Rounds in Communication Complexity Revisited

Noam Nisan[*]
Hebrew University

Avi Widgerson[*]
Hebrew University and Princeton University

Abstract The k-round two-party communication complexity was studied in the deterministic model by [14] and [4] and in the probabilistic model by [20] and [6]. We present new lower bounds that give (1) randomization is more powerful than determinism in k-round protocols, and (2) an *explicit* function which exhibits an exponential gap between its k and $(k-1)$-round randomized complexity.

We also study the three party communication model, and exhibit an exponential gap in 3-round protocols that differ in the starting player.

Finally, we show new connections of these questions to circuit complexity, that motivate further work in this direction.

[*]This work was partially supported by the Wolfson Research Awards (both authors) and the American-Israeli Binational Science Foundation grant 89-00126 (first author), administered by the Israel Academy of Sciences and Humanities

1 Introduction

1.1 The Two-Party Model

Papadimitriou and Sipser [14] initiated the study of how Yao's model [1] [19] of communication complexity is affected by limiting the two players to only k rounds of messages. They considered the following natural problem g_k: each of the players A and B is given a list of n pointers (each of $\log n$ bits), each pointing to a pointer in the list of the other. Their task is to follow these pointers, starting at some fixed $v_0 \in A$, and find the k^{th} pointer. This can easily be done in k rounds and complexity $O(k \log n)$: A starts and the players alter-

[1]In fact, they and [4] considered the stronger "arbitrary partition" model, but known simulation results of [4, 6, 9, 10] allow us to use Yao's standard "fixed partition" model without loss of generality

nately send the value of the next pointer. It is not clear how to use less than $n \log n$ bits if only $(k-1)$ rounds are allowed or in fact with k-rounds but player B starts. Indeed, [14] conjectured that the complexity is exponentially higher (for fixed k), namely that there is a strict hierarchy, and proved it for the case $k = 2$. The general case was resolved by Duris, Galil, and Schnitger [4] who gave an $\Omega(n/k^2)$ lower bound on the $(k-1)$ round complexity of g_k.

It is not difficult to see that allowing randomness g_k can be solved with high probability in $(k-1)$ rounds using only $O((n/k)\log n))$ communication bits. Another $\log n$ factor in the complexity can make this a Las Vegas (errorless) algorithm. This raises the question: *what is the relative power of randomness over determinism in k-round protocols?* Without limiting the number of rounds [12] showed a quadratic gap between Las Vegas and Determinism, and allowing error, the gap can be exponential.

We use simple information theoretic and probabilistic arguments to strengthen the lower bound of [4] in two ways. First we improve their $(k-1)$-round deterministic lower bound on g_k to $\Omega(n)$ (regardless of k), thus showing that randomness can be cheaper by a factor of $k/\log^2 n$ for k-round protocols. This result also provides the largest gap known for $k > \log n$ in the deterministic model - the previous one was obtained in [4] via counting arguments. The fact that the simulation on [10] is constructive, gives the same gap in the arbitrary partition model for an explicit function, resolving an open question of [4].

Second, we prove that the probabilistic upper bound above is not very far from optimal – we give an $\Omega(n/k^2)$ lower bound, establishing an exponential gap in the probabilistic setting between k and $(k-1)$-round protocols for an explicitly given function. The existance of such functions (with somewhat larger gap) was proved by Halstenberg and Reischuk [6], via complicated counting arguments. The only previous exponential gap for an explicit function was shown for $k = 2$ by Yao [20]. We stress the simplicity of our proof technique, in contrast to that of [6]. We have recently learned that similar techniques were used by Smirnov [16] to obtain an $\Omega(n/(k(\log n)^k))$ lower bound on g_k, which is much weaker than our bound, but gives the exponential gap.

Finally, we use the communication complexity characterization of circuit depth of [8] to establishe g_k as a "complete" problem for monotone depth-k Boolean circuits. (This result was independently discovered by Yanakakis [18]). Thus a simple deterministic reduction enables to derive the monotone constant-depth hierarchy of [7] from the constant-round hierarchy of [4]. (The reverse direction was proven in [7]). We speculate that our new probabilistic lower bound may serve to extend the monotone circuit hierarchy result to depth above $\log n$, via probabilistic reductions (as was done in [15]).

1.2 The Multi-Party Model

Chandra, Furst and Lipton [2] devised the multi-party communication complexity model. Here t players $P_1, P_2, \ldots, P_t$ are trying to compute a Boolean function $g(x_1, x_2, \ldots, x_t)$, where $x_i \in \{0,1\}^{n_i}$. (Until now all work in this model considered equal length inputs, i.e. $n_i = n$ for all i). The twist is that every player P_i sees *all* values x_j for $j \neq i$. This model turns out to capture diverse computational models. [2] used it to prove that majority requires superlinear length constant width branching programs. Babai, Nisan and Szegedy [1] gave $\Omega(n/2^t)$ lower bounds for explicit functions g, and used it for Turing machine, branching program and fomrulae lower bounds, as well as efficient pseudorandom generator for small space. Recently, Goldman and Hastad [5] used the results in [1] to prove lower bounds on constant-depth threshold circuits.

We consider only the 3-player model, and within it allow three rounds of communciation: one per player. We exhibit a function u whose complexity is $\Omega(\sqrt{n})$ if P_3 is the first to speak, but $O(\log n)$ otherwise. The proof uses properties of universal hash functions developed in [13, 11].

It is interesting that u acts on different size arguments; $u : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^{\log n} \rightarrow \{0,1\}$ so $n_1 = n_2 = n$, but $n_3 = \log n$. The following connection to circuit complexity makes such functions important. We show that improving our lower bound to $\Omega(n)$ for some explicit function g of this form gives the following size-depth trade-off: the function $f : \{0,1\}^{2n} \rightarrow \{0,1\}^n$ defined by $f(x_1, x_2)_{x_3} = g(x_1, x_2, x_3)$ cannot be computed by Boolean circuits of size $O(n)$ and depth $O(\log n)$ simultaneously. This result is obtained via Valiant's [17] method of depth-reduction in circuits.

2 The Two-Party Model

The four subsections of this section give the definitions, results, technical lemmas and some proofs, respectively in the two-party communication complexity model.

2.1 Definitions

Let $g : X_A \times X_B \rightarrow \{0,1\}$ be a function. The players A, B receive respectively inputs $x_A \in X_A$, $x_B \in X_B$. A k-round protocol specifies for each input a sequence of k messages, $m_1, m_2, \ldots, m_k$ sent alternately between the players such that at the end both know $g(x_A, x_B)$. The cost of a k-round protocol is $\sum_{i=1}^{k} |m_i|$ (where $|m_i|$ is the binary length of m_i), maximized over all inputs (x_A, x_B). Denote by $C^{A,k}(g)$ (resp. $C^{B,k}(g)$) the cost of the best protocol in which player A (resp. B) sends the first message, and $C^k(g) = \min\{C^{A,k}(g), C^{B,k}(g)\}$.

Let $T : X_A \times X_B \rightarrow \{0,1\}$ be the function computed by the two players following a protocol T. We introduce randomization by allowing T to be a random variable distributed over deterministic protocols. The cost is simply the expectation of the associated random variable. We say that randomized protocol makes ϵ-error if $\Pr[T(x_A, x_B) \neq g(x_A, x_B)] \leq \epsilon$ for every input $(x_A, x_B) \in X_A \times X_B$. Denote by $C_\epsilon^k(g)$ the cost of the best k-round ϵ-error protocol for g, and similarly define $C_\epsilon^{A,k}$, $C_\epsilon^{B,k}$. The case $\epsilon = 0$ (e.g. $C_0^k(g)$) denotes Las Vegas (errorless) protocols.

Finally, if we leave T a deterministic protocol, and choose the input uniformly at random, we can define the ϵ-error distributional complexity $D^k_\epsilon(g)$ to be the cost of the best k-round protocol for which $\Pr[T(x_A, x_B) \neq g(x_A, x_B)] \leq \epsilon$, under this distribution. The following lemmas are useful.

Lemma 1 [20] *For every* $g\,,\epsilon > 0$ $D^k_{2\epsilon}(g) \leq 2C^k_\epsilon(g)\,.$

Lemma 2 *For every* $\frac{1}{3} \geq \epsilon > \epsilon' > 0$ $C^k_{\epsilon'}(g) = O(C^k_\epsilon(g))\,.$

2.2 Results

Let $V_A\,,\,V_B$ be two disjoint sets (of vertices) with $|V_A| = |V_B| = n$ and $V = V_A \cup V_B\,.$ Let $F_A = \{f_A : V_A \to V_B\}$, $F_B = \{f_B : V_B \to V_A\}$ and $f = (f_A, f_B) : V \to V$ defined by $f(v) = \begin{cases} f_A(v) & v \in V_A \\ f_B(v) & v \in V_B \end{cases}$. For each $k \geq 0$ define $f^{(k)}(v)$ by $f^{(0)}(v) = v\,,\ f^{(k+1)}(v) = f(f^{(k)}(v))\,.$

Let $v_0 \in V_A\,.$ The functions we will be interested in computing is $g_k : F_A \times F_B \to V$ defined by $g_k(f_A, f_B) = f^{(k)}(v_0)$.

Remarks: In the following theorems note that the number of input bits to each player is $n \log n$, and that they hold for every value of k. We also note that one can make g_k a Boolean function by taking (say) the parity of the output vertex. All our upper and lower bounds apply to this Boolean function as well.

Theorem 1 *[14]* $C^{A,k}(g_k) = O(k \log n)\,.$

Theorem 2 $C^{B,k}(g_k) = \Omega(n)\,.$

Theorem 3 $C^{B,k}_{1/3}(g_k) = O((n/k)\log n)$ $C^{B,k}_0(g_k) = O((n/k)\log^2 n)$.

Theorem 4 $C^{B,k}_{1/3}(g_k) = \Omega(\frac{n}{k^2})$.

In the remainder we show the "completeness" of g_k for monotone depth k circuits. Let $g_k = g_{k,n}$ to stress that each player gets n vertices.

Definition: For a boolean function h define $L^d(h)$ to be the size of the minimal *monotone formula* of depth d and unbounded fanin that computes h. Define $LS^d(k,n)$ to be the maximum of $L^d(h)$ over all functions h that can be computed by monotone circuits of unbounded fanin depth k and total size n. Define $LF^d(k,n)$ to be the maximum of $L^d(h)$ over all functions h that can be computed by a formula of depth k and fanin n at each gate.

Theorem 5:

$$\log LS^d(k,n) \le C^d(g_{k,n}) \le \log LF^d(k,n)$$

The left inequality was proven in [7], and allowed them to deduce a lower bound on g_k from their circuit lower bound. The right inequality was independently discovered by Yanakakis [18]. It allows to recover the tight hierarchy theorem of [7] from the lower bound on g_k.

Let h_k be the complete function for depth k-circuits, i.e. an alternating and-or tree of depth k and fanin $n^{1/k}$ at each gate.

Corollary [7]: Any monotone circuit of depth $k-1$ for h_k requires size $2^{\Omega(n^{1/k}/k)}$.

2.3 Probability, Measure and Information Theory

Let Ω be a finite set (universe), $X \subseteq \Omega$. Denote by $\mu(X)$ the *density* of X in Ω, $\mu(X) = \frac{|X|}{|\Omega|}$. Let $P : X \to [0,1]$ a probability distribution on X, and $x \in X$ a random variable distributed according to P. The probability of any event $Y \subseteq X$ is denoted $\Pr_P[Y]$, and the subscript P is usually omitted. For $y \in X$, we write $\Pr[\{y\}] = P_y$. Then the *entropy* $H(P) = H(x) = \sum_{y\in X} P_y \log P_y$. The *information* on X (relative to Ω), is $I(x) = \log|\Omega| - H(x)$. If P is the uniform distribution U on X, then $H(x) = \log|X|$, and $I(x) = -\log \mu(X)$.

The following lemmas will be useful to us.

Lemma 3 *For every P*
$\Pr_P[\{y : P_y \le \alpha\}] \le \alpha|X|$.

Lemma 4 *For every P and if $x = (x_1, x_2, \ldots, x_m)$, (so $\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$ and x_i distributed over Ω_i), then $I(x) \ge \sum_{i=1}^{m} I(x_m)$.*

The next lemma (from [15]) shows that if $I(x)$ is very small, one can get good bounds on the probability of any event under P in terms of its probability under the uniform distribution U.

Lemma 5 [15] *For $Y \subseteq X$, let $q = \Pr_U[Y]$. Assume $\Delta = \sqrt{\frac{4I(x)}{q}} \le \frac{1}{10}$. Then $|\Pr_P[Y] - q| \le q\Delta$.*

Lemma 6 *If $X = \Omega = \{0,1\}$, $I(x) \le \delta \le \frac{1}{4}$, then $|P_0 - \frac{1}{2}|, |P_1 - \frac{1}{2}| \le 2\sqrt{\delta}$.*

2.4 Proofs

Proof of Theorems 1 and 3.

$C^{A,k}(g_k) \leq k \log n$ follows easily, since in round t the right player knows $f(v_{t-1}) = v_t$ and can send these $\log n$ bits to the second player.

The idea in beating the determinstic $\Omega(n)$ lower bound when the wrong player B starts is as follows: First B chooses a random subset $U \subseteq V_B$ with $|U| = 10n/k$, and sends to A $\{f_B(u) : u \in U\}$. Now it is A's turn and they start sending each other $v_1, v_2, \ldots$ as above, but lagging one round "behind schedule". However, with probability $\geq 2/3$, one of the v_i's will be in U, which allows them to save two rounds, and "finish on time". This gives $C_\epsilon^{B,k}(g_k) = O((k + n/k) \log n)$. This algorithm can be made Las-Vegas with an extra factor of $O(\log n)$ in the complexity.

Proof of theorems 2,4

Let $f = (f_A, f_B) \in F_A \times F_B$ be the input. Let T' be a deterministic k-round protocol for g_k in which B sends the first message. Note that at any round $t \geq 1$, if it is B's turn to speak, then $v_{t-1} = f^{(t-1)}(v_0) \in V_A$, and vice versa. It will be convenient to replace T' by a protocol T in which in any round $t \geq 1$, we replace the message m by the message (m, v_{t-1}). By induction on t, this is always possible for the player whose turn it is. In particular, it implies that $\geq \log n$ bits are sent per round. Thus if T' used C bits, T uses $\leq C + k \log n$ bits. We will assume T uses $\frac{\epsilon n}{2}$ bits, (ϵ will be chosen later), and obtain a contradiction.

Every node z of the protocol tree T can be labeled by the rectangle $F_A^z \times F_B^z$ of inputs arriving at z. By the structure of T, if z is at level $t \geq 1$ (the root is at level 0), then $v_0, v_1, \ldots, v_{t-1}$ are determined in $F_A^z \times F_B^z$.

We shall assume the input is chosen uniformly at random from $F_A \times F_B$, so in fact we shall bound from below the distributional comlexity. Thus the probability of arriving at z is $\mu(F_A^z \times F_B^z)$, and given that the input arrived at z, it is uniformly distributed in $F_A^z \times F_B^z$. The main lemma below intuitively shows that if the input arrived at z and the rectangle at z has nice properties, then with high (enough) probability the input will proceed to a child w of z which is equally nice. Nice means that both F_A^z, F_B^z are large enough, and that the player *not* holding v_{t-1} has very little information on $v_t = f(v_{t-1})$.

Denote by c_z the total number of bits sent by the players before arriving at z. Assume without loss of generality that A speaks at z. Let $f_A^z(f_B^z)$ be random variables uniformly distributed over $F_A^z(F_B^z)$. Recall that T uses $\leq \frac{\epsilon}{2}n$ bits, and let δ satisfy $\delta = \text{Max}\ 4\sqrt{\epsilon}, 400\,\epsilon$. Define z to be *nice* if it satisfies:

1. $I(f_A^z) \leq 2c_z$
2. $I(f_B^z) \leq 2c_z$
3. $I(f_B^z(v_{t-1})) \leq \delta$

Main Lemma:

If z is nice, and w a random child of z, then $\Pr[w \text{ not nice}] \leq 4\sqrt{\epsilon} + \frac{1}{n}$.

Proof: Assume A sends $c(\geq \log n)$ bits at z. (In general the possible messages in a particular step may differ in length. For simplicity, we assume here they don't. Handling the general case requires only a slight changes in the proof of claim 2 below). Thus $c_w = c_z + c$ for all children w of z. We will now give upper bounds separately on the probability of each of the three properties defining nice being false at a random child w.

Claim 1: $\Pr[I(f_B^w) > 2c_w] = 0$.

Proof: B sent nothing, so $\forall\, w\ F_B^w = F_B^z$ and

$$I(f_B^w) = I(f_B^z) \leq 2c_z < 2c_w. \quad \square$$

Claim 2: $\Pr[I(f_A^w) > 2c_w] \leq \frac{1}{n}$.

Proof: Z has 2^c children, and child w is chosen with probability $\mu(F_A^w)\big/\mu(F_A^z)$. Thus by Lemma 3 $\Pr[I(f_A^w) > 2c_w] \leq \Pr[\mu(F_A^w)\big/\mu(F_A^z) < 2^{-2c}] \leq 2^{-c} \leq 1/n$. $\square$

Claim 3: $\Pr[I(f_A^w(v_t)) > \delta] \leq 4\sqrt{\epsilon}$.

Proof: We may assume now that $I(f_A^w) \leq 2c_w \leq \epsilon n$. The random variable f_A^w is a vector of random variables $f_A^w(v)$ for all $v \in V_A$. Thus by Lemma 4, $\sum_{v\in V} I(f_A^w(v)) \leq I(f_A^w) \leq \epsilon n$. So if v_t was chosen *uniformly* from v_A, $\Pr_U[I(\delta_A^w(v_t)) > \delta] \leq \frac{\epsilon}{\delta}$ by Markov's inequality. But $v_t = f_B^z(v_{t-1})$, so v_t is distributed with $I(v_t) = I(f_B^z(v_{t-1})) \leq \delta$ as

we assumed z was nice. By Lemma 5 (and our choice of δ),
$\Pr[I(f_A^w(v_t)) > \delta] \le \frac{\epsilon}{\delta}\left(1 + \sqrt{\frac{4\delta}{\epsilon/\delta}}\right)$
$\le 4\sqrt{\epsilon}$. □

Now we can conclude the proofs of Theorems 2 and 4 form the main lemma. Consider any *nice* leaf ℓ of the protocol tree T, labeled by an answer (0 or 1). Say A spoke on the last round k. Then $I(v_k) = I(f_B^\ell(v_{k-1})) \le \delta$. So by Lemma 6, even if the algorithm gives one bit (say parity) of the answer, it is correct with probability $\le \frac{1}{2} + 2\sqrt{\delta}$.

Conclusion of Theorem 2 Take $\epsilon = 10^{-4}$. The root of T is nice, so by the main lemma and induction we have a positive probability ($\ge 2^{-k}$) of reaching a nice leaf, contradicting the fact that the protocol never errs. This proves only $C^{B,k}(g_k) = \Omega(n - k\log n)$, since we augmented an arbitrary T' to a nice protocol T.

The lower bound $C^{B,k}(g_k) = \Omega(n)$ (which is stronger when $k \ge \frac{n}{\log n}$) requires a more delicate argument that we sketch below. The idea is to follow the same steps of the proof with the following changes. (1) We stay with the original protocol T', as we cannot afford the players sending $\log n$ bits per round as in the nice protocol T. (2) We still fix the vertex v_{t-1} by the player sending the message at round t, but avoid paying $\log n$ bits for this information by removing this vertex from our universe. Thus the information I is measured relative to a smaller set of pointers at every round. (3) We prove a weaker main lemma, which is clearly sufficient in the deterministic case, namely that every nice node z has at least one nice child w. The details are left to the interested reader.

Conclusion of Theorem 4. Pick $\epsilon = 10^{-4} \cdot k^{-2}$. Thus the probability of not reaching a nice leaf is $\le k\frac{1}{25k} = \frac{1}{25}$, and the probability that the protocol answers correctly is less than $\frac{1}{25} + (\frac{1}{2} + \frac{2}{5\sqrt{k}}) < 0.95$. Thus we get $D^{B,k}_{1/20}(g_k) = \Omega(\frac{n}{k^2} - k\log n)$, or $\Omega(\frac{n}{k^2})$ for all $k < (\frac{n}{\log n})^{1/3}$. The theorem for this range of k follows from Lemmas 1 and 2. The higher range of values for k is handled by the trivial $\Omega(k)$ lower bound for k-round protocols, which is stronger in this range.

Proof of theorem 5: As mentioned above, the left inequality was proven in [7], so we prove only the right inequality. The proof is based on the Karchmer-Wigderson characterization of circuit depth in terms of communication complexity, which can be stated as follows. For every monotone function h on n variables with minterms $Min(h)$ and maxterms $Max(h)$ define a communication search problem $R_h^m \subset Min(h) \times Max(h) \times [n]$ in which player A gets a minterm $S \in Min(h)$, player B gets a maxterm $T \in Max(h)$, and their task is to find an element in $S \cap T$. Then monotone formulae for h and protocols for R_h^m are in 1-1 correspondence via the simple syntactic identification of $\vee$ gates with palyer A's moves and $\wedge$ gates with player B's moves. In particular, depth correspondes to the number of rounds, and logarithm of the size to the communication complexity.

In view of the above, all we need to give now is a reduction from computing $g_{k,n}$ to the computation of R_h^m for some function h which has a depth k formula of fanin n at each gate. Once this is done the players can solve R_h^m and hence $g_{k,n}$ in d rounds and $logLF^d(k,n)$ communication by simulating the guaranteed depth d circuit for h.

Let h be defined by a formula that is a complete n-ary tree of depth k, alternating levels of $\vee$ and $\wedge$ gates (say with $\vee$ at the root), and distinct n^k variables at the leaves. The players agree on a fixed labeling of the nodes of this tree in which the root is labeled v_0, the childern of every $\vee$ gates labeled by V_B, and children of every $\wedge$ gate labeled by V_A. Let f_A and f_B be the inputs to players A, B respectively. Player A constructs sets S_i of nodes from the ith level inductively as follows. S_0 contains the root. If level i contains $\vee$ gates, than for every gate in S_i labeled v he adds to S_{i+1} the unique child of this gate labeled $f_A(v)$. If level i contains $\wedge$ gates, then for every gate in S_i he adds all its children to S_{i+1}. In a similar way (exchanging the roles of gates) player B constructs his sets T_i. It is easy to verify that S_k is a minterm of h, T_k is a maxterm of h, and that they intersect at a unique leaf, whose label is $f^(k)(v_0)$. This completes the reduction, and hence the proof.

3 The Three-Party Model

Let $g : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \times \{0,1\}^{n_3} \rightarrow \{0,1\}$ be a function. Players P_1, P_2, P_3 are given (x_2, x_3), (x_1, x_3), (x_1, x_2) respectively with $x_i \in \{0,1\}^{n_i}$ and compute g from this information by exchanging messages according to a predetermined protocol. We consider only 3-round protocols in which each player speakes once. Let $M^i(g)$ denote the communication complexity when player P_i speaks first (and then the other two in arbitrary order), and $M^s(g)$ the complexity when they all speak simultaneously (an oblivious protocol). Clearly, for all $i \in \{1,2,3\}$ $M^i(g) \leq M^s(g)$.

Let $u : \{0,1\}^{2n} \times \{0,1\}^n \times \{0,1\}^{\log n} \rightarrow \{0,1\}$ be the following function. Interpret the first string x_1 as a 2-universal hash function ([3]) h, mapping $\{0,1\}^n$ to itself, the second string x_2 as an argument y to h, and the third x_3 as an index $j \in [n]$. Then $u(h, y, j) = h(y)_j$. The next two theorems exhibit an exponential gap between 3-round protocols that differ in the order in which players speak.

Theorem 6: $M^1(u) = M^2(u) = O(\log n)$

Theorem 7: $M^3(u) = \Omega(\sqrt{n})$.

Let $f : \{0,1\}^m \rightarrow \{0,1\}^n$ be an arbitrary function, and for any $m' < m$ define $g_f : \{0,1\}^{m'} \times \{0,1\}^{m-m'} \times \{0,1\}^{\log n} \rightarrow \{0,1\}$ by $g_f(x_1, x_2, x_3) = f(x_1 \circ x_2)_{x_3}$, where $\circ$ denotes concatenation. The next theorem gives the relationship of size-depth trade-offs in circuits to 3-round oblivious protocols.

Theorem 8: If f above can be computed by a circuit of size $O(n)$ and depth $O(\log n)$, then $M^s(g_f) = O(n/\log\log n)$.

Proof of Theorem 7

Restrict the value of j to be $j \in [\sqrt{n}]$. Thus we consider $h : \{0,1\}^n \rightarrow \{0,1\}^{\sqrt{n}}$ which is still a universal hash function. Assume $M^3(u) \leq \sqrt{n}/4$. This means that there is a new protocol to compute $z = h(y)$ in which P_3 sends $\sqrt{n}/4$ bits, and then players P_1 and P_2 can compute each bit of z separately, using altogether $n/4$ bits.

Pick values m_1, m_2, m_3 to the messages of P_1, P_2, P_3 in this new protocol with the largest "support", and take (h, y) uniformly at random. As $|m_3| \leq \sqrt{n}/4$, and $|m_1|, |m_2| \leq n/4$ we have

$\Pr[h(y) = z \,|\, m_1, m_2, m_3] \leq$
$2^{\sqrt{n}/4} \Pr[h(y) = z \,|\, m_1, m_2] \stackrel{(*)}{\leq}$
$2^{\sqrt{n}/4} \cdot 2^{-\sqrt{n}/2} = 2^{-\sqrt{n}/4} < 1.$

The inequality $(*)$ follows from Lemma 10 of [11] regarding the distribution of hash values when little information is given on each of h, y.

Proof of Theorem 8

Let $f : \{0,1\}^m \to \{0,1\}^n$ be computed by a circuit C of size $O(n)$ and depth $O(\log n)$. By a result of Valiant [17], there are $s = O(n/\log\log n)$ wires in C, $e_1, e_2, \ldots, e_s$ with the following property. For every input $x \in \{0,1\}^m$, and every $j \in [n]$, $f(x)_j$ is determined by the values $e_1(x), \ldots, e_s(x)$ on these wires, together with the values of x_i, $i \in S_j$ with $|S_j| \leq n^\epsilon$. To compute g_f, note that P_3 has access to $x = (x_1 \circ x_2)$ (which is the input to f) can compute the values on the wires. P_2 and P_3, now knowing $j = x_3$, exchange the necessary bits in S_j to complete the computation of $f(x)$.

References

[1] L. Babai, N. Nisan, M. Szegedy: *Multiparty protocols and logspace-hard pseudorandom sequences* Proc. of the 21st STOC (1989) 1-11.

[2] A. Chandra, M. Furst, R. Lipton: *Multiparti Protocols* Proc. of the 15th STOC (1983) 94-99.

[3] L. Carter, M. Wegman: *Universal hash functions* Journal of Comp. and Sys. Sci. 18 (1979) 143-154.

[4] P. Duris, Z. Galil, G. Schnitger: *Lower Bounds of Communication Complexity* Proc. of the 16^{th} STOC, (1984) 81–91

[5] M. Goldman, J. Hastad: *On the power of small depth threshold circuits* Proc. of the 31st FOCS, (1990) 610-618.

[6] B. Halstenberg, R. Reischuk: *On Diferent Modes of Communication* Proc. of the 20th STOC (1988) 162-172.

[7] M. Klawe, W.J. Paul, N. Pippenger, M. Yannakakis: *On Monotone Formulae with Restricted Depth* Proc. of the 16^{th} STOC, (1984) 480–487

[8] M. Karchmer, A. Wigderson: *Monotone Circuits for Connectivity Require Super-Logarithmic Depth* Proc of the 20^{th} STOC, (1988), 539–550

[9] T. Lam, L. Ruzzo: *Results on Communication Complexity Classes* Proc. of the 4th Structures in Complexity Theory conference, (1989) 148-157.

[10] L.A. McGeoch: *A Strong Seperation Between k and $k-1$ Round Communication Complexity for a Constructive Language* CMU Technical Report CMU-CS-86-157 (1986)

[11] Y. Mansour, N. Nisan, P. Tiwary: *The computational complexity of universal hashing* Proc. of the 22nd STOC (1990).

[12] K. Melhorn, E. Schmidt: *Las Vegas is better than Determinism in VLSI and Distributed Computation* Proc of the 14th STOC (1982), 330-337.

[13] N. Nisan: *Pseudorandom Generators for Space Bounded Computation* Proc. of the 22nd STOC, (1990) 204-212.

[14] P.H. Papadimitriou, M. Sipser: *Communication Complexity* Proc. of the 14th STOC, (1982) 330–337

[15] R. Raz, A. Wigderson: *Probabilistic Communication Complexity of Boolean Relations* Proc. of the 30th FOCS, (1989) 562–567

[16] D. V. Smirnov: *Shannon's information methods for lower bounds for probabilistic communication complexity*, Manuscript (in Russian) (1989)

[17] L. Valiant: *Graph theoretic arguments in low-level complexity* Technical Report CS 13-77, University of Edinburgh (1977).

[18] M. Yanakakis: Private communication.

[19] A.C.-C. Yao: *Some Complexity Questions Related to Distributive Computing* Proc of the 11th STOC, (1979) 209–213

[20] A.C.-C. Yao: *Lower Bounds by Probabilistic Arguments* Proc of the 24th FOCS, (1983) 420–428