the best one can hope for if k is a constant, since $\gtrsim \log n$ depth is required just to ensure that the output of the circuit depends on all the input coordinates.

More recently, Allender and Koucký [AK10] showed that the majority of n bits can be computed by a constant depth circuit with gates that compute the majority of n^ϵ bits, for and constant $\epsilon < 1$. All of this work is intimately connected to understanding the class TC_0 of constant depth circuits using threshhold gates (see [Juk12] for details).

Here we study a beautiful variant of this question suggested by Kulikov and Podolskii [KP17]—for what k is it possible to express the majority of n coordinates as a circuit of depth 2, where every gate computes the majority of at most k distinct coordinates? Kulikov and Podolskii found several non-trivial circuits of depth 2 that compute majority for small values of n. For example, they showed that one can compute the majority of 7 bits using a depth 2 circuit consisting of gates that compute the majority and read at most 5 bits. In this upper bound, they allowed some of the gates to read coordinates multiple times. In addition, they show that there is a depth 3 circuit computing majority with gates of fan-in $\lesssim n^{2/3}$. In very recent work subsequent to the present paper, Hrubeš and Rao [HR] showed that whenever n is divisible by 4, there is a depth 2 circuit with $k = n - 1$ computing the majority of n inputs. In this construction, they allowed the top gate to read a constant in addition to the gates in the middle layer.

Kulikov and Podolskii gave a simple argument to prove a lower bound of $k \gtrsim n^{2/3} = n^{0.66\ldots}$ for depth 2 circuits. We sketch the idea here. For the sake of finding a contradiction, suppose $k \ll n^{2/3}$ and we are given a depth 2 circuit computing the majority of n coordinates with gates that compute the majority of at most k coordinates. Since there are at most k gates in the middle layer, and each reads at most k variables, there must be an input variable x_u that is read by at most $\lesssim k^2/n$ gates. Moreover, one can ensure that x_u is only read by gates that read at least 3 inputs. Set $x_u = 1$ and the values of all other inputs read by gates that see x_u to -1. Since $k \ll n^{2/3}$, we will have set at most $k \cdot k^2/n \ll n$ input variables in this way. So, we can set all the remaining unset inputs in such a way that the overall input x is balanced and has majority 1. Then flipping the value of x_u does not change the value of the output of the circuit because none of the gates reading x_u change their values. Yet the majority of x does change after x_u is flipped, showing that the circuit does not compute majority. Using a more complicated argument, Kulikov and Podolskii proved a better lower bound of $k \gtrsim n^{0.7+o(1)}$. Our main result improves on these lower bounds.

A $\mathsf{Maj}_{\leq k} \circ \mathsf{Maj}_{\leq k}$ circuit is a depth two circuit consisting of gates that can only compute the majority function of at most k inputs. A majority gate with fan-in d takes in d *distinct* inputs and computes Maj_d, the majority of its d inputs. Here we do not allow the circuit to contain constants. We note that Kulikov and Podolskii do also prove lower bounds when the gates are not required to read distinct variables, and may read constants. In this case they prove that $k \gtrsim n^{13/19} \sim n^{0.68\ldots}$.

Our main theorem is:

Theorem 1. *If a* $\mathsf{Maj}_{\leq k} \circ \mathsf{Maj}_{\leq k}$ *circuit with* n *inputs computes* Maj_n*, then* $k \gtrsim n^{4/5}$.

It remains open whether our bound is tight or not. Next, we give an informal high-level overview of our proof. We prove Theorem 1 in Section 3.

1.1 Techniques

For ease of exposition, we ignore several key details here that are addressed in the actual proof in later sections. For the sake of finding a contradiction, suppose we are given a $\mathsf{Maj}_{\leq k} \circ \mathsf{Maj}_{\leq k}$ circuit computing Maj_n. It will be convenient for the purpose of the exposition here to assume that each gate in the circuit reads *exactly* k inputs, with $k = \epsilon \cdot n^{4/5}$ for some small constant $\epsilon > 0$. Our proof does not really require this assumption.

Kulikov and Podolskii achieve their lower bound by identifying a balanced input x with some nice properties. An input is *balanced* if it has $\lceil n/2 \rceil$ 1's, and $\mathsf{Maj}_n(x) = 1$. They define a random process depending on the circuit to modify the input x, obtaining a different input x' with $\mathsf{Maj}_n(x') = -1$. They guarantee that the values of all gates in the middle layer of the circuit remain the same whether the circuit reads x or x'. This contradicts the correctness of the circuit.

In our proof, we modify their approach, allowing the values of the gates in the middle layer to change. As in their proof, we find a balanced input x with $\mathsf{Maj}_n(x) = 1$. We then identify three distinct input variables x_u, x_v and x_w with the property that x_u is much more influential than x_v and x_w. More precisely, we guarantee that $x_u = -1$, $x_v = 1 = x_w$, and flipping the value of x_u changes the value of many gates in the circuit, but subsequently flipping x_v, x_w changes the value of very few gates in the circuit. This is stated formally in Lemma 6. Lemma 6 can be used to prove that the circuit cannot compute majority, because after we flip all three variables, the majority of the whole input should be -1, yet the number of gates in the middle layer with a positive value can only increase.

We find the input x and u, v, w in two steps. First, we identify an input x and a variable u for which $x_u = -1$ and x_u is very influential. To do this, we fix u to be the coordinate read most often in the circuit. Then we use a technique developed by Beck and Fiala [BF81] in the context of discrepancy theory. In a well studied problem in discrepancy theory, one is given a family $\mathcal{S}$ of subsets of $\{1, 2, \ldots, n\}$, and looks for an assignment $x \in \{+1, -1\}^n$ such that *every* subset $S \in \mathcal{S}$ of the coordinates is *close* to being balanced. In our work, we use similar ideas, combined with a concentration bound, to find an assignment x such that *many* of the gates that read x_u in the middle layer see inputs that are *perfectly* balanced. In the rest of this discussion, we say that a gate is balanced if it sees inputs that are balanced. One should expect most variables to be read by k^2/n gates. In a random assignment, one would expect about a $1/\sqrt{k}$ fraction, or about $k^{3/2}/n$ of these gates, to be balanced. Our proof finds $k/\sqrt{n}$ perfectly balanced gates that read x_u, which is larger since $k \ll n$. This step of our proof gives essentially optimal parameters. It is captured by Lemma 7.

In the second step, we need to identify variables x_v, x_w that are read by very few gates that are close to being balanced. Since we expect only $(k^2/n) \cdot (1/\sqrt{k}) = k^{3/2}/n \ll k/\sqrt{n}$ balanced gates to read a typical coordinate in a uniformly random assignment, we might hope to actually prove that $k \gtrsim n$ in this way. However, the assignment x found in the first step is not a uniformly random assignment—it was chosen specifically to make the gates reading x_u balanced. So, under this assignment there may be no variables that are read by few balanced gates. In order to complete the proof, we select a subset of roughly $n/(8k)$ of the gates reading x_u that are perfectly balanced. Since $n/(8k) \ll k/\sqrt{n}$ we can find such a set of gates that are balanced. These gates can read only $n/8$ of the inputs, so we fix the inputs to these gates, and resample the rest of the input variables subject to the constraint that the resulting string x is perfectly balanced. Only a small fraction of the inputs have been fixed in this process, so we are able to use an anticoncentration bound, double counting and the Cauchy-Schwartz inequality to show that most of the reset variables are read by $\lesssim k^{3/2}/n$ perfectly balanced gates, as one would expect in a random assignment. This fact is captured in Claim 8. Thus, our argument gives a contradiction, since $n/(8k) \gtrsim k^{3/2}/n$.

2 Preliminaries

For a positive integer r, let $[r]$ denote the set $\{1, \ldots, r\}$. We write $k \gtrsim n$ to denote $k = \Omega(n)$, and we write $k \lesssim n$ to denote $k = O(n)$.

We use the following strong form of Stirlings approximation of the factorial function:

Proposition 2 (Stirling's Approximation)**.** *For every positive integer n,*

$$\sqrt{2\pi n} \cdot n^n \cdot e^{-n} \leq n! \leq e \cdot \sqrt{n} \cdot n^n \cdot e^{-n}.$$

The following theorem proved by Hoeffding [Hoe56] will be useful. We say that a collection of random variables $Y_1, \ldots, Y_n \in \{0, 1\}$ are *identically distributed* if $\Pr[Y_1 = 1] = \Pr[Y_2 = 1] = \ldots = \Pr[Y_n = 1]$.

Theorem 3. *For every $p \in [0, 1]$ and positive integer n, let $Y_1, Y_2, \ldots, Y_n \in \{0, 1\}$ be independently distributed bits with $\mathbb{E}\left[\sum_{i=1}^n Y_i\right] = pn$. Suppose that a and b are integers with $a \leq pn \leq b$. Then among all such distributions, $\Pr\left[a \leq \sum_{i=1}^n Y_i \leq b\right]$ is minimized when $Y_1, \ldots, Y_n$ are identically distributed.*

Hoeffding's theorem and Stirling's approximation can be used to prove the following concentration estimate:

Theorem 4. *There exists a constant $c_0 > 0$ such that for any positive integer n, if we have independently distributed random variables $X_1, X_2, \ldots, X_n \in \{+1, -1\}^n$ such that the expectation $\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \mu$ is an integer, then these variables must safitisfy*

$$\Pr\left[\left|\mu - \sum_{i=1}^n X_i\right| \leq 1\right] \geq \frac{c_0}{\sqrt{n}}.$$

We defer the proof of Theorem 4 to Section 5. We use the following proposition, proved by Raz and Yehudayoff [RY09] (Lemma 5.3[2]). The proposition gives an anticonentration bound for the hypergeometric distribution.

Proposition 5. *There exists a constant $c_1 > 0$ such that for every positive integers n, h and integer s, ℓ, if $|s| \leq n/2, h \leq n/4$ and $X \in \{+1, -1\}^n$ is sampled uniformly from the set of strings with $\sum_{i=1}^n X_i = s$, then*

$$\Pr\left[\sum_{i=1}^h X_i = \ell\right] \leq \frac{c_1}{\sqrt{h}}.$$

[2] In [RY09], the authors assume in addition that $-n/2 \leq s \leq 0$. However, by symmetry the same statement holds when $0 \leq s \leq n/2$.

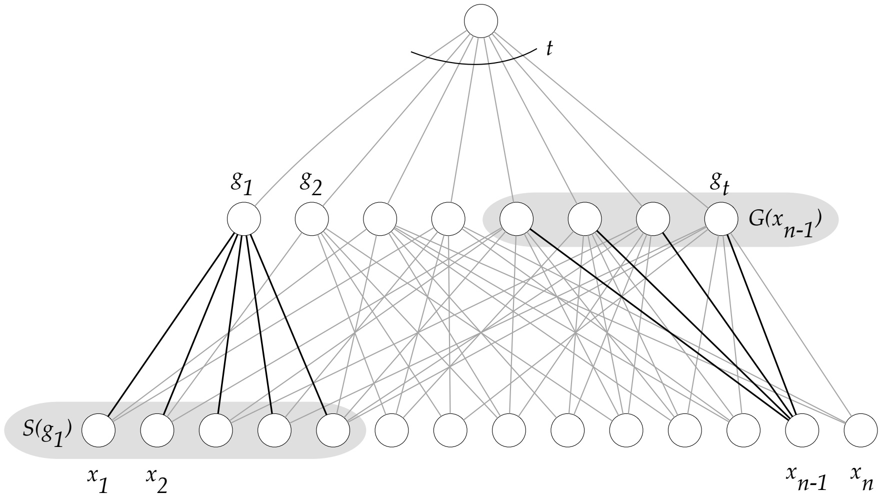


Figure 1: A circuit of depth 2.

3 Proving the main theorem

In order to prove Theorem 1, assume we are given a $\mathsf{Maj}_{\leq k} \circ \mathsf{Maj}_{\leq k}$ circuit, where $t\,(\leq k)$ is the fan-in of the top gate. The circuit takes n inputs and computes Maj_n. Let $g_1, \ldots, g_t$ denote the t gates that feed into the top gate. For each gate g that feeds into the top gate, let $S(g) \subseteq [n]$ denote the set of coordinates read by g. By definition,

$$g(x) = \begin{cases} 1 & \text{if } \sum_{j \in S(g)} x_j \geq 0, \\ -1 & \text{otherwise.} \end{cases}$$

Similarly, for each variable x_i, let $G(x_i)$ denote the set of gates that read x_i.

We say that the input to a gate g is *balanced* if $\sum_{j \in S(g)} x_j \in \{0,1\}$. Similarly, we say that the input to the circuit is balanced if $\sum_{i=1}^{n} x_i \in \{0,1\}$. A balanced x has majority $\mathsf{Maj}(x) = 1$, but flipping any x_i from 1 to -1 makes the majority -1.

It will be important to keep track of the *critical* gates, namely the gates that see inputs that are close to being balanced. For $x \in \{+1,-1\}^n$, $u \in [n]$, and integer r, define

$$\mathsf{Critical}(x,u,r) = \left\{ g \in G(x_u) : \left(\sum_{j \in S(g)} x_j \right) - 2r \in \{0,1\} \right\}.$$

These are the gates that read x_u and see inputs that are r steps away from being balanced. Set

$$\mathsf{C}(x,u,r) = |\mathsf{Critical}(x,u,r)|.$$

Lemma 6. *There exist constants $c, n_0 \geq 0$ such that for any positive integer $n \geq n_0$, any circuit as above with $k \leq c \cdot n^{4/5}$ has a balanced input $x \in \{+1,-1\}^n$ and three distinct coordinates $u, v, w \in [n]$ such that $x_u = -1$, $x_v = x_w = 1$ and*

$$\mathsf{C}(x,u,-1) \geq \sum_{\substack{r \in \{-1,0,1\} \\ z \in \{v,w\}}} \mathsf{C}(x,z,r).$$

Before proving Lemma 6, we first show how to use it to prove Theorem 1.

Proof of Theorem 1. We claim that x, u, v, and w as in Lemma 6 are witnesses to the fact that the circuit cannot compute majority correctly. Indeed, if the circuit correctly computes majority, since $\mathsf{Maj}_n(x) = 1$, we must have:

$$\sum_{i=1}^{t} g_i(x) \geq 0. \tag{1}$$

Now, consider $x' \in \{+1, -1\}^n$ obtained from x by flipping the value of x_u. Namely, for all $\ell \in [n]$, set

$$x'_\ell = \begin{cases} -x_\ell & \text{if } \ell = u, \\ x_\ell & \text{otherwise.} \end{cases}$$

We have $\sum_{i=1}^{n} x'_i \in \{2, 3\}$, and hence $\mathsf{Maj}_n(x') = 1$. Since $x'_\ell \geq x_\ell$ for every ℓ, we have $g_i(x') \geq g_i(x)$ for every i. We claim that exactly $\mathsf{C}(x, u, -1)$ gates change their values when moving from x to x'. If a gate g satisfies $\sum_{j \in S(g)} x_j \in \{-2, -1\}$ and $u \in S(g)$, then we must have $\sum_{j \in S(g)} x'_j \in \{0, 1\}$, so $g(x') > g(x)$. Moreover, only such gates can change their value. This together with (1) implies

$$\sum_{i=1}^{t} g_i(x') = \sum_{i=1}^{t} g_i(x) + 2 \cdot \mathsf{C}(x, u, -1) \geq 2 \cdot \mathsf{C}(x, u, -1). \tag{2}$$

Next, define $x'' \in \{+1, -1\}^n$ obtained from x' by flipping the values of x'_v, x'_w. Namely, for all $\ell \in [n]$, set

$$x''_\ell = \begin{cases} -x'_\ell & \text{if } \ell \in \{v, w\}, \\ x'_\ell & \text{otherwise.} \end{cases}$$

We have $\sum_{i=1}^{n} x''_i \in \{-2, -1\}$, implying that $\mathsf{Maj}_n(x'') = -1$. However, we shall prove that $\sum_{i=1}^{t} g_i(x'') \geq 0$, contradicting the correctness of the circuit. We shall estimate the number of gates g for which $g(x'') < g(x')$. Only the values of x''_v and x''_w were changed from their values in x', so only the gates that read v and w that are nearly balanced in x' can change their values. Formally, $g(x'') < g(x')$ implies

$$g \in \bigcup_{\substack{r \in \{0,1\} \\ z \in \{v,w\}}} \mathsf{Critical}(x', z, r).$$

Moreover, since x' and x differ in only one coordinate, and that coordinate was changed from -1 to 1, we have that for all $z \in \{v, w\}, r \in \{0, 1\}$:

$$\mathsf{Critical}(x', z, r) \subseteq \mathsf{Critical}(x, z, r) \cup \mathsf{Critical}(x, z, r - 1).$$

Hence $g(x'') < g(x')$ implies

$$g \in \bigcup_{\substack{r \in \{-1,0,1\} \\ z \in \{v,w\}}} \mathsf{Critical}(x, z, r).$$

Using (2), we get

$$\sum_{i=1}^{t} g_i(x'') \geq \sum_{i=1}^{t} g_i(x') - 2 \cdot \left| \bigcup_{\substack{r\in\{-1,0,1\}\\ z\in\{v,w\}}} \mathsf{Critical}(x,z,r) \right|$$
$$\geq 2 \cdot \mathsf{C}(x,u,-1) - 2 \cdot \sum_{\substack{r\in\{-1,0,1\}\\ z\in\{v,w\}}} \mathsf{C}(x,z,r) \geq 0,$$

where the last inequality follows from Lemma 6. This proves that the circuit cannot correctly compute Maj_n. □

It only remains to prove Lemma 6.

Proof of Lemma 6. Define

$$m = \sum_{i=1}^{n} |G(x_i)| = \sum_{i=1}^{t} |S(g_i)|.$$

This parameter m counts the total number of wires in the bottom layer of the circuit. The proof of the lemma proceeds in two steps.

Step 1 We find an assignment y and a coordinate u for which $\mathsf{C}(y,u,-1)$ is large.

Step 2 We modify this assignment y by randomly resetting a carefully chosen subset of its coordinates to obtain our final assignment x. We sample v and w uniformly from the set of coordinates that are reset in the above process. We ensure that $\mathsf{C}(x,u,-1)$ remains large, $x_v = x_w = 1$, and

$$\sum_{\substack{r\in\{-1,0,1\}\\ z\in\{v,w\}}} \mathsf{C}(x,z,r) \leq \mathsf{C}(x,u,-1)$$

with positive probability.

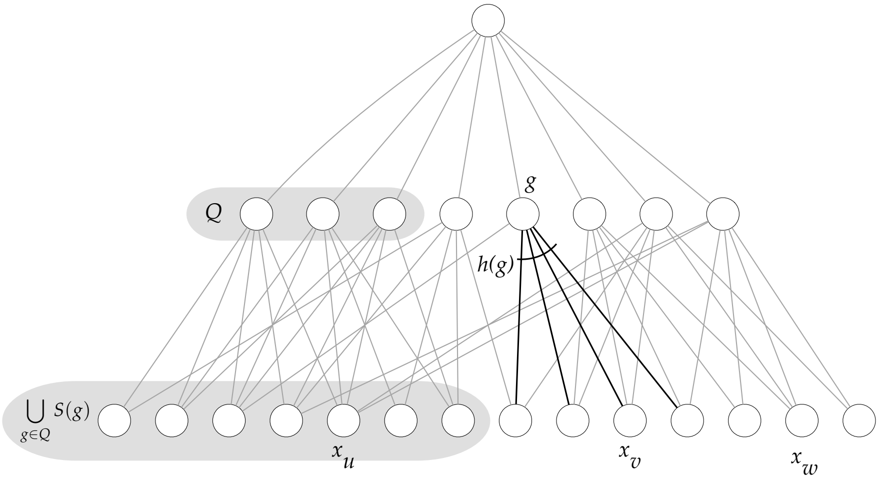
For Step 1, we make use of ideas from discrepancy theory [BF81], as well as a strong concentration bound. We prove:

Lemma 7. *There exist a constant $c_2 > 0$, such that for any circuit, there is a variable u, and vector $y \in \{+1,-1\}^n$ such that $y_u = -1$ and*

$$\mathsf{C}(y,u,-1) \geq c_2 \cdot \sqrt{\frac{m}{n}}.$$

Note that Lemma 7 does not give any upper bound on $\sum_{\substack{r\in\{-1,0,1\}\\ z\in\{v,w\}}} \mathsf{C}(y,z,r)$, for any choice of v and w. We defer the proof of Lemma 7 to Section 4.

In Step 2, we complete the proof of Lemma 6 by showing how to modify the assignment y promised by Lemma 7 to obtain the assignment x. Lemma 7 has already guaranteed that y contains a variable u for which $\mathsf{C}(y,u,-1)$ is large. We show how to modify y while preserving this



property. Let u and y be as in Lemma 7, and let $Q \subseteq \mathsf{Critical}(y, u, -1)$ be an arbitrary set of gates of size

$$|Q| = \min\left\{ \left\lceil c_2 \cdot \sqrt{\frac{m}{n}} \right\rceil, \lfloor n/8k \rfloor \right\}.$$

We shall modify y while preserving the inputs seen by the gates of Q. Consider the following random experiment. Pick a uniformly random balanced assignment from the set of assignments satisfying $x_i = y_i$, whenever $i \in \bigcup_{g\in Q} S(g)$. Namely, we fix the set of variables read by the gates in Q to their values under y, and pick the rest of the variables uniformly at random, subject to the constraint that the resulting assignment is a balanced input to the circuit. Since $|Q| \leq n/8k$, at most $n/8$ variables are read by the gates in Q. The resulting assignment x sets the remaining variables uniformly at random with a fixed sum. Next, pick two independent coordinates v and w uniformly at random from the set of variables not read by the gates of Q.

The resulting assignment x is always balanced. Moreover, since the input to the gates of Q has not been changed, we have $x_u = -1$ and

$$\mathsf{C}(x, u, -1) \geq |Q| = \min\left\{ \left\lceil c_2 \cdot \sqrt{\frac{m}{n}} \right\rceil, \lfloor n/8k \rfloor \right\}. \tag{3}$$

We claim that the expected numbers of critical gates for v and w are small:

Claim 8. *There exists a constant c_3 such that for any x, v and w sampled as described above, any positive integer n, and for any $r \in [n]$, we have*

$$\mathbb{E}\left[\mathsf{C}(x, v, r)\right] = \mathbb{E}\left[\mathsf{C}(x, w, r)\right] \leq c_3 \cdot \frac{\sqrt{tm}}{n}.$$

To prove the claim, first observe that since v and w are identically distributed, both expectations are the same. It suffices to prove the bound for v. Fix $r \in [n]$ arbitrarily. Let s be the sum of values of the variables outside those read by the gates in Q. For each gate g, define $h(g) = |S(g) \setminus (\bigcup_{g'\in Q} S(g'))|$ to be the number of variables that are not read by the gates in Q.

We claim that the probability that gate g is r away from balanced is at most $O(1/\sqrt{h(g)})$, when $h(g) > 0$. Indeed, the induced distribution on assignments x corresponds to a distribution on the coordinates not read by the gates of Q whose sum is s. Since $k \cdot |Q| \leq \frac{n}{8}$, we have

$$|s| \leq k \cdot |Q| + 1 \leq \frac{7n}{16} \leq \frac{n - k \cdot |Q|}{2}.$$

Moreover, we have

$$h(g) \leq k \leq \frac{n - k \cdot |Q|}{4},$$

for large enough n. Therefore, we can apply Proposition 5 to conclude that the probability that the gate $g \notin Q$ is r away from balanced can be bounded by $c_1/\sqrt{h(g)}$. Since the probablity that the gate belongs to $G(v)$ is at most $h(g)/(n - k \cdot |Q|)$, we have

$$\mathbb{E}\left[\mathsf{C}(x, v, r)\right] \leq \sum_{g : h(g) > 0} \frac{h(g)}{n - k \cdot |Q|} \cdot \frac{c_1}{\sqrt{h(g)}} \leq \frac{c_3}{n} \cdot \sum_{g : h(g) > 0} \sqrt{h(g)},$$

for a constant c_3. By the Cauchy-Schwartz inequality, we can bound this by

$$\leq \frac{c_3}{n} \cdot \sqrt{t} \cdot \sqrt{\sum_{g : h(g) > 0} h(g)} \leq c_3 \cdot \frac{\sqrt{tm}}{n},$$

proving the claim.

Claim 8 and linearity of expectation imply that

$$\mathbb{E}\left[\sum_{\substack{r \in \{-1,0,1\} \\ z \in \{v,w\}}} \mathsf{C}(x, z, r)\right] \leq 6c_3 \cdot \frac{\sqrt{tm}}{n}.$$

By Markov's inequality, when $c_4 = 60c_3$,

$$\Pr\left[\sum_{\substack{r \in \{-1,0,1\} \\ z \in \{v,w\}}} \mathsf{C}(x, z, r) > c_4 \cdot \frac{\sqrt{tm}}{n}\right] \leq 1/10. \tag{4}$$

On the other hand, since x is a balanced assignment and $k \cdot |Q| \leq n/8$,

$$\Pr[x_v = x_w = 1] \geq \left(\frac{(n/2) - k \cdot |Q|}{n}\right)^2 \geq \left(\frac{n/2 - n/8}{n}\right)^2 \geq 1/9.$$

Since u and v are uniform over a set of size at least $n - k \cdot |Q| \geq 7n/8$, we have

$$\Pr[u = v] \leq \frac{1}{n - k \cdot |Q|} \leq \frac{8}{7n}.$$

Thus, by the union bound,

$$\Pr[v \neq w, x_v = x_w = 1] \geq \frac{1}{9} - \frac{8}{7n} > 1/10,$$

for n large enough. Combining this bound with (4), there exist x, v and w such that

$$\sum_{\substack{r\in\{-1,0,1\}\\ z\in\{v,w\}}} \mathsf{C}(x,z,r) \leq c_4 \cdot \frac{\sqrt{tm}}{n},$$

x is balanced, $v \neq w$, and $x_v = x_w = 1$, as required.

Given (3), to complete the proof of the lemma we show that $|Q| \geq c_4 \cdot \frac{\sqrt{tm}}{n}$. Recall that $|Q| = \min\left\{\left\lceil c_2 \cdot \sqrt{m/n}\right\rceil, \lfloor n/8k \rfloor\right\}$. When $k \leq c \cdot n^{4/5}$, we have that:

$$c_2 \cdot \sqrt{\frac{m}{n}} = c_2 \cdot \sqrt{\frac{n}{t}} \cdot \frac{\sqrt{tm}}{n} > c_4 \cdot \frac{\sqrt{tm}}{n},$$

for n large enough. Also

$$\left\lfloor \frac{n}{8k} \right\rfloor \geq \left\lfloor \frac{n^{1/5}}{8c} \right\rfloor > c_4 \cdot \frac{\sqrt{k^3}}{n} \geq c_4 \cdot \frac{\sqrt{tm}}{n},$$

for c chosen small enough. Thus, $|Q| \geq c_4 \cdot \frac{\sqrt{tm}}{n}$, as required. □

4 Using discrepancy theory to prove Lemma 7

By averaging, there is an input variable x_u such that $|G(x_u)| \geq m/n$. Let x_u be such a variable. For $g \in G(x_u)$, let $R(g) = S(g) \setminus \{u\}$. Consider the system of $\ell\,(= |G(x_u)|)$ linear equations in the $n-1$ variables $x_1, \ldots, x_{u-1}, x_{u+1}, \ldots, x_n$:

$$\sum_{j\in R(g)} x_j = 0 \qquad \text{for } g \in G(x_u).$$

This system can be written as $B(x) = 0$, where we define $B(x)$ to be a vector with $B(x)_g = \sum_{j\in R(g)} x_j$. This linear system has a real-valued solution, since we can set $x = 0$. There are ℓ constraints and $n-1$ variables. If $\ell < n-1$, there must be a non-zero solution as well, since B cannot be an injective map. By scaling this solution, we can find a non-trivial solution x' such that $max_{i\in[n]\setminus\{u\}}|x'_i| = 1$. If there is a coordinate $r \in [n]$ with $|x'_r| = 1$, we add an equation to the system fixing the value of $x_r = x'_r$ to its value, to obtain a new system of constraints. Formally, we define a new linear map with $B'(x)_g = \sum_{j\in R(g)} x_j$, and for each variable r as above, $B'(x)_r = x'_r$. We let b be a column vector with $b_g = 0$, and $b_r = x'_r$. We obtain a new system of equations $B'(x) = b$, with $\ell' > \ell$ constraints. Again, if $\ell' < n-1$, there must be at least two solutions. Indeed, $B'(x) = 0$ must have a solution $x'' \neq 0$, and so for every $\epsilon > 0$, we have $B'(x' + \epsilon \cdot x'') = B'(x') = b$. In x'', for every coordinate r that has been fixed to $+1$ or -1 above, we have $x''_r = 0$. Increasing the value of ϵ from 0 to ∞, we must discover another solution of the form $x' + \epsilon \cdot x''$ that sets some new variable to $+1$ or -1, and keeps all the remaining variables with magnitude at most 1.

This process can always be continued until the number of variables of magnitude less than 1 is at most ℓ. Let $x \in [-1,1]^n$ denote the resulting real-valued solution when the above process cannot be continued. Let $T \subseteq [n]$ denote the set of at most ℓ coordinates that were not set to $+1$ or -1 in the above process. Define independent random variables $X_1, \ldots, X_{u-1}, X_{u+1}, \ldots, X_n \in \{+1, -1\}$

in such a way that $\mathbb{E}[X_i] = x_i$, for all i. By construction, x satisfies all of our original linear constraints, and for every $g \in G(x_u)$, we have:

$$\mathbb{E}\left[\sum_{j \in T \cap R(g)} X_j\right] = -\sum_{j \in R(g) \setminus T} x_j,$$

Define $\mu_i = \mathbb{E}\left[\sum_{j \in T \cap R(g)} X_j\right]$. By Theorem 4, we have that

$$\Pr\left[\left|\mu_i - \sum_{j \in T \cap R(g)} X_j\right| \leq 1\right] \geq \frac{c_0}{\sqrt{|T|}} \geq \frac{c_0}{\sqrt{\ell}}.$$

Thus, the expected number of equations satisfying this bound is at least

$$\frac{c_0}{\sqrt{\ell}} \cdot \ell \geq c_0 \cdot \sqrt{\frac{m}{n}}.$$

Therefore, there must be an assignment $z \in \{+1, -1\}^n$ such that for at least $c_0\sqrt{m/n}$ gates g, we have $\sum_{j \in R(g)} z_j \in \{-1, 0, 1\}$. Either at least half of these equations satisfy $\sum_{j \in R(g)} z_j \in \{0, 1\}$, or at least half of them satisfy $\sum_{j \in R(g)} z_j \in \{-1, 0\}$. In the first case, we set $y_u = -1$, and $y_i = -z_i$ for $i \neq u$ to obtain the desired assigment. This ensures that whenever $\sum_{j \in R(g)} z_j \in \{0, 1\}$, we have $\sum_{j \in S(g)} y_j \in \{-2, -1\}$. In the second case, we set $y_u = -1$ and $y_i = z_i$ for $i \neq u$ to obtain the desired assignment, ensuring that whenever $\sum_{j \in R(g)} z_j \in \{-1, 0\}$, we have $\sum_{j \in S(g)} y_j \in \{-2, -1\}$. In either case, we have found an assignment y satisfying the requirements of the lemma.

5 Proof of Theorem 4

Define $Y_1, \ldots, Y_n$ by setting $Y_i = \frac{X_i + 1}{2}$. Then $Y_1, \ldots, Y_n$ are independent bits, and $\mathbb{E}[\sum_{i=1}^n Y_i] = \frac{\mu + n}{2} = pn$, for $p = \frac{\mu + n}{2n}$. We have

$$\Pr\left[\left|\mu - \sum_{i=1}^n X_i\right| \leq 1\right] = \Pr\left[\left|pn - \sum_{i=1}^n Y_i\right| \leq \frac{1}{2}\right].$$

When $\mu + n$ is even, we set $a = b = pn$. Then by Theorem 3, this probability is minimized when all the bits are identically distributed. We can use Stirling's approximation, given in Proposition 2, to give a lower bound for this probability:

$$\Pr\left[\sum_{i=1}^n Y_i = pn\right] = \binom{n}{pn} \cdot p^{pn} \cdot (1-p)^{(1-p)n}$$
$$= \frac{n! \cdot p^{pn} \cdot (1-p)^{(1-p)n}}{(pn)! \cdot ((1-p)n)!},$$

which can be bounded using Stirling's approximation by:

$$\geq \frac{\sqrt{2\pi n} \cdot n^n \cdot e^{-n} \cdot p^{pn} \cdot (1-p)^{(1-p)n}}{e \cdot \sqrt{pn} \cdot (pn)^{pn} \cdot e^{-pn} \cdot e \cdot \sqrt{(1-p)n} \cdot ((1-p)n)^{(1-p)n} \cdot e^{-(1-p)n}}$$
$$= \frac{\sqrt{2\pi}}{e^2\sqrt{np(1-p)}} \cdot \frac{n^n \cdot p^{pn} \cdot (1-p)^{(1-p)n}}{p^{pn} \cdot n^{pn} \cdot (1-p)^{(1-p)n} \cdot n^{(1-p)n}} \geq \frac{\sqrt{2\pi}}{e^2\sqrt{n}},$$

since $p(1-p) \leq 1$.

When $\mu+n$ is odd, we set $a = pn - \frac{1}{2}$, $b = pn + \frac{1}{2}$. Again, Theorem 3 implies that the probability is minimized when all the bits are identically distributed. Setting $p' = p + \frac{1}{2n}$, we get $pn + \frac{1}{2} = p'n$, and we can lower bound this probability by:

$$\begin{aligned}\Pr\left[\sum_{i=1}^{n} Y_i = pn - \frac{1}{2}\right] + \Pr\left[\sum_{i=1}^{n} Y_i = pn + \frac{1}{2}\right] \geq & \qquad \Pr\left[\sum_{i=1}^{n} Y_i = pn + \frac{1}{2}\right] \\ \geq & \qquad \binom{n}{p'n} \cdot p^{p'n} \cdot (1-p)^{(1-p')n},\end{aligned}$$

which can be bounded using Stirling's approximation by:

$$\begin{aligned}&\geq \frac{\sqrt{2\pi}}{e^2\sqrt{n}} \cdot \frac{n^n \cdot p^{p'n} \cdot (1-p)^{(1-p')n}}{p'^{p'n} \cdot n^{p'n} \cdot (1-p')^{(1-p')n} \cdot n^{(1-p')n}} \\ &\geq \frac{\sqrt{2\pi}}{e^2\sqrt{n}} \cdot \left(\frac{p}{p'}\right)^{p'n},\end{aligned}$$

since $\frac{1-p}{1-p'} \geq 1$. Using the fact that $pn \geq 1/2$ and $e^x \geq 1 + x$, we can bound

$$\left(\frac{p}{p'}\right)^{p'n} = \left(\frac{1}{1+\frac{1}{2pn}}\right)^{pn+1/2} \geq \left(\frac{1}{e^{\frac{1}{2pn}}}\right)^{pn+\frac{1}{2}} = \frac{1}{e^{\frac{1}{2}+\frac{1}{4pn}}} \geq \frac{1}{e},$$

establishing the theorem when $\mu + n$ is odd.

6 Acknowledgements

Thanks to Pavel Hrubeš, Vladimir Podolskii and Amir Yehudayoff for useful conversations.

References

[AK10] Eric Allender and Michal Koucký. Amplifying lower bounds by means of self-reducibility. *J. ACM*, 57(3):14:1–14:36, 2010.

[AKS83] M. Ajtai, J. Komlós, and E. Szemerédi. Sorting in $c \log n$ parallel steps. *Combinatorica*, 3:1–19, 1983.

[BF81] József Beck and Tibor Fiala. "integer-making" theorems. *Discrete Applied Mathematics*, 3(1):1–8, 1981.

[Hoe56] Wassily Hoeffding. On the distribution of the number of successes in independent trials. *Annals of Mathematical Statistics*, 27:713–721, 1956.

[HR] Pavel Hrubeš and Anup Rao. Personal communication.

[Juk12] Stasys Jukna. *Boolean Function Complexity - Advances and Frontiers*, volume 27 of *Algorithms and combinatorics*. Springer, 2012.

[KP17] Alexander S. Kulikov and Vladimir V. Podolskii. Computing majority by constant depth majority circuits with low fan-in gates. In *34th Symposium on Theoretical Aspects of Computer Science*, volume 66 of *LIPIcs*, pages 49:1–49:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

[RY09] Ran Raz and Amir Yehudayoff. Lower bounds and separations for constant depth multilinear circuits. *Computational Complexity*, 18(2):171–207, 2009.

[Val84] L. G. Valiant. Short monotone formulae for the majority function. *Journal of Algorithms*, 5(3):363–366, September 1984.