Simplified Lower Bounds on the Multiparty Communication Complexity of Disjointness

Anup Rao[*1] **and Amir Yehudayoff**[†2]

1 **Department of Computer Science and Engineering**
University of Washington, Seattle, USA
anuprao@cs.washington.edu
2 **Department of Mathematics**
Technion-IIT, Israel
amir.yehudayoff@gmail.com

Abstract

We show that the deterministic number-on-forehead communication complexity of set disjointness for k parties on a universe of size n is $\Omega(n/4^k)$. This gives the first lower bound that is linear in n, nearly matching Grolmusz's upper bound of $O(\log^2(n) + k^2 n/2^k)$. We also simplify the proof of Sherstov's $\Omega(\sqrt{n}/(k2^k))$ lower bound for the randomized communication complexity of set disjointness.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases communication complexity, number-on-forehead model, set disjointness, lower bounds

Digital Object Identifier 10.4230/LIPIcs.CCC.2015.88

1 Introduction

Given a family of k sets $\mathcal{F} = (X_1, \ldots, X_k)$ over the universe $[n]$, the *disjointness* function is defined as

$$\mathsf{Disjoint}(\mathcal{F}) = \begin{cases} 1 & \text{if } \bigcap_{i=1}^{k} X_i = \emptyset, \\ 0 & \text{otherwise.} \end{cases}$$

We study the communication complexity of computing disjointness in the number-on-forehead model [9]. We consider k parties that attempt to compute $\mathsf{Disjoint}(\mathcal{F})$ by exchanging messages about $X_1, \ldots, X_k$, until one of the parties announces the value of $\mathsf{Disjoint}(\mathcal{F})$. The i'th party can see all of the inputs except for X_i, and can send messages that depend on the inputs she sees and all previous messages. All messages are visible to all parties. The communication complexity is the minimum number of bits that needs to be transmitted to compute $\mathsf{Disjoint}(\mathcal{F})$. In a randomized communication protocol, the parties use shared randomness to pick a deterministic communication protocol, and then run the chosen deterministic protocol. The protocol computes $\mathsf{Disjoint}(\mathcal{F})$ correctly if it outputs $\mathsf{Disjoint}(\mathcal{F})$ with probability at least $2/3$, for every family $\mathcal{F}$. For formal definitions of multiparty communication complexity and its significance, we refer the reader to [23].

* Supported by an Alfred P. Sloan Fellowship, the National Science Foundation under agreement CCF-1016565, an NSF Career award, and by the Binational Science Foundation under agreement 2010089.
† Horev Fellow – supported by the Taub Foundation. Supported by the Israel Science Foundation and by the Binational Science Foundation under agreement 2010089.

30th Conference on Computational Complexity (CCC'15).
Editor: David Zuckerman; pp. 88–101

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Grolmusz [19] gave a beautiful deterministic protocol showing that $\mathsf{Disjoint}(\mathcal{F})$ can be computed deterministically with communication $O(\log^2(n) + k^2 n/2^k)$. Chattopadhyay [11] used similar ideas to give a protocol with communication $O(k\log(n) + n/2^k)$. This paper is about proving lower bounds on the communication complexity.

1.1 Motivation and related work

Lower bounds on multiparty communication complexity are important because several computational models such as circuits, branching programs, and propositional proofs can be used to obtain efficient communication protocols. Strong enough communication complexity lower bounds for the computation of any explicit function can therefore be used to prove lower bounds on these models [3, 15, 2, 30, 40]. In particular, lower bounds on the communication complexity of disjointness have many applications (see the recent survey [13]). Such lower bounds imply lower bounds on proof systems [6], circuit lower bounds [21, 32, 28, 39], lower bounds on communication for problems related to combinatorial auctions [16, 26, 25, 17, 20, 29], and oracle separations for complexity classes [1].

Attempts to prove lower bounds for disjointness have led to many interesting ideas. When the number of parties is $k = 2$, Kalyanasundaram and Schnitger [22] proved that $\Omega(n)$ communication is required in the randomized setting. Alternate proofs and tight bounds have since been obtained [31, 4, 8] using methods involving information theory. These methods have found many other applications that we do not discuss here.

When k is large, Tesson [38] and Beame, Pitassi, Segerlind and Wigderson [7] proved that the deterministic communication complexity is $\Omega(\log(n)/k)$. Then Sherstov [34, 35] introduced the *pattern matrix method* for proving lower bounds in the case $k = 2$. The method was used to separate certain circuit classes by relating their complexity to analytic properties of boolean functions, like their approximate degree. This technique was generalized to $k > 2$ by Chattopadhyay [10], Lee and Shraibman [24], and Chattopadhyay and Ada [12]. These last two papers proved lower bounds of the type $\Omega\big(n^{1/(k+1)}/2^{2^{O(k)}}\big)$ on the randomized communication complexity. Beame and Huynh-Ngoc [5] extended these methods further to prove that the randomized communication complexity is at least $2^{\Omega\left(\sqrt{\log(n)/k}\right)}2^{-k}$. Finally, Sherstov [36, 37] proved the best known lower bounds prior to our work, showing that the randomized communication complexity is at least $\Omega(\sqrt{n}/(k2^k))$. In fact, Sherstov proved lower bounds for a broader class of functions, as we discuss below.

These results use powerful techniques such as Fourier analysis, Gowers norms, directional derivatives, and bounds on the approximate degree. The last two works of Sherstov are the main inspiration for our work.

1.2 Results

In what follows, k is the number of players in the number-on-forehead model, and n is the size of the universe. For an integer m, we denote by $[m]$ the set $\{1, 2, \ldots, m\}$, and for two real numbers a and b, we denote by $[a, b]$ the interval $\{x \in \mathbb{R} : a \leq x \leq b\}$.

Our work follows the ideas in the recent papers of Sherstov [36, 37]. We prove a linear lower bound on the deterministic multiparty communication complexity o f disjointness:

▶ **Theorem 1.1.** *The deterministic communication complexity of disjointness is* $\Omega\left(\frac{n}{4^k}\right)$.

Given our interpretation of Sherstov's work in [36], the proof of Theorem 1.1 is short. We also simplify the proof of the randomized lower bound from [37]:

▶ **Theorem 1.2** ([37]). *The randomized communication complexity of disjointness is* $\Omega\left(\frac{\sqrt{n}}{k2^k}\right)$.

Sherstov proved lower bounds for functions of the type $f(\mathsf{Disjoint}(\mathcal{F}_1), \ldots, \mathsf{Disjoint}(\mathcal{F}_m))$, where f is a multivariate function, and $\mathcal{F}_1, \ldots, \mathcal{F}_m$ are families on disjoint parts of the universe. In our proof, we focus on the case where f is symmetric. We symmetrize his proof by viewing f as a univariate function $f : \{0, 1, \ldots, m\} \to \{0, 1\}$, rather than as a multivariate function.

The proof begins by bounding the *discrepancy* of the parity of several independent instances of disjointness. Here we use two different bounds that Sherstov proved [36, 37], as black boxes. The first bound, stated as Theorem 2.1 in this paper, is used for the deterministic case, and the second, stated as Theorem 2.2, is used in the randomized case.

The rest of Sherstov's proof of Theorem 1.2 is a method to control the error in an approximation of the function f, using the bounds on the discrepancy. In the symmetrized proof, this corresponds to a bound on the error in an approximation of the Kronecker delta function (i.e. the univariate function f that is the indicator of m). We bound the error via the following theorem, which shows that every polynomial that is not correlated with any parity has a low-degree approximation:

▶ **Theorem 1.3.** *Let m be a power of 2. For $j \in [m]$, let J denote the smallest power of 2 such that $J \geq j$. Let $Y_1, \ldots, Y_m \in \{0, 1\}$ be distributed uniformly and independently. Suppose f is a real univariate polynomial of degree at most m, and $\delta \geq 0$ is such that for every $j \geq d > 0$,*

$$\left|\mathbb{E}\left[f((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j}\right]\right| \leq 2^{-12J}\delta. \tag{1}$$

Then there exists a polynomial g of degree at most $d - 1$ such that $|g(x) - f(x)| \leq \delta$ for all $x \in [0, m]$.

To prove Theorem 1.3, we define a useful basis $b_0(x), \ldots, b_m(x)$ for the space of polynomials, where each b_i is of degree i. Given this basis, the polynomial g is just the projection of f to the space spanned by $b_0, \ldots, b_{d-1}$. This basis may be of independent interest. For the analogous part of the proof, Sherstov finds a low-degree approximation of f using a different basis. We prove Theorem 1.3 in Section 3.

Finally, we state a corollary that may be useful in other applications.

▶ **Corollary 1.4.** *There exists a function $\ell(k) \leq O(k^2 4^k)$ with the following property. Suppose each family $\mathcal{F}_i$, $i \in [m]$, is supported on a disjoint universe of size ℓ. Let $f : \{0, 1, \ldots, m\} \to \{0, 1\}$ be an arbitrary function. If the randomized communication complexity of $f(\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i))$ is C, then there exists a polynomial g of degree at most $C - 1$ such that $|f(x) - g(x)| \leq 1/3 + 2^{-3C}$ for all $x \in \{0, 1, \ldots, m\}$.*

The proof of the corollary follows easily from the ideas in Section 2.2.

2 The lower bounds

Without loss of generality, we assume that $n = m\ell$, where m is a power of 2 and ℓ is a function of k to be determined. Any family $\mathcal{F} = (X_1, \ldots, X_k)$ can be described using the m families $\mathcal{F}_1, \ldots, \mathcal{F}_m$, each over a universe of size ℓ, defined as

$$\mathcal{F}_i = (X_1 \cap [(i-1)\ell + 1, i\ell], \ldots, X_k \cap [(i-1)\ell + 1, i\ell]). \tag{2}$$

Distribution μ on $\mathcal{F}_1 \subseteq 2^{[\ell]}$

Let $S_1, \ldots, S_{k-1} \subseteq [\ell]$ be uniformly random sets conditioned on $|S_1 \cap S_2 \cap \ldots \cap S_{k-1}| = 1$. Let $S_k \subseteq [\ell]$ be uniform and independent. Set

$$\mathcal{F}_1 = (S_1, \ldots, S_{k-1}, S_k).$$

Figure 1 The distribution μ.

Moreover,

$$\mathsf{Disjoint}(\mathcal{F}) = \begin{cases} 1 & \text{if } \sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i) = m, \\ 0 & \text{otherwise.} \end{cases}$$

In order to prove Theorems 1.1 and 1.2, we consider distributions on families $\mathcal{F}$, where each $\mathcal{F}_i$ is independent and identically distributed. Sherstov shows that there are distributions of this type under which every protocol with small communication complexity must have low correlation with $(-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)}$.

2.1 The lower bound for deterministic protocols

Consider the distribution μ given in Figure 1 as a way to sample each $\mathcal{F}_i$. The following theorem is an easy consequence of Theorem 4.2 in [36], and the fact that every communication protocol can be expressed as a sum of cylinder intersections:

▶ **Theorem 2.1** ([36]). *If each family $\mathcal{F}_i$ is sampled independently according to μ, and π is a k party protocol with communication complexity C, then*

$$\left| \mathbb{E}\left[\pi(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)} \right] \right| \leq 2^C \cdot \left(\frac{2^{k-1}-1}{\sqrt{\ell}} \right)^m .$$

The proof of Theorem 2.1 involves ideas analogous to [3] and some subtle reasoning about the distribution μ. For completeness, we give a full exposition of the proof in Appendix B. Given Theorem 2.1, Theorem 1.1 easily follows:

Proof of Theorem 1.1. Let π be a deterministic protocol that computes $\mathsf{Disjoint}(\mathcal{F})$ with communication complexity C. When $\mathsf{Disjoint}(\mathcal{F}) = 1$, we have $(-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)} = (-1)^m$ and $\pi(\mathcal{F}) = 1$. On the other hand, when $\mathsf{Disjoint}(\mathcal{F}) = 0$, we have $\pi(\mathcal{F}) = 0$. In addition, $\Pr[\mathsf{Disjoint}(\mathcal{F}_i) = 1] = 1/2$ for all $i \in [m]$, which implies $\Pr[\mathsf{Disjoint}(\mathcal{F}) = 1] = 2^{-m}$. Thus,

$$\left| \mathbb{E}\left[\pi(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)} \right] \right| = 2^{-m}. \tag{3}$$

Now set $\ell = 16(2^{k-1}-1)^2$. Theorem 2.1 and (3) imply that

$$2^C \cdot ((2^{k-1}-1)/\sqrt{\ell})^m \geq 2^{-m} \;\Rightarrow\; C \geq m = \Omega\left(\frac{n}{4^k}\right).$$

◀

Distribution γ on $\mathcal{F}_1 \subseteq 2^{[\ell]}$

We define a distribution on $k \times \ell$ boolean matrices. Given a matrix M sampled from this distribution, define the family $\mathcal{F}_1$ by setting the i'th set $S_i = \{j \in [\ell] : M_{i,j} = 1\}$. Let t be such that $\ell = 2^{k-1} + t(2^k - 1)$. Sample M as follows:

1. For each $v \in \{0,1\}^k$ such that $v \neq 1^k$, let M have t columns equal to v.
2. Let $b \in \{0,1\}$ be uniformly random. For each $u \in \{0,1\}^k$ such that $\sum_{i=1}^k u_i = b \mod 2$, add a column to M that is equal to u.
3. Permute the columns of M using a uniformly random permutation of $[\ell]$.

Figure 2 The distribution γ.

2.2 The lower bound for randomized protocols

The proof of Theorem 1.1 does not give anything meaningful in the randomized setting, since it may be the case that a randomized protocol has no correlation with $(-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)}$. To prove lower bounds on the randomized communication, following Sherstov, we use a more complicated distribution on inputs, as well as approximation theory.

For the rest of this section, we work with the distribution γ described in Figure 2. Note that under this distribution, $\Pr[\mathsf{Disjoint}(\mathcal{F}_1) = 1] = 1/2$. A crucial feature of this distribution is the following symmetric structure: if $\rho : [\ell] \to [\ell]$ is a uniformly random permutation independent of $\mathcal{F}_1$, then the families $(\mathcal{F}_1, \rho(\mathcal{F}_1))$ have the same joint distribution as two independent samples $(\mathcal{F}_1, \mathcal{F}_1')$ from γ conditioned on $\mathsf{Disjoint}(\mathcal{F}_1) = \mathsf{Disjoint}(\mathcal{F}_1')$. Here by $\rho(\mathcal{F}_1)$ we mean the family obtained by permuting the underlying universe. In analogy with Theorem 2.1, Sherstov shows (Corollary 4.19 in [37]) that no protocol can be significantly correlated with $(-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)}$ under the distribution γ:

▶ **Theorem 2.2** ([37]). *If each family $\mathcal{F}_i$ is sampled independently according to γ, and π is a protocol with communication complexity C, then*

$$\left|\mathbb{E}\left[\pi(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i)}\right]\right| \leq 2^C \cdot \left(\frac{c_0 k^2 4^k}{\ell}\right)^{m/4},$$

where $c_0 > 0$ is a universal constant.

The proof of Theorem 2.2 is delicate, mainly due to the symmetric structure of the distribution γ (especially if one wishes to optimize the dependence on k). This symmetric structure is, on the other hand, very useful, and we shall exploit it next.

Given any protocol π computing $\mathsf{Disjoint}(\mathcal{F})$, define f_π as the unique degree m polynomial so that for all $t \in \{0, 1, \ldots, m\}$,

$$f_\pi(t) = \Pr\left[\pi(\mathcal{F}) = 1 \,\middle|\, \sum_{i=1}^m \mathsf{Disjoint}(\mathcal{F}_i) = t\right]. \tag{4}$$

Since the protocol computes $\mathsf{Disjoint}(\mathcal{F})$ with probability at least $2/3$, we have that $|f_\pi(t)| \leq 1/3$, for $t = 0, 1, \ldots, m-1$, and $|1 - f_\pi(m)| \leq 1/3$. The following well known theorem [18, 33, 27] shows that any such function must have degree $\sqrt{m}/3$:

▶ **Theorem 2.3** ([18, 33, 27]). *Let $\epsilon \in (0, 1/2)$. If $f : [0, m] \to \mathbb{R}$ is a polynomial such that $|f(t)| \leq \epsilon$ for $t = 0, 1, \ldots, m-1$, and $|1 - f(m)| \leq \epsilon$, then the degree of f is at least $\sqrt{m(1-2\epsilon)/3}$.*

Protocol $\tau_{\pi,j}(\mathcal{F}_1,\dots,\mathcal{F}_j)$

1. Let J denote the smallest power of 2 such that $J \geq j$. Note that m/J is an integer, since m is assumed to be a power of 2.
2. Using public randomness, sample $J-j$ families $\mathcal{F}_{j+1},\mathcal{F}_{j+2},\dots,\mathcal{F}_J$ according to γ, conditioned on the event that $\mathsf{Disjoint}(\mathcal{F}_{j+1}) = \mathsf{Disjoint}(\mathcal{F}_{j+2}) = \mathsf{Disjoint}(\mathcal{F}_J) = 0$.
3. Let $\mathcal{G} = (\mathcal{F}_1,\dots,\mathcal{F}_1,\mathcal{F}_2,\dots,\mathcal{F}_2,\dots,\mathcal{F}_J,\dots,\mathcal{F}_J)$ be the m families obtained by repeating each family $\mathcal{F}_i$ exactly m/J times.
4. Let $\rho_1,\rho_2,\dots,\rho_m : [\ell] \to [\ell], \eta : [m] \to [m]$ be independent uniformly random permutations chosen using public randomness.
5. Output $\pi(\rho_1(\mathcal{G}_{\eta(1)}),\rho_2(\mathcal{G}_{\eta(2)}),\dots,\rho_m(\mathcal{G}_{\eta(m)}))$.

Figure 3 The protocol $\tau_{\pi,j}$.

Theorem 2.3 is proved via a clever reduction to Markov's bound on the magnitude of derivatives in bounded polynomials. We include the short proof in Appendix A. We remark that Theorem 2.3 is tight — one can use Chebyshev polynomials to give a polynomial f of degree $O(\sqrt{m})$ satisfying the constraints. We shall prove that if the communication of π is much less than $\sqrt{n}/(k2^k)$, then Theorems 2.2 and 1.3 imply that f_π can be approximated by a polynomial whose degree is much less than $\sqrt{m}$, contradicting Theorem 2.3.

We analyze the behavior of π under several carefully chosen input distributions, and use the symmetric structure of the distributions together with Sherstov's correlation bounds to show that f_π has low correlation with parity. We then appeal to Theorem 1.3 to conclude that f_π has a low degree approximation. We formalize this plan by describing m protocols $\tau_{\pi,1},\dots,\tau_{\pi,m}$, each simulating π with a different distribution on inputs.

Define the protocol $\tau_{\pi,j}$ as in Figure 3. The protocol $\tau_{\pi,j}$ takes j families of sets $\mathcal{F}_1,\dots,\mathcal{F}_j$. If each of $\mathcal{F}_1,\dots,\mathcal{F}_j$ is in the support of γ, then the protocol $\tau_{\pi,j}$, using shared public randomness and no communication, generates m families $\mathcal{H}_1,\dots,\mathcal{H}_m$ that are independently distributed according to γ, conditioned on

$$\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{H}_i) = (m/J)\sum_{i=1}^{j} \mathsf{Disjoint}(\mathcal{F}_i).$$

This distribution of $\mathcal{H}_1,\dots,\mathcal{H}_m$ is as stated due to the symmetric structure of γ, which is discussed in the second paragraph of this section. Finally, $\tau_{\pi,j}$ simulates π on $\mathcal{H}_1,\dots,\mathcal{H}_m$.

The key properties of $\tau_{\pi,j}$ are summarized in the following lemma.

▶ **Lemma 2.4.** *Let $j \leq m$.*

1. *The communication complexity of $\tau_{\pi,j}$ equals that of π.*
2. *Let $\mathcal{F}_1,\dots,\mathcal{F}_j$ be fixed families of sets, each in the support of γ. Then,*

$$\Pr[\tau_{\pi,j}(\mathcal{F}_1,\dots,\mathcal{F}_j) = 1] = f_\pi\left(\left(\sum_{i=1}^{j} \mathsf{Disjoint}(\mathcal{F}_i)\right) m/J\right).$$

Lemma 2.4 and Theorem 2.2 together imply that the correlation of f_π with parity is small:

▶ **Lemma 2.5.** *Let J be the smallest power of 2 so that $J \geq j$. If the communication complexity of π is C, and $Y_1, \ldots, Y_j \in \{0,1\}$ are uniformly random and independent, then*

$$\mathbb{E}\left[f_\pi((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j}\right] \leq 2^C \cdot \left(\frac{c_0 k^2 4^k}{\ell}\right)^{j/4},$$

where $c_0 > 0$ is a universal constant.

Proof. If $\mathcal{F}_1$ is distributed according to γ, then $\mathsf{Disjoint}(\mathcal{F}_1)$ is a uniformly random bit. Thus,

$$\begin{aligned}
&\mathbb{E}\left[f_\pi((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j}\right] \\
&= \mathbb{E}\left[f_\pi\left(\left(\sum_{i=1}^{j} \mathsf{Disjoint}(\mathcal{F}_i)\right) m/J\right) \cdot (-1)^{\sum_{i=1}^{j} \mathsf{Disjoint}(\mathcal{F}_i)}\right]
\end{aligned}$$

Using Lemma 2.4,

$$= \mathbb{E}\left[\tau_{\pi,j}(\mathcal{F}_1, \ldots, \mathcal{F}_j) \cdot (-1)^{\sum_{i=1}^{j} \mathsf{Disjoint}(\mathcal{F}_i)}\right] \leq 2^C \cdot \left(\frac{c_0 k^2 4^k}{\ell}\right)^{j/4},$$

where the last inequality is by Theorem 2.2, since the communication complexity of $\tau_{\pi,j}$ is equal to that of π. ◀

Given Lemma 2.5, the proof is completed as follows:

Proof of Theorem 1.2. We set $\ell = 2^{16 \cdot 4} c_0 k^2 4^k$, so that the right hand side of Lemma 2.5 is 2^{C-16j}. Fix any randomized protocol π that computes $\mathsf{Disjoint}(\mathcal{F})$ on the distribution induced by γ with C bits of communication. Let f_π be as defined in (4).

By Lemma 2.5, f_π satisfies the hypothesis of Theorem 1.3, with $d = C$. Thus we conclude that there is a degree $C-1$ polynomial g that agrees with f_π up to an error of 2^{-3C}. Theorem 2.3 implies that

$$C \geq \sqrt{m(1 - 2(1/3 + 2^{-3C}))/3} \;\Rightarrow\; C \geq \Omega(\sqrt{n}/(k2^k)).$$

◀

3 Approximating functions that are not correlated with parity

Here we prove Theorem 1.3, which shows that if a polynomial has low correlation with parity, then it can be approximated by a low degree polynomial. We restate the theorem for convenience.

▶ **Theorem 1.3 (restated).** *Let m be a power of 2. For $j \in [m]$, let J denote the smallest power of 2 such that $J \geq j$. Let $Y_1, \ldots, Y_m \in \{0,1\}$ be distributed uniformly and independently. Suppose f is a real univariate polynomial of degree at most m, and $\delta \geq 0$ is such that for every $j \geq d > 0$,*

$$\left|\mathbb{E}\left[f((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j}\right]\right| \leq 2^{-12J}\delta. \tag{5}$$

Then there exists a polynomial g of degree at most $d-1$ such that $|g(x) - f(x)| \leq \delta$ for all $x \in [0, m]$.

In what follows, let $Y_1, \ldots, Y_m \in \{0,1\}$ be independent and uniformly random bits, and let I, J be the smallest powers of 2 such that $I \geq i$ and $J \geq j$.

To prove the theorem, we define a useful basis for the space of polynomials. Let $b_0(x) = 1$. For $i > 0$, let[1]

$$b_i(x) = 2^i \binom{xI/m}{i} = \frac{2^i x(x - m/I)(x - 2m/I)\ldots(x - (i-1)m/I)}{i! \cdot (m/I)^i}.$$

Since b_i is of degree i, the polynomials $b_0, \ldots, b_m$ form a basis for the space of polynomials of degree at most m. To prove Theorem 1.3, we express f in this basis and then argue that all coefficients corresponding to high degree terms are negligible. The polynomials in our basis can be bounded by the following lemma:

▶ **Lemma 3.1.** *For every $i \in \{0, 1, \ldots, m\}$, $\max_{x \in [0,m]} |b_i(x)| \leq 8^i$.*

Proof. We show that the maximum of b_i is attained when $x = m$, and so

$$\max_{x \in [0,m]} |b_i(x)| = |b_i(m)| = 2^i \binom{mI/m}{i} \leq 2^i \cdot 2^I \leq 8^i.$$

Note that the magnitude of b_i is symmetric around the point $(i-1)m/(2I)$,

$$|b_i(x + (i-1)m/(2I))| = |b_i(-x + (i-1)m/(2I))|.$$

So the maximum is attained with $x \in [(i-1)m/(2I), m]$. For any such x that is not a root of b_i,

$$\left|\frac{b_i(x + m/I)}{b_i(x)}\right| = \left|\frac{x + m/I}{x - (i-1)m/I}\right| \geq \left|\frac{x + m/I}{x}\right| > 1,$$

proving that the maximum is attained with $x \in [m - m/I, m]$. For such x, every term $(x - jm/I)$ with $j \in \{0, 1, \ldots, i-1\}$ in $b_i(x)$ is non-negative, and so the maximum is attained when $x = m$. ◀

The basis polynomials behave nicely under the random experiments from (1):

▶ **Lemma 3.2.** *For all $i \in \{0, 1, 2, \ldots, m\}$ and $j \in [m]$,*

$$\left|\mathbb{E}\left[b_i((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1 + \ldots + Y_j}\right]\right| \begin{cases} = 0 & \text{if } i < j, \\ = 1 & \text{if } i = j, \\ = 0 & \text{if } j < i \leq J, \\ \leq 8^i & \text{if } J < i. \end{cases}$$

Proof. When $i < j$, the polynomial $b_i(y_1 + \ldots + y_j)$ has degree i in the variables $y_1, \ldots, y_j$. Since every monomial must exclude one of the j variables, the contribution of each of the monomials to the expectation is 0. When $i = j$, $b_i((y_1 + \ldots + y_i)m/I) = 2^i \binom{y_1 + \ldots + y_i}{i}$ is non-zero only when $y_1 = y_2 = \ldots = y_i = 1$. Thus the expectation is $2^{-i} \cdot 2^i \binom{i}{i} = 1$ in this case. When $j < i \leq J$, we have $I = J$. Since for $r \in [i-1]$, $b_i(rm/I) = 2^i \binom{r}{i} = 0$, the expectation is 0. When $i > J$, by Lemma 3.1, the expectation is at most 8^i. ◀

Theorem 1.3 now follows by straightforward induction:

[1] Here and below we think of $\binom{x}{i} = \frac{x(x-1)(x-2)\ldots(x-(i-1))}{i!}$ as a real polynomial in the variable x.

Proof of Theorem 1.3. Write $f(x) = \sum_{j=0}^{m} a_j b_j(x)$, and let $g(x)$ be the degree $d-1$ polynomial $g(x) = \sum_{j=0}^{d-1} a_j b_j(x)$. To prove the theorem, we show that $|g(x) - f(x)| \leq \delta$ for all $x \in [0, m]$. Lemma 3.2 and (1) imply that for $j = d, \ldots, m$,

$$|a_j| - \sum_{i=J+1}^{m} 8^i |a_i| \leq |\,\mathbb{E}\left[f((Y_1 + \ldots + Y_j)m/J) \cdot (-1)^{Y_1+\ldots+Y_j}\right]\,| \leq 8^{-4J}\delta$$

$$\Rightarrow |a_j| \leq 8^{-4J}\delta + \sum_{i=J+1}^{m} 8^i |a_i|. \tag{6}$$

We now prove by induction that for $j = m, m-1, \ldots, d$,

$$\sum_{t=j}^{J} |a_t| \leq 8^{-3J}\delta. \tag{7}$$

When $m/2 < j \leq m$, (6) implies

$$\sum_{t=j}^{m} |a_t| \leq (m/2)8^{-4m}\delta \leq 8^{-3m}\delta.$$

In the general case (6) implies

$$(2/J)\sum_{t=j}^{J} |a_t| \leq 8^{-4J}\delta + \sum_{t=J+1}^{m} 8^t |a_t| \leq 8^{-4J}\delta + \sum_{r=\log(J)+1}^{\log(m)} 8^{2^r} \sum_{t=1+2^{r-1}}^{2^r} |a_t|$$

Applying the induction hypothesis, we get

$$\leq 8^{-4J}\delta + \sum_{r=\log(J)+1}^{\log(m)} 8^{2^r} 8^{-3\cdot 2^r}\delta \leq 8^{-4J}\delta + 8^{-4J}\delta \sum_{q=0}^{\infty} 8^{-q} \leq 8^{-3J}(2/J)\delta,$$

which proves the general case of (7).

Finally, for every $x \in [0, m]$, Lemma 3.1 and (7) imply

$$|g(x) - f(x)| \leq \sum_{j=d}^{m} |a_j||b_j(x)| \leq \sum_{r=\lceil \log d \rceil}^{\log m} 8^{-3\cdot 2^r}\delta \cdot 8^{2^r} \leq 8^{-2d}\delta \sum_{q=0}^{\infty} 8^{-2q} \leq \delta.$$

◀

Acknowledgements. We thank Paul Beame, Pavel Hrubeš, and Alexander Sherstov for useful discussions.

References

1 Scott Aaronson and Avi Wigderson. Algebrization: A new barrier in complexity theory. *TOCT*, 1(1), 2009.

2 László Babai, Thomas P. Hayes, and Peter G. Kimmel. The cost of the missing bit: Communication complexity with help. *Combinatorica*, 21(4):455–488, 2001.

3 László Babai, Noam Nisan, and Mario Szegedy. Multiparty protocols, pseudorandom generators for logspace, and time-space trade-offs. *J. Comput. Syst. Sci.*, 45(2):204–232, 1992.

4 Ziv Bar-Yossef, T. S. Jayram, Ravi Kumar, and D. Sivakumar. An information statistics approach to data stream and communication complexity. *Journal of Computer and System Sciences*, 68(4):702–732, 2004.

5 Paul Beame and Dang-Trinh Huynh-Ngoc. Multiparty communication complexity and threshold circuit size of AC^0. In *FOCS*, pages 53–62, 2009.

6 Paul Beame, Toniann Pitassi, and Nathan Segerlind. Lower bounds for Lovász–Schrijver systems and beyond follow from multiparty communication complexity. *SIAM Journal on Computing*, 37(3):845–869, 2007.

7 Paul Beame, Toniann Pitassi, Nathan Segerlind, and Avi Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.

8 Mark Braverman and Ankur Moitra. An information complexity approach to extended formulations. In *STOC*, pages 161–170. ACM, 2013.

9 Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton. Multi-party protocols. In *STOC*, pages 94–99, 1983.

10 Arkadev Chattopadhyay. Discrepancy and the power of bottom fan-in in depth-three circuits. In *FOCS*, pages 449–458. IEEE Computer Society Press, 2007.

11 Arkadev Chattopadhyay. *Circuits, Communication and Polynomials*. PhD thesis, University of Toronto, 2008.

12 Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *CoRR*, abs/0801.3624, 2008.

13 Arkadev Chattopadhyay and Toniann Pitassi. The story of set disjointness. *SIGACT News*, 41(3):59–85, 2010.

14 Elliot W. Cheney. *Introduction to Approximation Theory*. McGraw-Hill Book Co., New York, 1966.

15 Fan R. K. Chung and Prasad Tetali. Communication complexity and quasi randomness. *SIAM Journal on Discrete Mathematics*, 6(1):110–123, February 1993.

16 Vincent Conitzer and Tuomas Sandholm. Communication complexity as a lower bound for learning in games. In Carla E. Brodley, editor, *ICML*, volume 69 of *ACM International Conference Proceeding Series*. ACM, 2004.

17 Shahar Dobzinski and Noam Nisan. Limitations of VCG-based mechanisms. *Combinatorica*, 31(4):379–396, 2011.

18 H. Ehlich and K. Zeller. Schwankung von polynomen zwischen gitterpunkten. *Mathematische Zeitschrift*, 86:41–44, 1964.

19 Vince Grolmusz. The bns lower bound for multi-party protocols in nearly optimal. *Inf. Comput.*, 112(1):51–54, 1994.

20 Sergiu Hart and Yishay Mansour. The communication complexity of uncoupled nash equilibrium procedures. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 345–353. ACM, 2007.

21 Johan Håstad and Mikael Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

22 Bala Kalyanasundaram and Georg Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, November 1992.

23 Eyal Kushilevitz and Noam Nisan. *Communication complexity*. Cambridge University Press, 1997.

24 Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

25 Noam Nisan. The communication complexity of approximate set packing and covering. *Lecture Notes in Computer Science*, 2380:868–875, 2002.

26 Noam Nisan and Ilya Segal. The communication requirements of efficient allocations and supporting prices. *J. Economic Theory*, 129(1):192–224, 2006.

27 Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.

28 Noam Nisan and Avi Wigderson. Rounds in communication complexity revisited. *SIAM Journal on Computing*, 22(1):211–219, February 1993.

29 Christos H. Papadimitriou, Michael Schapira, and Yaron Singer. On the hardness of being truthful. In *FOCS*, pages 250–259. IEEE Computer Society, 2008.

30 Ran Raz. The BNS-chung criterion for multi-party communication complexity. *Computational Complexity*, 9(2):113–122, 2000.

31 Alexander A. Razborov. On the distributional complexity of disjointness. *Theor. Comput. Sci.*, 106(2):385–390, 1992.

32 Alexander A. Razborov and Avi Wigderson. $n^{\omega}(logn)$ lower bounds on the size of depth-3 threshold circuits with AND gates at the bottom. *Information Processing Letters*, 45(6):303–307, 1993.

33 Theodore J. Rivlin and Elliott W. Cheney. A comparison of uniform approximations on an interval and a finite subset thereof. *SIAM Journal on Numerical Analysis*, 3(2):311–320, June 1966.

34 Alexander A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM Journal of Computing*, 38(6):2113–2129, 2009.

35 Alexander A. Sherstov. The pattern matrix method. *SIAM Journal of Computing*, 40(6):1969–2000, 2011.

36 Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In *STOC*, pages 525–548, 2012.

37 Alexander A. Sherstov. Communication lower bounds using directional derivatives. In *STOC*, pages 921–930, 2013.

38 Pascal Tesson. *Computational complexity questions related to finite monoids and semigroups*. PhD thesis, McGill University, 2003.

39 Emanuele Viola. Pseudorandom bits for constant-depth circuits with few arbitrary symmetric gates. *SIAM Journal on Computing*, 36(5):1387–1403, 2007.

40 Emanuele Viola and Avi Wigderson. Norms, XOR lemmas, and lower bounds for polynomials and protocols. *Theory of Computing*, 4(1):137–168, 2008.

A Approximation theory

The proof relies on a fundamental theorem of Markov, relating the degree of a bounded polynomial to the maximum value of its derivative.

▶ **Theorem 1.1** (Markov's Theorem [14]). *Let $g : [-1,1] \to [-1,1]$ be computed by a polynomial of degree d. Then $|g'(y)| \leq d^2$ for every $y \in [-1,1]$.*

Markov's theorem allows us to prove the statement about approximation that we need:

Proof of Theorem 2.3. Let d be the degree of f and let $D = \max_{x\in[0,m]} |f'(x)|$. We can bound f using D as follows. The value $|f(j)|$ is at most $1+\epsilon$ for $j \in \{0,1,\ldots,m\}$, and so $|f(x)| \leq 1+\epsilon+D/2$ for $x \in [0,m]$. On the other hand, $D \geq \frac{f(m)-f(m-1)}{1} \geq 1-2\epsilon$.

Now consider the degree d polynomial $g : [-1,1] \to [-1,1]$ given by $g(y) = \frac{f(my/2+m/2)}{1+\epsilon+D/2}$. Since $g'(y) = \frac{(m/2)f'(my/2+m/2)}{1+\epsilon+D/2}$, there is a $y \in [-1,1]$ such that $|g'(y)| = \frac{Dm/2}{1+\epsilon+D/2}$. By Theorem 1.1,

$$d^2 \geq \frac{Dm/2}{1+\epsilon+D/2} \geq \frac{m(1/2-\epsilon)}{1+\epsilon+1/2-\epsilon} = \frac{2m(1/2-\epsilon)}{3},$$

so

$$d \geq \sqrt{m(1-2\epsilon)/3}.$$

◀

B Bounding the discrepancy for the deterministic case

Here we give the proof of Theorem 2.1 [36]. Let $\mathcal{F} = (T_1, \ldots, T_k)$. We shall need to analyze the discrepancy on a more general class of distributions. Let each family $\mathcal{F}_i$ be sampled independently according to the distribution μ, on a universe of size ℓ_i. Let $g(\mathcal{F})$ be a cylinder intersection, that is, $g(\mathcal{F}) = \prod_{i=1}^{k} g_i(\mathcal{F})$ where each g_i is 0/1 valued and does not depend on T_i. We shall prove that

$$\left|\mathbb{E}\left[g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]\right| \leq \prod_{i=1}^{m} \frac{2^{k-1}-1}{\sqrt{\ell_i}}, \tag{8}$$

which implies Theorem 2.1, since every communication protocol with communication C can be expressed as a sum of 2^C cylinder intersections. We prove (8) by induction on k.

When $k = 2$, convexity implies that

$$\begin{aligned}
&\left|\mathbb{E}\left[g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]\right|^2 \\
&\leq \mathop{\mathbb{E}}_{T_2}\left[g_1(\mathcal{F}) \mathop{\mathbb{E}}_{T_1}\left[g_2(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]^2\right] \\
&\leq \mathop{\mathbb{E}}_{T_2}\left[\mathop{\mathbb{E}}_{T_1}\left[g_2(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]^2\right] \\
&= \mathop{\mathbb{E}}_{T_2, T_1, T_1'}\left[g_2(\mathcal{F}) \cdot g_2(\mathcal{F}') \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i) + \mathsf{Disjoint}(\mathcal{F}_i')}\right] \\
&\leq \mathop{\mathbb{E}}_{T_1, T_1'}\left[\left|\mathop{\mathbb{E}}_{T_2}\left[(-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i) + \mathsf{Disjoint}(\mathcal{F}_i')}\right]\right|\right],
\end{aligned}$$

where here $\mathcal{F} = (T_1, T_2)$ and $\mathcal{F}' = (T_1', T_2)$. Now for every fixing of T_1, T_1', the inner expectation is 1 when $T_1 = T_1'$, and otherwise it is 0. Thus,

$$\left|\mathbb{E}\left[g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]\right|^2 \leq \Pr[T_1 = T_1'] = \prod_{i=1}^{m} \frac{1}{\ell_i},$$

proving the base case.

When $k > 2$, we again use convexity to bound

$$\begin{aligned}
&\left|\mathbb{E}\left[g(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]\right|^2 \\
&\leq \mathop{\mathbb{E}}_{T_2, \ldots, T_k}\left[g_1(\mathcal{F}) \mathop{\mathbb{E}}_{T_1}\left[\prod_{j=2}^{k} g_j(\mathcal{F}) \cdot (-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i)}\right]^2\right] \\
&\leq \mathop{\mathbb{E}}_{T_1, T_1'}\left[\left|\mathop{\mathbb{E}}_{T_2, \ldots, T_k}\left[\left(\prod_{j=2}^{k} g_j(\mathcal{F}) \cdot g_j(\mathcal{F}')\right)(-1)^{\sum_{i=1}^{m} \mathsf{Disjoint}(\mathcal{F}_i) + \mathsf{Disjoint}(\mathcal{F}_i')}\right]\right|\right],
\end{aligned} \tag{9}$$

where here $\mathcal{F} = (T_1, T_2, \ldots, T_k)$ and $\mathcal{F}' = (T_1', T_2, \ldots, T_k)$. Recall that the first $k-1$ sets of $\mathcal{F}_i$ and $\mathcal{F}_i'$ each intersect in exactly one element. Let $Z = (Z_1, Z_2, \ldots, Z_m)$, where Z_i is the indicator random variable for the event that these two elements are not the same in $\mathcal{F}_i$ and $\mathcal{F}_i'$. Let $Q = T_1 \setminus T_1'$, $Q' = T_1' \setminus T_1$, and denote by Q_i, Q_i' the intersection of these sets with the i'th part of the universe (see (2)). Let R denote all the intersections of the sets $T_2, \ldots, T_k$

with the elements that are not in Q, Q'. By convexity of the absolute value function,

$$(9) \leq \mathop{\mathbb{E}}_{T_1,T_1',R,Z}\left[\left|\mathop{\mathbb{E}}_{T_2,\ldots,T_k}\left[\left(\prod_{j=2}^{k} g_j(\mathcal{F})\cdot g_j(\mathcal{F}')\right)(-1)^{\sum_{i=1}^{m}\mathsf{Disjoint}(\mathcal{F}_i)+\mathsf{Disjoint}(\mathcal{F}_i')}\right]\right|\right]$$
$$\leq \mathop{\mathbb{E}}_{Z,Q,Q'}\left[\prod_{i:Z_i=1}\frac{(2^{k-2}-1)^2}{\sqrt{|Q_i|\cdot|Q_i'|}}\right], \tag{10}$$

where the last inequality follows from the fact that after fixing T_1, T_1', Z, R, the inner expectation can be bounded by the inductive hypothesis applied to the families where $Z_i = 1$, over the disjoint universes Q_i, Q_i', and the cylinder intersection defined by $\prod_{j=2}^{k} g_j(\mathcal{F})g_j(\mathcal{F}')$. Apply the arithmetic-mean-geometric-mean inequality to conclude that

$$(10) \leq \mathop{\mathbb{E}}_{Z,Q,Q'}\left[\prod_{i:Z_i=1}(2^{k-2}-1)^2\frac{1}{2}\left(\frac{1}{|Q_i|}+\frac{1}{|Q_i'|}\right)\right]. \tag{11}$$

Since (even conditioned on the value of Z) the size of Q_i is distributed identically to the size of Q_i', we have

$$(11) = \mathop{\mathbb{E}}_{Z,Q}\left[\prod_{i:Z_i=1}\frac{(2^{k-2}-1)^2}{|Q_i|}\right]$$
$$\leq \prod_{i=1}^{m}\left(\Pr[Z_i=0]+\mathop{\mathbb{E}}_{Z_i,Q_i}\left[\frac{Z_i(2^{k-2}-1)^2}{|Q_i|}\right]\right), \tag{12}$$

where here we adopt the convention that $Z_i/|Q_i|$ is 0 when $Z_i = 0, |Q_i| = 0$. We shall prove that for all i,

$$\Pr[Z_i = 0] \leq \frac{2^{k-1}-1}{\ell_i}, \tag{13}$$

and

$$\mathop{\mathbb{E}}_{Z_i,Q_i}\left[\frac{Z_i}{|Q_i|}\right] \leq \frac{2(2^{k-1}-1)}{\ell_i(2^{k-2}-1)}. \tag{14}$$

Inequalities (13) and (14) imply that

$$(12) \leq \prod_{i=1}^{m}\left(\frac{2^{k-1}-1}{\ell_i}+\frac{2(2^{k-1}-1)(2^{k-2}-1)^2}{\ell_i(2^{k-2}-1)}\right)$$
$$= \prod_{i=1}^{m}\left(\frac{2^{k-1}-1+(2^{k-1}-1)(2^{k-1}-2)}{\ell_i}\right) = \prod_{i=1}^{m}\frac{(2^{k-1}-1)^2}{\ell_i},$$

as required.

It only remains to prove (13) and (14). Fix i for the rest of the proof. We start with (13). Let $S_1, \ldots, S_k$ be the intersections of $T_1, \ldots, T_k$ with the i'th part of the universe, and let S_1' be the intersection of T_1' with the i'th part of the universe. Observe that for all $w \neq \emptyset$,

$$\Pr\left[Z_i = 0 \,\middle|\, \cap_{j=2}^{k-1} S_j = w\right] = \frac{1}{|w|}.$$

The total number of choices for sets $S_1, \ldots, S_{k-1}$ can be counted as the number of ways to pick the common intersection point, times the number of configurations for the rest of the

universe: $\ell_i \cdot (2^{k-1}-1)^{\ell_i - 1}$. Of these, the number of configurations with $\cap_{j=2}^{k-1} S_j = w$ can be counted as the number of choices for the common intersection point in w, times the number of configurations for the rest of the universe: $|w| \cdot (2^{k-1}-2)^{\ell_i - |w|}$. So the probability that the two intersection points are the same is

$$\begin{aligned}\Pr[Z_i = 0] &= \frac{1}{\ell_i \cdot (2^{k-1}-1)^{\ell_i - 1}} \cdot \sum_{w \neq \emptyset} \frac{|w| \cdot (2^{k-1}-2)^{\ell_i - |w|}}{|w|} \\ &\leq \frac{1}{\ell_i \cdot (2^{k-1}-1)^{\ell_i - 1}} \cdot (2^{k-1}-2+1)^{\ell_i} = \frac{2^{k-1}-1}{\ell_i},\end{aligned}$$

proving (13).

Next we prove (14). Let $p = \frac{2^{k-2}-1}{2(2^{k-1}-1)}$. Let $V = S_1 \cap \ldots \cap S_{k-1}$ be the intersection set of size 1. We claim that for every non-empty set q, and singleton v,

$$\Pr[Z_i = 1, Q_i = q, V = v] \begin{cases} \leq p^{|q|-1}(1-p)^{\ell_i - |q|}/\ell_i & \text{when } v \subseteq q, \\ = 0 & \text{otherwise.} \end{cases} \tag{15}$$

When V is not contained in Q_i, the value of Z_i is always 0. On the other hand, when $v \subseteq q$,

$$\begin{aligned}\Pr[Z_i = 1, Q_i = q, V = v] &\leq \Pr[Q_i = q, V = v] \\ &= (1/\ell_i) \cdot \Pr[Q_i = q | V = v] \\ &\leq p^{|q|-1}(1-p)^{\ell_i - |q|}/\ell_i,\end{aligned}$$

since every element $e \notin v$ is included in Q_i with probability p, independent of all other such elements. Indeed such an element is included in S_1 with probability $\frac{2^{k-2}-1}{2^{k-1}-1}$, and given that it is included in S_1, it is excluded from S_1' with probability $1/2$.

When $|Q_i| = 0$, we have that $Z_i = 0$, and so $Z_i/|Q_i| = 0$ by our convention. Thus (15) gives

$$\begin{aligned}\mathop{\mathbb{E}}_{Z,Q_i}\left[\frac{Z_i}{|Q_i|}\right] &\leq \sum_{v:|v|=1} \sum_{q: v \subseteq q} \frac{p^{|q|-1}(1-p)^{\ell_i - |q|}/\ell_i}{|q|} \\ &= \frac{1}{p\ell_i} \sum_{q \neq \emptyset} p^{|q|}(1-p)^{\ell_i - |q|} \\ &\leq \frac{1}{p\ell_i} \cdot (1-p+p)^{\ell_i} = \frac{2(2^{k-1}-1)}{\ell_i(2^{k-2}-1)},\end{aligned}$$

proving (14).