# Lecture 2: Multiparty number-on-the-forehead complexity

Troy Lee

Oct. 4, 2012

## 1 Basic Definition

Now we get into the exotic models. Say that there are $k$ players who want to compute a function $f : (\{-1,+1\}^n)^k \to \{-1,+1\}$. To make things interesting, on input $(x_1, \dots, x_k)$ player $i$ is given $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$. That is, player $i$ knows everything *except* $x_i$, which is figuratively written on his forehead. How much communication is needed to compute $f$ correctly (or with high probability) on every input in this case? Usually it is assumed that messages are written "on the blackboard", that is that every player sees every message. The cost is the total number of bits written on the blackboard for the worst case input.

The overlap in information between the players is what makes showing lower bounds in this model very challenging. We want to show lower bounds not just because it is difficult but because there are a wealth of applications. Let's see one of these first to motivate the model.

## 2 Application to circuit complexity

One of the principal motivations for studying multiparty number-on-the-forehead communication complexity is that lower bounds in this model imply circuit complexity lower bounds. A key observation in this connection is due to Håstad and Goldmann.

**Lemma 1** (Håstad and Goldmann [HG91]). *Suppose that a function $f$ can be computed by a depth-2 circuit whose top gate is an arbitrary symmetric function of fan-in $s$ and whose bottom gates compute arbitrary functions of fan-in at most $k-1$. Then, under any partition of the input variables, the $k$-party number-on-the-forehead complexity of $f$ is at most $k\log(s)$. Furthermore, this can be achieved by a simultaneous protocol.*

*Proof.* As each bottom gate has fan-in at most $k-1$, under any partition of the input variables, there is some player who sees the entire input to this gate. By a scheme arranged beforehand, the players partition these gates among themselves so that each gate is computed by some player. Each player then announces the number of gates assigned to him which evaluate to true. This takes $\log s$ bits of communication. Once the players know the total number of bottom gates which evaluate to true, they can compute $f$. The total communication is $k \log(s)$. $\square$

Functions that can be computed by quasipolynomial size depth-2 circuits with a symmetric top gate and bottom gates of polylogarithmic size fan-in is a surprisingly rich class. Indeed, Allender [All89] shows that this class can compute all of $\mathrm{AC}^0$. Further work by Yao [Yao90] shows that the probabilistic version of this class can compute all of $\mathrm{ACC}^0$ and Beigel and Tarui [BT94] improve this to a deterministic simulation. We record this statement for reference.

**Theorem 2** (Beigel and Tarui)**.** *Any language in* $\mathrm{ACC}^0$ *can be computed by a depth-2 circuit of size* $2^{\log^{O(1)}(n)}$ *with a symmetric gate at the top and AND gates of fan-in* $\log^{O(1)} n$ *at the bottom.*

As a consequence, showing that a function $f$ requires super-polylogarithmic communication for super-polylogarithmic many players in the simultaneous number-on-the-forehead model will show that $f$ is not in $\mathrm{ACC}^0$. A recent major result showed that NEXP is not contained in $\mathrm{ACC}^0$ [Wil11]. This was not shown via NOF communication complexity.

# 3 Grolmusz Protocol

The best lower bounds for explicit functions we have are of the form $n/2^k$. For example for the Generalized Inner Product Function

$$\mathrm{GIP}_k(x_1, \ldots, x_k) = \bigoplus (x_1 \wedge \cdots \wedge x_k)$$

we have a lower bound of $\Omega(n/2^{2k})$ as shown in the seminal paper of Babai, Nisan, and Szegedy [BNS92]. In fact this lower bound is nearly tight, as shown by a very cool protocol of Grolmusz [Gro94].

Let $f : \{0,1\}^n \to \{0,1\}$ be a symmetric function. In other words, $f(x) = f(y)$ whenever $x$ and $y$ have the same number of ones. Let $x_1, \ldots, x_k \in \{0,1\}^n$ be the inputs to the $k$-players. We will think of the input as being described by a $k$-by-$n$ matrix $X$ whose $i^{th}$ row is $x_i$. The key step in the protocol of Grolmusz is the following lemma:

**Lemma 3** (Grolmusz [Gro94])**.** *Let $f$ be a symmetric function. Suppose the players know that some string $r \in \{0,1\}^k$ does not appear in the input matrix $X$. Then they can evaluate $f(x_1 \wedge \ldots \wedge x_k)$ with $k \log n$ bits of communication.*

*Proof.* Notice that if the players can count the number of all-one columns in $X$ then they can compute $f$. By rearranging the rows of $X$ as necessary, we may assume that the missing column $r$ is of the form $0^\ell 1^{k-\ell}$ where $\ell \in \{1, \ldots, k\}$. If $\ell = 0$ then the all-one column does not appear, and the players can immediately evaluate $f$.

More generally, let $e_i = 0^i 1^{k-i}$ for $i \in \{0, \dots, k\}$. Thus the players want to count the number of times $e_0$ appears as a column of $X$. Let $E_i$ be the number of times the string $e_i$ appears as a column of $X$.

Although the first player cannot distinguish between a column of the form $e_0$ or $e_1$ as he does not see the first bit, he can exactly compute $E_0 + E_1$. Player 1 announces this number with $\log n$ bits. Similarly, player 2 announces $E_1 + E_2$. The players continue this way until they reach player $\ell$, who will announce $E_{\ell-1}$ as by assumption $e_\ell$ does not appear as a column of $X$. With this knowledge, the players can then solve for $E_0$ and evaluate $f$. $\square$

**Theorem 4** (Grolmusz [Gro94]). *Let $f$ be a symmetric function. Then*

$$D_k(f(x_1 \wedge \dots \wedge x_k)) \leq k(\log(n) + 1) \left\lceil \frac{n}{2^{k-1} - 1} \right\rceil.$$

*Proof.* The first player will play a special role in the protocol. He (mentally) partitions the input matrix $X$ into $\left\lceil \frac{n}{2^{k-1}-1} \right\rceil$ many blocks of columns of size $2^{k-1} - 1$. By counting, in each of these blocks of columns, there is some $k - 1$ bit string which does not appear, and can be identified by the first player. The first player announces these strings, and then the players perform the protocol given in the lemma. The total communication is for each block is $k(\log(n) + 1$ and overall is

$$k(\log(n) + 1) \left\lceil \frac{n}{2^{k-1} - 1} \right\rceil.$$

$\square$

In the special case where $f$ is the parity function, this communication can be reduced to

$$k \left\lceil \frac{n}{2^{k-1} - 1} \right\rceil$$

as in the lemma the players do not need to say $E_i + E_{i+1}$ but just the parity of this number.

Notice that the protocol of Grolmusz is nearly simultaneous, but not quite as the first player must announce the missing columns to the other players. Babai et al. [BGKL03] have shown that any function $f(x_1 \wedge \dots \wedge x_k)$ for symmetric $f$ indeed has a simultaneous protocol with $O(\log^{O(1)}(n))$ bits of communication whenever the number of players is larger than $\log n + 1$.

# 4 Lower Bounds

Yesterday we considered an XOR game $G = (f, \mu)$ and looked at the maximum bias achievable by Alice and Bob for $G$. We denoted this as

$$\beta_{(f,\mu)} = \max_{\substack{a \in \{-1,+1\}^{|X|} \\ b \in \{-1,+1\}^{|Y|}}} \sum_{x,y} f(x, y)\mu(x, y)a(x)b(y).$$

Later on, we argued that it was more natural to consider the object $Q(x, y) = f(x, y)\mu(x, y)$ which can be an arbitrary function $Q : X \times Y \to \mathbb{R}$ with the property that $\ell_1(Q) = 1$. (Here $\ell_1(Q) = \sum_{x,y} |Q(x, y)|$.) It will be useful to extend our notation for this. Define

$$\beta(Q) = \max_{\substack{a \in \{-1,+1\}^{|X|} \\ b \in \{-1,+1\}^{|Y|}}} Q(x, y)a(x)b(y).$$

The same argument that we gave yesterday works in the multiplayer number-on-the-forehead model as well. In other words,

$$2^{R_\epsilon(f)} \geq \max_{\substack{Q \\ \ell_1(Q)=1}} \frac{\langle f, Q\rangle - 2\epsilon}{\beta(Q)},$$

where now $\beta(Q)$ looks at the maximal correlation of $Q$ with a strategy in a number-on-the-forehead XOR game. Here the strategy of player $i$ is a function $C_i(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_k) \in \{-1, +1\}$. This is known as a *cylinder* and the product of cylinders is known as a cylinder intersection. Cylinder intersections in the number-on-the-forehead world are the analog of rectangles in normal 2-party communication complexity.

It gets pretty unwieldy so let's just look at the correlation of $Q$ with a cylinder intersection for 3-players.

$$\beta(Q) = \max_{C_1,C_2,C_3} \sum_{x_1,x_2,x_3} Q(x_1, x_2, x_3)C_1(x_2, x_3)C_2(x_1, x_3)C_3(x_1, x_2)$$

Now we don't know how to turn to matrix (or tensor) theory to help us bound this thing like we did yesterday. Babai, Nisan, and Szegedy [BNS92] came up with a way of upper bounding this correlation, and 20 years later it is basically still the only thing we know how to do. It's called apply Cauchy-Schwarz! We will use Cauchy-Schwarz in the form that $\mathbb{E}[ab] \leq \sqrt{\mathbb{E}[a^2]}\sqrt{\mathbb{E}[b^2]}$.

Fix $C_1, C_2, C_3$ which realize the maximum in $\beta(Q)$.

$$\begin{aligned}
\frac{\beta(\Psi)}{\text{size}(\Psi)} &= \mathbb{E}_{x_1,x_2,x_3} \Psi(x_1, x_2, x_3)C_1(x_2, x_3)C_2(x_1, x_3)C_3(x_1, x_2) \\
&= \mathbb{E}_{x_1,x_2} C_3(x_1, x_2)\mathbb{E}_{x_3} \Psi(x_1, x_2, x_3)C_1(x_2, x_3)C_2(x_1, x_3) \\
&\leq \sqrt{\mathbb{E}_{x_1,x_2} C_3(x_1, x_2)^2} \sqrt{\mathbb{E}_{x_1,x_2,x_3^0,x_3^1} \prod_{\ell \in \{0,1\}} \Psi(x_1, x_2, x_3^\ell)C_1(x_2, x_3^\ell)C_2(x_1, x_3^\ell)}
\end{aligned}$$

Note that the first term is at most one, so we forget about it. Now we move on to round 2 of Cauchy-Schwarz!

$$\begin{aligned}\left(\frac{\beta(\Psi)}{\text{size}(\Psi)}\right)^2 &\le \mathbb{E}_{x_1,x_2,x_3^0,x_3^1}\prod_{\ell\in\{0,1\}}\Psi(x_1,x_2,x_3^\ell)C_1(x_2,x_3^\ell)C_2(x_1,x_3^\ell)\\ &= \mathbb{E}_{x_1,x_3^0,x_3^1}C_2(x_1,x_3^0)C_2(x_1,x_3^1)\mathbb{E}_{x_2}\prod_{\ell\in\{0,1\}}\Psi(x_1,x_2,x_3^\ell)C_1(x_2,x_3^\ell)\\ &\le \left(\mathbb{E}_{x_1,x_2^0,x_2^1,x_3^0,x_3^1}\prod_{\ell\in\{0,1\}^2}\Psi(x_1,x_2^{\ell_1},x_3^{\ell_2})C_1(x_2^{\ell_1},x_3^{\ell_2})\right)^{1/2}\end{aligned}$$

Finally we get our bound!

$$\left(\frac{\beta(\Psi)}{\text{size}(\Psi)}\right)^4 \le \mathbb{E}_{x_2^0,x_2^1,x_3^0,x_3^1}\left|\mathbb{E}_{x_1}\prod_{\ell\in\{0,1\}^2}\Psi(x_1,x_2^{\ell_1},x_3^{\ell_2})\right| \tag{1}$$

In the general $k$-player case we will have to do $k-1$ rounds of Cauchy-Schwarz, and so the left hand side will be raised to the power $2^{k-1}$. This is why all the bounds we have decay like $1/2^k$.

## 4.1 Pattern Tensors

We will again study the communication complexity of composed functions. The outer function $f:\{-1,+1\}^n \to \{-1,+1\}$ can be arbitrary, but the inner function $g$ will be of a very particular type. The inner function will take $k$ arguments $g:\{-1,+1\}^{N^{k-1}}\times[N]\times\cdots\times[N]$. The first argument can be thought of as a $k-1$ dimensional tensor, and the other $k-1$ arguments as indices into the sides of this tensor. We define

$$g(x,y_1,\ldots,y_{k-1}) = x[y_1,\ldots,y_{k-1}].$$

**Definition 5** (Pattern Tensor)**.** *Fix an integer $N$. As a whole, the pattern tensor $A_{f,N}$ is then defined such that $A_{f,N}[x,y_1,\ldots,y_{k-1}]$ is equal to*

$$f(g(x^1,y_1^1,\ldots,y_{k-1}^1),\ldots,g(x^n,y_1^n,\ldots,y_{k-1}^n)).$$

*Here $x=(x^1,\ldots,x^n)$ is a $k$-dimensional sign tensor with dimensions $n\times N\times\cdots\times N$, $x^i$ is the $k-1$ dimensional tensor achieved by constraining the first index of $x$ to be equal to $i$. The $y_j$'s are vectors of indices in $[N]^n$, and $y_j^i$ is the $i^{th}$ element of the $j^{th}$ vector.*

Here is the main theorem we will talk about today.

**Theorem 6** ([LS09, CA08])**.** *Let $f:\{-1,+1\}^n\to\{-1,+1\}$ be a boolean function, then*

$$R_\epsilon(A_{f,N}) \ge \frac{\deg_{3\epsilon}(f)}{2^{k-1}} - O(1),$$

*provided $N \ge \frac{2e(k-1)2^{2^{k-1}}n}{\deg_{3\epsilon}(f)}$.*

This theorem has the following application to Disjointness.

**Theorem 7** ([LS09, CA08]). *For every $k$ and $n$*

$$R_{1/4}(\mathrm{DISJ}_{k,n}) \geq \Omega\left(\frac{n^{1/(k+1)}}{2^{2^k}}\right).$$

*Proof.* By Theorem 6 $R_{1/4}(A_{\mathrm{OR}_n,N}) \geq \sqrt{n}/2^k$ provided $N > 2e(k-1)2^{2^{k-1}}\sqrt{n}$. If we can solve $\mathrm{DISJ}_{k,m}$ on instances of size $m = nN^{k-1}$ then we can solve $A_{\mathrm{OR}_n,N}$. Solving shows that $\sqrt{n} = \Omega(m^{1/(k+1)}/2^{2^k})$. □

The proof of Theorem 6 is similar to what we saw yesterday. Let $\psi$ be a witness to the fact that $f$ cannot be approximated to within less than $\epsilon$ by degree $d$ polynomials. That is, let $\psi$ satisfy the following properties:

1. $\langle \psi, f\rangle \geq \epsilon$
2. $\ell_1(\psi) = 1$
3. $\langle \chi_T, \psi\rangle = 0$ for all characters $|T| \leq d$

Our witness $Q$ to the hardness of the pattern tensor $A_{f,N}$ will be the pattern tensor of $\psi$, that is $Q = A_{\psi,N}$.

Let us check what $\langle A_{f,N}, Q\rangle$ is.

$$\begin{aligned}
\langle A_{f,N}, Q\rangle &= \sum_{x_1,y_1,\ldots,y_{k-1}} A_{f,N}(x_1,y_1,\ldots,y_{k-1})Q(x_1,y_1,\ldots,y_{k-1}) \\
&= \sum_{z\in\{-1,+1\}^n} f(z)\psi(z)\prod_{i=1}^{n}\sum_{x_1^i,y_1^i,\ldots,y_{k-1}^i}[x_1^i(y_1^i,\ldots,y_{k-1}^i) = z_i] \\
&= \frac{\mathrm{size}(Q)}{2^n}\sum_{z\in\{-1,+1\}^n} f(z)\psi(z) \geq \epsilon\frac{\mathrm{size}(Q)}{2^n}
\end{aligned}$$

A similar calculation shows that $\ell_1(Q) = \frac{\mathrm{size}(Q)}{2^n}$. Now we turn to the interesting part—the upper bound on $\beta(Q)$.

**Binomial coefficient bound** We will need the following quick upper bound on binomial coefficients in the proof:

$$\binom{n}{t} \leq \left(\frac{en}{t}\right)^t \tag{2}$$

Of course we have a lower bound of $(n/t)^t$ so this is not too bad. Here is a simple way to see it.

$$\binom{n}{t} \leq \frac{n^t}{t!}.$$

Now plug in $t! \geq (t/e)^t$. A nice way to see this is that $e^t = \sum_s t^s/s!$. All the terms in the sum are positive so $e^t \geq t^t/t!$.

**Lemma 8** ([Cha07, LS09, CA08]). *Let $v : \{-1, +1\}^n \to \mathbb{R}$ be a function satisfying:*

1. $\|v\|_1 \leq 1$,
2. $\hat{v}_T = 0$ *for every* $T \subset [n]$ *with cardinality* $|T| \leq d$.

*Take* $Q = Q_{v,N}$ *be the pattern $k$-tensor corresponding to $v$. Then*

$$\mu^*(Q) \leq \frac{\text{size}(Q)}{2^{n+d/2^{k-1}}},$$

*provided that* $N \geq \frac{2e(k-1)2^{2^{k-1}}n}{d}$.

*Proof.* We will just prove the $k = 2$ case. The general case follows in the same way, with a bit more involved combinatorics.

Consider the definition of a pattern tensor. In the two dimensional (matrix) case, the input $x$ is an $n \times N$ sign matrix, and $y$ is a vector in $[N]^n$. As in Equation (1) we have

$$\left(\frac{\beta(Q)}{\text{size}(Q)}\right)^2 \leq \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x Q[x, y^0]Q[x, y^1]\right|. \tag{3}$$

To estimate the inner expectations over $x$, we use the Fourier representation $v = \sum_T \hat{v}_T \chi_T$ of $v$.

We can express $Q$ as a linear combination $Q = \sum_T \hat{v}_T \chi_{T,N}$, where $\chi_{T,N}$ is the pattern matrix corresponding to the character $\chi_T$. Now the right hand side of (3) becomes

$$\mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \sum_{T,T'} \hat{v}_T \hat{v}_{T'} \chi_{T,N}[x, y^0]\chi_{T',N}[x, y^1]\right|.$$

By linearity of expectation and the triangle inequality this is bounded by

$$\sum_{T,T'} |\hat{v}_T \hat{v}_{T'}| \, \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \chi_{T,N}[x, y^0]\chi_{T',N}[x, y^1]\right|.$$

We now use the properties of $v$. First, $\|v\|_1 \leq 1$ and therefore $|\hat{v}_T| \leq \frac{1}{2^n}$. In addition $\hat{v}_T = 0$ for every set $T$ with $|T| \leq d$. We therefore arrive at the following expression

$$\frac{1}{2^{2n}} \sum_{T,T':|T|,|T'|>d} \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \chi_{T,N}[x, y^0]\chi_{T',N}[x, y^1]\right|.$$

This is equal to

$$\frac{1}{2^{2n}} \sum_{T,T':|T|,|T'|>d} \mathbb{E}_{y^0,y^1} \left|\mathbb{E}_x \prod_{i \in T} x[i, y^0[i]] \prod_{j \in T'} x[j, y^1[j]]\right|.$$

The expectation inside the absolute value is equal to 0 if $T \neq T'$, and also if $T = T'$ but there is an element $i \in T$ such that $y^0[i] \neq y^1[i]$. The value of this expectation is 1 in all other cases. Our expression is therefore equal to

$$\frac{1}{2^{2n}} \sum_{T:|T|>d} \Pr_{y^0,y^1} \left[\forall i \in T\ ,\ y^0[i] = y^1[i]\right]$$

For a set $T$ of cardinality $|T| = t$, we have

$$\Pr_{y^0,y^1} \left[\forall i \in T\ ,\ y^0[i] = y^1[i]\right] \leq N^{-t}.$$

Therefore, the sum of these probabilities above can be bounded as follows

$$\begin{aligned} \frac{1}{2^{2n}} \sum_{t=d+1}^{n} \binom{n}{t} N^{-t} &\leq \frac{1}{2^{2n}} \sum_{t=d+1}^{n} \left(\frac{en}{dN}\right)^t \\ &\leq \frac{1}{2^{2n+d}}, \end{aligned}$$

assuming $N \geq 2en/d$. Plugging this back into (3) we get the desired result. □

**Putting everything together** Let's now see how we get Theorem 6. We have established the following properties of our witness $Q$:

1. $\langle A_{f,N}, Q\rangle \geq \epsilon \frac{\text{size}(Q)}{2^n}$
2. $\ell_1(Q) = \frac{\text{size}(Q)}{2^n}$
3. $\beta(Q) \leq \frac{\text{size}(Q)}{2^n} 2^{-d/2^{k-1}}$ provided that $N \geq 2e(k-1)2^{2^{k-1}} n/d$.

Putting these three items into the generalized discrepancy bound gives the result!

# 5 More recent results

The bound on disjointness whose proof we sketched becomes trivial for $\log\log n$ many players. Beame and Huynh showed lower bounds on $k$-party disjointness that are non-trivial up to $(\log n)^{1/3}$ many players [BHN08]. More recently Sherstov has shown a bound of $(n/4^k)^{1/4}$ [She12].

**Open Problem** Show that the 3-party NOF complexity of Disjointness is $\Omega(n)$.

# References


[All89] E. Allender. A note on the power of threshold circuits. In *Proceedings of the 30th IEEE Symposium on Foundations of Computer Science*, pages 580–584, 1989.

[BGKL03] L. Babai, A. Gál, P. Kimmel, and S. Lokam. Simultaneous messages vs. communication. *SIAM Journal on Computing*, 33(1):137–166, 2003.

[BHN08] P. Beame and D. Huynh-Ngoc. Multiparty communication complexity of $\mathrm{AC}^0$. Technical Report TR-08-082, ECCC, 2008.

[BNS89] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols and Logspace-hard pseudorandom sequences. In *Proceedings of the 21st ACM Symposium on the Theory of Computing*, pages 1–11. ACM, 1989.

[BNS92] L. Babai, N. Nisan, and M. Szegedy. Multiparty protocols, pseudorandom generators for Logspace, and time-space trade-offs. *Journal of Computer and System Sciences*, 45:204–232, 1992.

[BT94] R. Beigel and J. Tarui. On ACC. *Computational Complexity*, 4:350–366, 1994.

[CA08] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. Technical Report TR-08-002, ECCC, 2008.

[Cha07] A. Chattopadhyay. Discrepancy and the power of bottom fan-in depth-three circuits. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science*, pages 449–458. IEEE, 2007.

[Gro94] V. Grolmusz. The BNS lower bound for multi-party protocols is nearly optimal. *Information and computation*, 112(1):51–54, 1994.

[HG91] J. Håstad and M. Goldmann. On the power of small-depth threshold circuits. *Computational Complexity*, 1:113–129, 1991.

[LS09] T. Lee and A. Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[She08] A. Sherstov. The pattern matrix method for lower bounds on quantum communication. In *Proceedings of the 40th ACM Symposium on the Theory of Computing*, pages 85–94. ACM, 2008.

[She09] A. Sherstov. Separating $\mathrm{AC}^0$ from depth-2 majority circuits. *SIAM Journal on Computing*, 38(6):2113–2129, 2009.

[She12] A. Sherstov. The multiparty communication complexity of set disjointness. In *Proceedings of the 44th ACM Symposium on the Theory of Computing*. ACM, 2012.

[Wil11] R. Williams. Non-uniform ACC circuit lower bounds. In *Proceedings of the 26th IEEE Conference on Computational Complexity*. IEEE, 2011.

[Yao90] A. Yao. On ACC and threshold circuits. In *Proceedings of the 31st IEEE Symposium on Foundations of Computer Science*, pages 619–627. IEEE, 1990.