Number-on-Forehead Complexity

Dan Mitropolsky

April 27, 2022

Multiparty Communication Complexity

How do we define communication complexity for k parties?

Given $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k \to \mathcal{Z}$

Definition 1 (*Number-in-hand model, "NOH"*) Player i sees input $x_i \in \mathcal{X}_i$ only.

Definition 2 (*Number-on-forehead model, "NOF"*) Player i sees every input $x_j \in \mathcal{X}_j$ for $j \neq i$.

Notation

- Often by f we mean a function family $f_{n,k} : (\{0,1\}^n)^k \to \{0,1\}$.
- Write $D_k(f)$ for *deterministic* communication complexity of $f_{n,k}$
- For distribution μ over $(\{0,1\}^n)^k$ and $\epsilon > 0$, write $R_k^{\epsilon,\mu}(f)$ for communication complexity of $f_{n,k}$ where inputs drawn μ, and the (deterministic) protocol can err on at most ϵ fraction of inputs.

Motivating Example: EQ

- Consider $EQ_k : (\{0,1\}^n)^k \to \{0,1\}$.
- For $k = 2$, NIH $=$ NOF model.
- For $k = 2$, $D_2(EQ) = n$ (maximal).
- But for $k > 3$, in NOF model, $D_k(EQ) = 2$.
- In NIH mode, CC is $\Omega(n)$.
- in NOF, we can *exploit overlap of information* for efficiency!

Motivating Example: EQ

Solution: Player 1 sends 1 iff other players' inputs equal. Player 2 sends 1 iff other players' input equal. $EQ(x_1, \ldots, x_k) = 1 \iff$ both bits are 1.

Lay of the land

Lower Bounds

Very little is known!!!

"one good method" gives $n/4^k$-type bounds

- *Generalized Inner Product*: $D_k(\mathrm{GIP}) \geq \Omega(\frac{n}{4^k})$
- *Disjointness*: $D_k(\mathrm{DISJ}) \geq \Omega(\frac{n}{4^k})$
- *Exactly-n, $k = 3$*: $D_3(\textsc{Exactly-}n) \geq \Omega(\log \log \log n)$

Upper Bounds

We know a few surprising efficient protocols!

- *Generalized Inner Product*: $D_k(\mathrm{GIP}) \leq O(k\frac{n}{2^k})$
- *Exactly-n, $k = 3$*: $D_3(\textsc{Exactly-}n) \leq \sqrt{\log n}$.

Connection to ACC^0

Another reason to care about NOF!

Definition $\mathrm{AC}^0[m]$ is the class of languages that can be computed by a family of circuits $\{C_n\}$ such that each $C_n : \{0,1\}^n \to \{0,1\}$ is: constant depth, size poly(n), gates are $\{\wedge, \vee, \neg, \mod m\}$ with unbounded fan-in.

Definition $\mathrm{ACC}^0 = \bigcup_{m \geq 2} \mathrm{ACC}^0[m]$

Theorem (Beigel and Tarui '94) For $L \in \mathrm{ACC}^0$, $\exists c, d$ s.t. L be computed by depth 2 circuits, size $2^{\log^d n}$, top gate is *symmetric*, and bottom layer consists of $\wedge$ gates with fan-in $\log^c n$.

Definition The output of a *symmetric* gate is determined by the *number* of 0 and 1 inputs.

Connection to ACC^0

Theorem 1 (Beigel and Tarui '94) For $L \in ACC^0$, $\exists c, d$ s.t. L be computed by depth 2 circuits, size $2^{\log^d n}$, top gate is *symmetric*, and bottom layer consists of $\wedge$ gates with fan-in $\log^c n$.

Theorem 2 (Hådstad and Goldmann '91) Suppose $f : \{0,1\}^n \to \{0,1\}$ can be computed by circuits with: depth 2, top gate is symmetric with fan-in s, bottom layer consists of $\wedge$ gates with fan-in $\leq k-1$. Then (under *any* partition of n into k parties), $D_k(f) \leq k \log(s)$.

Proof Each $\wedge$ gate can be computed by some party. Partition gates among parties, each sends how many are 1.

Corollary (Theorems 1+2) For any function f in ACC^0, c, d s.t. under any partition of n bits to $k = \log^c n + 1$ parties, NOF $D_k(f) \leq (\log^c n + 1) \log^d n = \log^{O(1)} n$.

Connection to ACC^0

Corollary (Theorems 1+2) For any function f in ACC^0, c, d s.t. under any partition of n bits to $k = \log^c n + 1$ parties, NOF $D_k(f) \leq (\log^c n + 1) \log^d n = \log^{O(1)} n$.

- **Usefulness of NOF:** If we could show some f such that for any $k = \log^c n + 1$, it requires $D_k(f) > \log^{O(1)} n$, this would show $f \notin \mathsf{ACC}^0$!!!
- No such lower bounds, yet...
- **major** goal in circuit complexity
- we know $\mathsf{NEXP} \subsetneq \mathsf{ACC}^0$ (but not with this method– Ryan Williams, 2011)

Limited lower bounds

- If we could show some $f_{n,k}$ such that for any $k = \log^c n$, it requires $D_k(f) > \log^{O(1)} n$, this would show $f \notin ACC^0$!!!

Lower Bounds	Upper Bounds
• *Generalized Inner Product*: $D_k(\mathrm{GIP}) \geq \Omega(\frac{n}{4^k})$	• *Generalized Inner Product*: $D_k(\mathrm{GIP}) \leq O(k\frac{n}{2^k})$

- Lower bound is non-trivial only when $k < \log n$

Generalized Inner Product

Definition For $x_1, \ldots, x_k \in (\{0,1\}^n)^k$,
$\mathrm{GIP}_{n,k}(x_1, \ldots, x_k) = \bigoplus_{i=1}^{n} (x_1)_i \wedge \cdots \wedge (x_n)_i$

That is, $\mathrm{GIP}_{n,k}(x_1, \ldots, x_k) =$ number of coordinates that all equal 1, mod 2.

Proposition Viewing $\mathrm{GIP}_{n,k}$ for $k = \log^c n$ and vectors of size $n/(\log^c n)$ as a function on n bits, $\mathrm{GIP}_{n,k} \in \mathrm{ACC}^0$. In fact, $\mathrm{GIP}_{n,k} \in \mathrm{AC}^0[2]$.
Proof bottom layer has $n/(\log^c n)$ AND-gates, computing $(x_1)_i \wedge \cdots \wedge (x_k)_i$ for each coordinate; top layer is a mod 2 gate $\square$
Circuit in proof already in Beigel-Tarui form :)

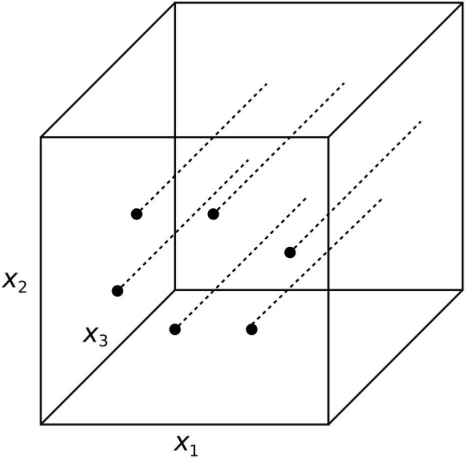
Cylinders

Definition. A *cylinder* C_i in the i-th coordinate is a subset of the input space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$ that does not depend on the i-th coordinate: if $(x_1, \ldots, x_i, \ldots, x_k) \in C_i$ then for all $x_i' \in \mathcal{X}_i, (x_1, \ldots, x_i', \ldots, x_k) \in C_i$.

Definition. A *cylinder intersection* C is an intersection of cylinders.

If C_i, C_i' are cylinders in the i-th coordinate, so is $C_i \cap C_i'$. So any cylinder intersection C can be written $\cap_{i=1}^k C_i$.

Cylinders



Cylinders and Protocols

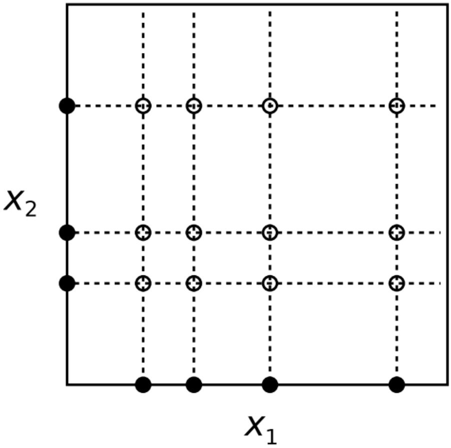
Proposition. For a NOF protocol P with communication c, the set of inputs that induce communication transcript $t \in \{0,1\}^c$ is a cylinder intersection.

Proof sketch. Bit by bit. At step i when player j speaks, whether they write bit c_i depends only on the inputs of every *other* player.

Corollary. If P is a deterministic NOF protocol computing $f : \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k \to \mathcal{Z}$ with c bits of communication, P partitions $\mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_k$ into at most 2^c monochromatic cylinder intersections.

- Cylinder intersections are the analogue of *rectangles*.
- For $k = 2$, cylinder intersection $=$ rectangle.
- Cylinder intersections are complex combinatorial objects. Limited understanding of cylinder intersections $=$ limited NOF bounds.

Cylinder intersection for $k = 2$



Discrepancy

In this section, $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_k \to \pm 1$, and by abuse of notation, $C(x_1, \ldots, x_k) = 1$ if $x_1, \ldots, x_k \in C$, else 0.

Definition. For distribution μ over $\mathcal{X}_1 \times \cdots \times \mathcal{X}_k$, function f, cylinder intersection C, the *discrepancy* of f w.r.t μ and C:

$$\mathrm{disc}_\mu(f, C) = |\mathbb{E}_{x_1,\ldots,x_k \sim \mu}[f(x_1, \ldots, x_k) C(x_1, \ldots, x_k)]|$$

Definition. The *discrepancy* of f wrt μ is

$$\mathrm{disc}_\mu(f) = \max_C \mathrm{disc}_\mu(f, C)$$

Intuition: "average" of f over cylinders. Close to 0 means "well-spread" over ± 1.

Discrepancy Method

Definition. The *discrepancy* of f wrt μ is $\text{disc}_\mu(f) = \max_C \text{disc}_\mu(f, C)$

Theorem (Discrepancy Method; Babai, Nisan, Szegedy '92)
For any f

$$R_k^{\epsilon,\mu} \geq \log\left(\frac{1-2\epsilon}{\text{disc}_\mu(F)}\right)$$

Proof identical to $k = 2$ case (we did it in class!)

Intuition: *upper bound* on discrepancy: for any cylinder intersection, f is "well spread" over ± 1, hard to partition monochromatically. Gives *lower bound* on NOF.

Lower bound for GIP

Theorem (Discrepancy Method) For any f

$$R_k^{\epsilon,\mu} \geq \log\left(\frac{1-2\epsilon}{\mathrm{disc}_\mu(F)}\right)$$

Theorem $\mathrm{disc}_U(\mathrm{GIP}) \;\leq\; \exp(-n/4^k)$

Theorem (GIP lower bound)

$$R_k^{\epsilon,U}(\mathrm{GIP}) \geq n/4^k + \log(1-2\epsilon)$$

And in particular,

$$D_k(\mathrm{GIP}) \geq n/4^k$$

Overview of Two Theorems

Theorem 1 (Goal) $\mathrm{disc}_U(\mathrm{GIP}) \leq \exp(-n/4^k)$

For **two inputs** to f, that is $(x_1^0, \ldots, x_k^0)$ and $(x_1^1, \ldots, x_k^1)$, for a vector $b \in \{0,1\}^k$, x^b denotes the mixed input $(x_1^{b_1}, \ldots, x_k^{b_k})$.

Theorem 2 (Cube-measure bound for discrepancy)
For *any* f,

$$\mathrm{disc}_U(f)^{2^k} \leq \mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\left[\prod_{b\in\{0,1\}^k} f(x^b)\right]$$

Theorem 3 (Cube-measure of GIP)

$$\mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\left[\prod_{b\in\{0,1\}^k} \mathrm{GIP}(x^b)\right] \leq e^{-n/2^{k-1}}$$

$\mathrm{disc}_U(\mathrm{GIP}) \leq (e^{-n/2^{k-1}})^{1/2^k} \leq 2^{-n/4^k}$ to get Theorem 1.

Cube-measure of GIP (Theorem 3)

Theorem 3 (Cube-measure of GIP)

$$\mathop{\mathbb{E}}_{\substack{(x_1^0,\dots,x_k^0)\\(x_1^1,\dots,x_k^1)}}\Big[\prod_{b\in\{0,1\}^k} \mathrm{GIP}(x^b)\Big] \le e^{-n/2^{k-1}}$$

Proof:

$$= \mathbb{E}\Big[\prod_{b\in\{0,1\}^k}\prod_{i=1}^{n}(-1)^{x_{1,i}^{b_1}\wedge\cdots\wedge x_{k,i}^{b_k} i}\Big]$$

$$= \mathbb{E}\Big[\prod_{i=1}^{n}\prod_{b\in\{0,1\}^k}(-1)^{x_{1,i}^{b_1}\wedge\cdots\wedge x_{k,i}^{b_k}}\Big]$$

Because the inputs are uniform, the coordinates are *independent*, hence

$$= \prod_{i=1}^{n}\mathbb{E}\Big[\prod_{b\in\{0,1\}^k}(-1)^{x_{1,i}^{b_1}\wedge\cdots\wedge x_{k,i}^{b_k}}\Big]$$

Cube-measure of GIP (Theorem 3)

Theorem 3 (Cube-measure of GIP)

$$\mathbb{E}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\Big[\prod_{b\in\{0,1\}^k} \mathrm{GIP}(x^b)\Big] \leq e^{-n/2^{k-1}}$$

$$= \left(\mathbb{E}_{\substack{x_1^0,\ldots,x_k^0\in\{0,1\}\\x_1^1,\ldots x_k^1\in\{0,1\}}}\Big[\prod_{b\in\{0,1\}^k} (-1)^{x_0^{b_1}\wedge\cdots\wedge x_k^{b_k}}\Big]\right)^n$$

- if for all $j \in [k]$, $x_j^0 \neq x_j^1$ then the product is -1
- prob of above $= 1/2^k$
- if for some j, $x_j^0 = x_j^1$, then product is 1

$$= \left((1 - 1/2^k) - 1/2^k\right)^n$$

$$= (1 - 1/2^{k-1})^n$$

$$=\leq e^{-n/2^{k-1}}$$

Bounding discrepancy

Theorem 2

For *any* f,

$$\mathrm{disc}_U(f)^{2^k} \leq \mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\left[\prod_{b\in\{0,1\}^k} f(x^b)\right]$$

Recall: $\mathrm{disc}_U(f) = \max_C \mathrm{disc}_U(f, C) = \max_C |\mathbb{E}_{x_1,\ldots,x_k}[f(x_1,\ldots,x_k)C(x_1,\ldots,x_k)]|$

- *the* main technique (and limitation) for NOF lower-bounds
- uses repeated Cauchy-Schwarz to get rid of cylinder intersections, replacing them with product over double-expectation

Cauchy-Schwarz Lemma: $\mathbb{E}[Z]^2 \leq \mathbb{E}[Z^2]$.

Bounding Discrepancy

Proof is by induction: assume true for any function on $k-1$ players. For f on k players, for the maximizing cylinder $C = \cap_{i=1}^{k} C_i$,

$$\mathrm{disc}_U(f) = \Big| \mathop{\mathbb{E}}_{x_1,\dots,x_k} [f(x_1,\dots,x_k)\Pi_{i=1}^{k} C_i(x_1,\dots,x_k)] \Big|$$

Since C_k does not depend on x_k,

$$= \Big| \mathop{\mathbb{E}}_{x_1,\dots,x_{k-1}} [C_k(x_1,\dots,x_{k-1},\cdot) \mathop{\mathbb{E}}_{x_k}[f(x_1,\dots,x_k)\Pi_{i=1}^{k-1} C_i(x_1,\dots,x_k)]]$$

By Cauchy-Schwarz, and $C_k(\cdots) \leq 1$

$$\mathrm{disc}_U(f)^2 \leq \mathop{\mathbb{E}}_{x_1,\dots,x_{k-1}} [(\mathbb{E}_{x_k}[f(x_1,\dots,x_k)\Pi_{i=1}^{k-1} C_i(x_1,\dots,x_k)])^2]$$

Bounding Discrepancy

$$\mathrm{disc}_U(f)^2 \leq \mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1}}[(\mathbb{E}_{x_k}[f(x_1,\ldots,x_k)\Pi_{i=1}^{k-1}C_i(x_1,\ldots,x_k)])^2]$$

$$= \mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1},x_k^0,x_k^1}[f(\ldots,x_k^0)f(\ldots,x_k^1)\Pi_{i=1}^{k-1}C_i(\ldots,x_k^0)C_i(\ldots,x_k^1)]$$

$$= \mathop{\mathbb{E}}_{x_k^0,x_k^1}[\mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1}}[f^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})\Pi_{i=1}^{k-1}C_i^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})]$$

Raise both sides to power of 2^{k-1}. By Cauchy-Schwarz,

$$\mathrm{disc}_U(f)^{2^k} \leq$$

$$\mathop{\mathbb{E}}_{x_k^0,x_k^1}[\Big(\mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1}}[f^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})\Pi C_i^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})\Big)^{2^{k-1}}]$$

Bounding Discrepancy

$$\mathrm{disc}_U(f)^{2^k} \leq$$

$$\mathop{\mathbb{E}}_{x_k^0, x_k^1}\Big[\Big(\mathop{\mathbb{E}}_{x_1,\ldots,x_{k-1}}[f^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})\Pi C_i^{x_k^0,x_k^1}(x_1,\ldots,x_{k-1})\Big)^{2^{k-1}}\Big]$$

Inner expectation upper bounded by $\mathrm{disc}_U(f^{x_k^0,x_k^1})$. This is a function on $k-1$ parties, so by induction,

$$\leq \mathop{\mathbb{E}}_{x_k^0,x_k^1}\ \mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_{k-1}^0)\\(x_1^1,\ldots,x_{k-1}^1)}}\Big[\prod_{b\in\{0,1\}^{k-1}} f^{x_k^0,x_k^1}(x^b)\Big]$$

$$= \mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\Big[\prod_{b\in\{0,1\}^{k-1}} f(x^b,x_k^0)f(x^b,x_k^1)\Big]$$

$$= \mathop{\mathbb{E}}_{\substack{(x_1^0,\ldots,x_k^0)\\(x_1^1,\ldots,x_k^1)}}\Big[\mathop{f}_{b\in\{0,1\}^k}(x^b)\Big] \ \square$$

Upper bounds: Exactly-n

Definition Exactly-n is a 3-party function $f : [n]^3 \to \{0,1\}$ where $f(x,y,z) = 1$ iff $x + y + z = n$.

- Remember, this is NOF: Alice sees y, z, Bob sees x, y, Charlie sees x, y.
- Trivial $\log n + 1$ protocol where Alice sends y

Main theorem: $D_3(f) \leq \sqrt{\log n}$

Colorings

Main theorem: $D_3(f) \leq \sqrt{\log n}$

Definition. A *coloring* is a mapping from $[n]$ to a color set C. It is "3-AP-free" if for any sequence $a, a+b, a+2b \in [n]$, they do not have the same color.

Examples– 3-AP free?

1 2 3 4 5 6

1 2 3 4 5 6

Theorem (Behrend 1946) There is a 3-AP-free coloring of $[n]$ with $2^{O(\sqrt{logn})}$ colors.

Proof (main theorem)

Theorem (Behrend 1946) There is a 3-AP-free coloring of $[n]$ with $2^{O(\sqrt{\log n})}$ colors.

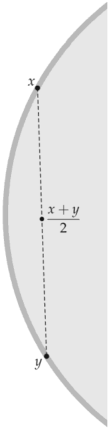
Proof of main theorem.

- Let $x' = n - y - z$, $y' = n - x - z$.
- Observe: $x - x' = y - y' = x + y + z - n$.
- $x + 2y', x' + 2y, x + 2y$ is a 3-AP (with jump $x + y + z - n$)
- They are all equal iff $x + y + z = n$.
- All three numbers in $[-2n, 2n]$ and can be computed by Bob, Alice, and Charlie, respectively.
- Using the coloring for $[4n]$, send colors and check if same: $D_3(f) \leq \log(2^{O(\sqrt{\log 4n})}) = O(\sqrt{\log n})$

Behrend's theorem

Theorem (Behrend 1946) There is a 3-AP-free coloring of $[n]$ with $2^{O(\sqrt{\log n})}$ colors.

Intuition: a 3-AP is sequence $x, \frac{x+y}{2}, y \in [n]$. Suppose we had homomorphism from $[n]$ to $\mathbb{R}^d$, and color by vector length.



Behrend's theorem

Theorem (Behrend 1946) There is a 3-AP-free coloring of $[n]$ with $2^{O(\sqrt{\log n})}$ colors.

Proof.

- Choose d, r such that $4|d$ and $d^r > n$. Let $v(x) \in \mathbb{R}^r$ be the base-d representation of x.
- If $\|v(x)\|^2$ coloring worked, $d^2 r = O(d^2 \log(n))$ colors would suffice. Unfortunately, doesn't work...
- Even if $\|v(x)\|^2 = \|v(y)\|^2$, not necessarily true that $v(\frac{x+y}{2}) = \frac{v(x)+v(y)}{2}$
- Idea: add "extra info" to coloring to force this homomorphic property.

Behrend's Coloring

Theorem (Behrend 1946) There is a 3-AP-free coloring of $[n]$ with $2^{O(\sqrt{log n})}$ colors.

- $v(x) =$ base-d representation of x.
- Let $w(x) \in \mathbb{R}^d$ be the *approximation* of $v(x)$: $w(x)_i$ is largest number $jd/4$ for $j \in \{0, 1, 2, 3, 4\}$ such that $jd/4 \leq x_i$.
- Color v by $(v(x), w(x))$
- At most $5^r = 2^{O(r)}$ values for $w(x)$, and d^2r for $v(x)$
- overall have $2^{O(r)+\log d}$ colors. Use $r = \sqrt{\log n}, d = 2^{\sqrt{\log n}}$ to get $2^{O(\sqrt{\log n})}$.

Behrend's theorem, ctd.

- Suppose $a, a+b, a+2b \in [n]$ have same color.
- $\|v(a)\| = \|v(a+b)\| = \|v(a+2b)\|$.
- **Will show** that w's are the same implies $v(a+b) = \frac{v(a)+v(a+2b)}{2}$, contradiction with line above!!
- Let $W(x)$ be the number represented by $w(x)$ (that is, $\sum_{i=0}^{r} w(x)_i d^i$.
- The base-d representation of $x - W(x)$ is $v(x) - w(x)$.
- $W(a) = W(a+b) = W(a+2b)$

$$
\begin{aligned}
a + 2b + a &= 2(a+b) \\
a + 2b - W(a+2b) + a - W(a) &= 2(a + b - W(a+b)) \\
v(a+2b) - w(a+2b) + v(a) - w(a) &= 2(v(a+b) - w(a+b)) \\
v(a+2b) + v(a) &= 2v(a+b)
\end{aligned}
$$

References I

Anil Ada.
Notes on communication complexity.
Excellent notes,
https://www.Fcs.mcgill.ca/~rraada/CCnotes.pdf.

László Babai, Noam Nisan, and Mario Szegedy.
Multiparty protocols and logspace-hard pseudorandom sequences (extended abstract).
In *STOC 1989*, 1989.

Troy Lee.
Lecture 2: Multiparty number-on-the-forehead complexity, 2012.
Excellent notes,
https://www.csc.kth.se/utbildning/kth/kurser/DD2441/semteo12/lecturenotes/NotesLec13.pdf.

References II

Shachar Lovett.
Cse 291: Communication complexity, winter 2019, multi-party protocols, 2019.
Excellent notes, https://cseweb.ucsd.edu/classes/wi19/cse291-b/5-multiparty.pdf.

Toniann Pitassi.
Foundations of communication complexity, lecture 5, 2014.

First part covers connection to ACC, https://www.cs.toronto.edu/~toni/Courses/PvsNP/Lectures/lecture5.pdf.

References III

Anup Rao and Amir Yehudayoff.
Communication Complexity: and Applications.
Cambridge University Press, 2020.
Chapters 4 and 5 in `https://yehudayoff.net.technion.ac.il/files/2016/03/book.pdf`.