On Communication Complexity of Vector–valued Functions

Rudolf Ahlswede and Ning Cai

Universität Bielefeld
Fakultät für Mathematik
Postfach 100131
33501 Bielefeld
Germany

Abstract.

New upper and lower bounds on the 2–way communication complexity of abstract functions $g : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ give tight bounds, when applied to vector–valued functions $f^n = (f_1, \dots, f_n) : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$, if the alphabets are small. For the set–intersection function an optimal protocol is presented. It is based on a simple new idea applicable also to abstract functions. The 2–way communication complexities of all other Boolean functions are also determined. The results are extended to meets in abstract lattices and to a probabilistic model.

keywords and phrases

2–way communication complexity, vector–valued functions, Kronecker product, prefix codes, correlated sources, rank, alternating partitions.

1. Introduction.

Let $\mathcal{X}, \mathcal{Y}$ and $\mathcal{Z}$ be finite sets. For any function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ we consider the (vector–valued) functions $f^n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$ defined by

$$f^n(x^n, y^n) = \big(f(x_1, y_1), \dots, f(x_n, y_n)\big) \tag{1.1}$$

for $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ and $y^n = (y_1, \dots, y_n) \in \mathcal{Y}^n$ and study their 2–way communication complexity $C(f^n; 1 \leftrightarrow 2)$, that is, the minimal number of bits which have to be exchanged for any argument (x^n, y^n) between a person $P_\mathcal{X}$ knowing x^n and a person $P_\mathcal{Y}$ knowing y^n so that both can calculate $f^n(x^n, y^n)$.

We also consider cases of non–identical component functions, that is, we are given sequences $(\mathcal{X}_t)_{t=1}^\infty$, $(\mathcal{Y}_t)_{t=1}^\infty$ and $(\mathcal{Z}_t)_{t=1}^\infty$ of finite sets, a sequence $(f_t)_{t=1}^\infty$ of functions $f_t : \mathcal{X}_t \times \mathcal{Y}_t \to \mathcal{Z}_t$ and the vector–valued functions $f^n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$, where $\mathcal{X}^n = \prod_{t=1}^n \mathcal{X}_t$, $\mathcal{Y}^n = \prod_{t=1}^n \mathcal{Y}_t$, $\mathcal{Z}^n = \prod_{t=1}^n \mathcal{Z}_t$, and $f^n(x^n, y^n) = \big(f_1(x_1, y_1), \dots, f_n(x_n, y_n)\big)$.

In particular, we sometimes assume that

$$\mathcal{X}_t = \mathcal{Y}_t = \mathcal{Z}_t = \{0, 1, \dots, \alpha_t - 1\} . \tag{1.2}$$

We denote the set of all protocols for calculating f^n by $\mathcal{Q}_{f^n}$ and for any protocol $Q \in \mathcal{Q}_{f^n}$ $\ell_Q(x^n, y^n)$ is the number of bits exchanged for inputs x^n and y^n . In the worst case we get

$$L(Q) \triangleq \max_{x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n} \ell_Q(x^n, y^n), \tag{1.3}$$

the length of the protocol Q . In this terminology

$$C(f^n; 1 \leftrightarrow 2) = \min_{Q \in \mathcal{Q}_{f^n}} L(Q) . \tag{1.4}$$

For abstract functions $g : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ with $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ finite Yao [1] introduced a k–decomposition as a partition $\mathcal{P} = \{U_1 \times V_1, \dots, U_k \times V_k\}$ of $\mathcal{X} \times \mathcal{Y}$ into g–monochromatic rectangles. Those are rectangels $U \times V$ $(U \subset \mathcal{X}, V \subset \mathcal{Y})$ on which g is constant. Let $\mathcal{D}_g$ be the set of all such decompositions. For the decomposition number

$$D(g) \triangleq \min_{\mathcal{P} \in \mathcal{D}_g} |\mathcal{P}| \tag{1.5}$$

Yao's inequality (in the improved form of [9]) states

$$C(g; 1 \leftrightarrow 2) \geq \log_2 D(g) . \tag{1.6}$$

The function table (or matrix) associated with g shall be denoted by M_g . We refer to the k–decompositions above also as k–decomposition of M_g . Mehlhorn and Schmidt

[2] gave an important lower bound for $D(g)$. They associated with M_g the class of matrices $\{\Delta_z : z \in \mathcal{Z}\}$, where

$$\Delta_z(x,y) = \begin{cases} 1 & \text{iff } M_g(x,y) = z \\ 0 & \text{otherwise,} \end{cases} \tag{1.7}$$

and established the inequality

$$D(g) \geq \sum_{z \in \mathcal{Z}} \text{rank}_{\mathbb{F}}(\Delta_z), \tag{1.8}$$

where the rank is taken over any field $\mathbb{F}$. Here we always choose the field of rationals and omit therefore the index $\mathbb{F}$.

There is a simple, but useful lower bound on $D(g)$. We call the set $\{(x_1,y_1),\dots,(x_i,y_i)\}$ an independent set of Δ_z if $\Delta_z(x_t,y_t)=1$ for $t=1,2\dots,i$, but no two members of the set are in the same monochromatic rectangle of Δ_z . The maximum cardinality of such a set is the independence number $\text{ind}(\Delta_z)$. Clearly,

$$D(g) \geq \sum_z \text{ind}(\Delta_z) \triangleq I(g) \ . \tag{1.9}$$

The matrix $\begin{pmatrix} 1\,1\,0\,0 \\ 0\,0\,1\,1 \\ 1\,0\,1\,0 \\ 0\,1\,0\,1 \end{pmatrix}$ with independence number 4 and rank 3 is an example for " $\text{rank}(\Delta_z) \not\geq \text{ind}(\Delta_z)$ ".

For the orientation of the reader we mention that other authors (for example [11]) speak of fooling sets and sizes instead of independent sets and independent numbers.

<u>There are three seemingly basic observations in the present paper.</u>

1.) <u>The protocol $\overline{Q}$:</u>

For the set intersection function Λ^n (Boolean "and") the decomposition number is 3^n (Lemma 1 in Section 2) and we found a protocol $\overline{Q}$ whose length achieves this bound. (Theorem 1 in Section 2). Other Boolean functions are easy to analyze. Moreover, all functions $f^n = (f,\dots,f)$ with $|\mathcal{X}| = |\mathcal{Y}| = 2$, $|\mathcal{Z}| \leq 4$ fall into 4 classes, which can be reduced to Boolean cases (Corollary 1 in Section 3). Here and also for the analysis of meets in abstract lattices variations of the basic protocol $\overline{Q}$ are used. (Theorems 7,8,9 in Section 9)

2.) <u>General upper bound:</u>

Partitions arising in protocols are more special than k–decompositions . For a function g we denote the class of those "protocol" generated partitions by $\mathcal{A}_g$ and introduce the protocol partition number

$$D^*(g) = \min_{\mathcal{P} \in \mathcal{A}_g} |\mathcal{P}| \ . \tag{1.10}$$

Whereas obviously

$$C(g;1\leftrightarrow 2)\geq \log D^*(g) \tag{1.11}$$

we found an equally simple upper bound involving the depth $d(\mathcal{P})$ (defined in Section 4):

$$C(g;1\leftrightarrow 2)\leq \min_{\mathcal{P}\in\mathcal{A}_g}\big(\log|\mathcal{P}|+d(\mathcal{P})\big)\ . \tag{1.12}$$

Explicit definitions, the result (Theorem 2), and proof can be found in Section 4. Consequences for vector–valued functions, in particular with constant alphabets, gives Theorem 3 in Section 5. Theorem 4 in Section 6 states a probabilistic version of (1.12) and again in a probabilistic vector–valued setting (for so called correlated sources) there is also an asymptotic result (Theorem 5, Section 7).

3.) Lower bound for vector–valued functions:

The independence number, I, and the rank function show multiplicative behaviour under the Kronecker product (Lemmas 2,3 in Section 8). Our new observation is that the bounds can be combined on a letter by letter basis, yielding a bound better than the maximum of the two original bounds (Theorem 6 in Section 8). Finally, this bound in conjunction with our upper bound (1.12) enables us to determine $C(f^n;1\leftrightarrow 2)$ within $o(n)$ for small alphabets. The precise results states Theorem 10 in Section 10.

2. Exact solutions for all Boolean functions.

In the Boolean case, that is, $\mathcal{X}=\mathcal{Y}=\mathcal{Z}=\{0,1\}$, one readily verifies that every Boolean function can be transformed by exchanges of 0 and 1 into one of the following functions (defined by tables)

$$B_1:\ \begin{array}{cc} & \begin{array}{cc}0&1\end{array}\\ \begin{array}{c}0\\1\end{array} & \begin{pmatrix}0&0\\0&0\end{pmatrix}\end{array},\ B_2:\begin{pmatrix}0&0\\1&1\end{pmatrix},\ B_3:\begin{pmatrix}0&0\\0&1\end{pmatrix},\ B_4:\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

Here B_1 is a constant–function, B_2 is the projection on the first argument, B_4 is the symmetric difference Δ , and finally B_3 stands for the logical “and” Λ , which is also called intersection function.

Now obviously $C(B_1^n;1\leftrightarrow 2)=0$ and $C(B_2^n;1\leftrightarrow 2)=n$, because $P_{\mathcal{X}}$ has to inform $P_{\mathcal{Y}}$ about x^n and this also suffices. Furthermore, $C(B_4^n;1\leftrightarrow 2)=2n$ because $P_{\mathcal{X}}$ (resp. $P_{\mathcal{Y}}$) can recover y^n (resp. x^n) from $\Delta(x^n,y^n)$ and x^n (resp. y^n). The analysis of B_3^n is less obvious, but the answer is again “smooth”. We summarize our findings.

Lemma 1. *For the decomposition number we have*

$D(B_3^n)=3^n$.

Theorem 1.

For the four types of Boolean functions we have for the two–way communication complexities

$$C(B_i^n; 1 \leftrightarrow 2) = \lceil n \log i \rceil \quad for \quad i = 1, 2, 3, 4. \tag{2.1}$$

Proof of Lemma 1:

For fixed z^n and $x^n \geq z^n$, that is, $x_t \geq z_t$ for $t = 1, 2, \dots, n$, consider the sets

$$S(x^n, z^n) = \big\{(x^n, y^n) : x^n \wedge y^n = z^n\big\} \ . \tag{2.2}$$

They partition the set

$$S(z^n) = \bigcup_{x^n \geq z^n} S(x^n, z^n) \tag{2.3}$$

into 2^{n-k} monochromatic rectangles, when $k = w(z^n)$ counts the 1's in z^n .

Since there are $\binom{n}{k}$ sequences z^n with weight $w(z^n) = k$, we can therefore partition the whole table into $\sum_{k=0}^{n} \binom{n}{k} 2^{n-k} = 3^n$ monochromatic rectangles.

Converseley, it suffices to show that the set $S(z^n)$ requires in any partition $2^{n-w(z^n)}$ rectangles.

But this is the case, because

$$\big\{(x^n, \overline{x^n}) : x_t = 1, \quad \text{if} \quad z_t = 1, 1 \leq t \leq n\big\},$$

is an independent (fooling) set of size $2^{n-w(2^n)}$ in $S(z^n)$, if $\overline{x^n} = (1-x_1, \dots, 1-x_n)$.

Proof of Theorem 1:

After the discussion preceding Lemma 1 we have to consider only the function B_3^n . Using inequality (1.6) Lemma 1 implies

$$C(B_3^n; 1 \leftrightarrow 2) \geq \lceil n \log 3 \rceil. \tag{2.4}$$

Since the lower bound in (1.6) is based on prefix codes it is natural to look how close to the truth protocols based on such codes lead in the present case. Surprisingly, this gives an exact bound.

To fix ideas let us recall a trivial protocol: $\mathcal{P}_{\mathcal{X}}$ sends x and $\mathcal{P}_{\mathcal{Y}}$ sends $g(x, y)$. This shows that $C(g; 1 \leftrightarrow 2) \leq \lceil \log |\mathcal{X}| \rceil + \lceil \log |\mathcal{Z}| \rceil$. There is a smarter 2 rounds protocol!

Knowing x $\mathcal{P}_{\mathcal{X}}$ also knows $\big\{g(x, y) : y \in \mathcal{Y}\big\}$. In case this set is large, the length $\ell(x)$ of the codeword for x should be small, and in case this set is small, x can get a longer codeword in order to minimize the worst case total number of bits. $\mathcal{P}_{\mathcal{Y}}$ knows when to start transmission if $\mathcal{P}_{\mathcal{X}}$ uses a prefix code. In the present case, if $\mathcal{P}_{\mathcal{X}}$ describes x^n , then $\mathcal{P}_{\mathcal{Y}}$ must describe all y_t's of y^n where x_t is 1 . The goal is to have a prefix code encoding x^n with $\ell(x^n)$ bits such that $L = \max_{x^n \in \mathcal{X}^n} \big(\ell(x^n) + w(x^n)\big)$ is minimal. By Kraft's inequality L must satisfy $\sum_{x^n} 2^{-(L-w(x^n))} \leq 1$.

Since $\sum_{x^n} 2^{w(x^n)} = 3^n$ the minimal L equals $\lceil n \log 3 \rceil$.

3. Exact solutions for the cases $|\mathcal{X}| = |\mathcal{Y}| = 2$.

In case $|\mathcal{Z}| \leq 2$ one readily verifies that the functions are equivalent to one of the Boolean functions $B_1, \dots, B_4$. Also, in case $|\mathcal{Z}| = 4$ for an f taking really 4 values $C(f_n; 1 \leftrightarrow 2)$ equals $2n$, because here anyone of the processors can recover the other input from the value of f_n . Thus we are left with the case $\mathcal{Z} = \{0, 1, 2\}$ and w.l.o.g. with the tables

$$M_g : \begin{pmatrix} 0 & 0 \\ 2 & 1 \end{pmatrix}, \qquad M_h : \begin{pmatrix} 0 & 1 \\ 2 & 0 \end{pmatrix} .$$

With the map φ_j , which sends 2 to j and i to i otherwise, we can write $\Lambda = \varphi_0(g)$ and $\Delta = \varphi_1(h)$ and thus by Theorem 1

$$\begin{aligned} C(g_n; 1 \leftrightarrow 2) &\geq C(B_3^n; 1 \leftrightarrow 2) \geq \lceil n \log 3 \rceil \\ C(h_n; 1 \leftrightarrow 2) &\geq C(B_2^n; 1 \leftrightarrow 2) \geq 2n \ . \end{aligned}$$

On the other hand, $C(h^n; 1 \leftrightarrow 2) \leq 2n$ follows with the simple protocol, where $P_{\mathcal{X}}$ sends x^n and $P_{\mathcal{Y}}$ sends y^n .

The bound $C(g^n; 1 \leftrightarrow 2) \leq \lceil n \log 3 \rceil$ can be derived with the protocol $\overline{Q}$ we used for Λ^n . In step c.) $P_{\mathcal{Y}}$ knows x^n and y^n and can therefore calculate $g^n(x^n, y^n)$. As before he also calculates $x^n \wedge y^n$ and proceeds again with d.). In step e.) $P_{\mathcal{X}}$ knows again $x^n \wedge y^n$ and of course x^n . This enables him to determine $g^n(x^n, y^n)$, because $x_t = 0$ implies $g(x_t, y_t) = 0$, $x_t = 1$ and $x_t \wedge y_t = 1$ implies $g(x_t, y_t) = 1$ and, finally, $x_t = 1$ and $x_t \wedge y_t = 0$ implies $y_t = 0$ and thus $g(x_t, y_t) = 2$.

We summarize our findings.

Corollary 1: The functions $f^n = (f, \dots, f)$ with $|\mathcal{X}| = |\mathcal{Y}| = 2$, $|\mathcal{Z}| \leq 4$ fall into classes $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3, \mathcal{Z}_4$. The members in $\mathcal{Z}_i$ have the same 2–way complexity as B_i .

4. A general upper bound via protocol generated partitions.

The present approach lifts the upper bound proof of Theorem 1 to a general function $f : \mathcal{X} \times Y \to \mathcal{Z}$ (all sets finite with table or matrix M_f). Whereas Yao considered the set $\mathcal{D}_f$ of all (abstract) partitions of M_f into monochromatic rectangles we use here the set $\mathcal{A}_f$ of protocol generated partitions of M_f into monochromatic rectangles. They are defined as follows. Suppose the partitions $\mathcal{P}_1 = \{A_1 \times B_1, \dots, A_k \times B_k\}$ and $\mathcal{P}_2 = \{A'_1 \times B'_1, \dots, A'_n \times B'_n\}$ divide a rectangle into subrectangles and for all $A'_i \times B'_i$ there exist $i_1, \dots, i_j$ such that $A'_i = A_{i_1} \cup A_{i_2} \cup \dots \cup A_{i_j}$ and $B_{i_1} = \dots = B_{i_j} = B'_i$, then we call $\mathcal{P}_1$ a row–refinement of $\mathcal{P}_2$ and write $\mathcal{P}_2 \rhd_r \mathcal{P}_1$. The column–refinement is defined analogously. We say that a partition $\mathcal{P}$ of M_f into monochromatic rectangles is protocol generated, if there are partitions $\mathcal{P}_i (i = 1, 2, \dots, k)$ with

$$\{M_f\} = \mathcal{P}_0 \rhd_1 \mathcal{P}_1 \cdots \rhd_k \mathcal{P}_k = \mathcal{P}, \quad \text{where}$$

$\rhd_i$ means a row–refinement iff $\rhd_{i-1}$ is a column–refinement. The minimal k of such a representation of $\mathcal{P}$ is called the depth $d(\mathcal{P})$.

Obviously, any $\mathcal{P} \in \mathcal{A}_f$ can be produced by a two–way communication protocol. A basic quantity is $D^*(f) = \min\{|\mathcal{P}| : \mathcal{P} \in \mathcal{A}_f\}$. The lower bound implicit in [9] is

$$C(f; 1 \leftrightarrow 2) \geq \log D^*(f) \; . \tag{4.1}$$

Here is our simple upper bound.

Theorem 2.

For all $\mathcal{P} \in \mathcal{A}_f$ there is a protocol Q with $L(Q) \leq \log |\mathcal{P}| + d(\mathcal{P})$. In particular,

$$C(f; 1 \leftrightarrow 2) \leq \min_{\mathcal{P} \in \mathcal{A}_f} \big(\log |\mathcal{P}| + d(\mathcal{P}) \big) \; .$$

Proof: We proceed by induction in d . If $d(\mathcal{P}) = k$, then there are partitions $\mathcal{P}_1, \dots, \mathcal{P}_{k-1}$ with $\{M_f\} \rhd_1 \mathcal{P}_1 \rhd_2 \mathcal{P}_2 \cdots \rhd \mathcal{P}_{k-1} \rhd_k \mathcal{P}$. Assuming w.l.o.g. that $\rhd_1$ is a row–refinement we have $\mathcal{P}_1 = \{A_i \times \mathcal{Y}\}_{i=1}^m$. Suppose now that $A_i \times \mathcal{Y}$ contains a_i subrectangles in $\mathcal{P}$. Then $|\mathcal{P}| = \sum_i a_i$ and the partition $\mathcal{P}_{(i)}$ of $A_i \times \mathcal{Y}$ (into the a_i subrectangles) induced by $\mathcal{P}$ has depth not greater than $d(\mathcal{P}) - 1$.

By induction hypothesis for all $A_i \times \mathcal{Y}$ there are protocols $Q_{(i)}$ with

$$L(Q_{(i)}) \leq \log |\mathcal{P}_{(i)}| + d(\mathcal{P}_{(i)}) \leq \log a_i + d(\mathcal{P}) - 1 \; . \tag{4.2}$$

Since $\left(\frac{a_1}{\sum_j a_j}, \frac{a_2}{\sum_j a_j}, \dots, \frac{a_m}{\sum_j a_j} \right)$ is a probability distribution, by the Noiseless Coding Theorem there is a binary prefix–code $\{c_1, c_2, \dots, c_m\}$ with $\text{length}(c_i) \leq \log \frac{\sum_j a_j}{a_i} + 1 = \log \sum_j a_j - \log a_i + 1$. With $b_i \triangleq \lfloor \log \sum_j a_j - \log a_i \rfloor + 1$ bits $P_\mathcal{X}$ can send i , the label of $A_i \times \mathcal{Y}$, to $P_\mathcal{Y}$. By (4.2) $b_i + L(Q_{(i)}) \leq \log \sum_j a_j + d(\mathcal{P}) = \log |\mathcal{P}| + d(\mathcal{P})$.

5. Application of upper bound to vector–valued functions.

The functions $f^n : \mathcal{X}^n \times \mathcal{Y}^n \to \mathcal{Z}^n$, $f^n = (f_1, f_2, \dots, f_n)$ were defined in the Introduction. We shall apply the bound of Theorem 2. A simple observation is that for partitions $\mathcal{P}_t = \{D_{t,j}\}_j \in \mathcal{A}_{f_t}$ $(t = 1, 2, \dots, n)$ we have

$$\mathcal{P}^n = \left\{ \prod_{t=1}^{n} D_{t,j_t} \right\}_{(j_1,\dots,j_n)} \in \mathcal{A}_{f^n} \ . \tag{5.1}$$

Furthermore, we show next that $d(\mathcal{P}^n)$ is controlled by

$$d(\mathcal{P}^n) \le \max_{1 \le t \le n} \big\{ 2 \cdot \min\left(|\mathcal{X}_t|, |\mathcal{Y}_t|\right) - 2 \big\} \triangleq \theta, \text{ say } . \tag{5.2}$$

This is the case, because we can cut first M_{f^n} row–wise by cutting all component spaces for which the first round in their protocol partition are row–refinements. In the next step there must be column–refinements in all components, etc. We thus reach $\mathcal{P}^n$ after at most $\theta + 1$ rounds.

The one extra round is not needed, if $f_t = f \, (t = 1, 2, \dots, n)$ Theorem 2 in conjunction with (5.1) and (5.2) imply therefore the following "single–letter" upper bound.

Theorem 3. *For* $n \in \mathbb{N}$

$$C(f^n; 1 \leftrightarrow 2) \le \sum_{t=1}^{n} \log D^*(f_t) + \theta + 1 \ .$$

If $f_t = f$ *for* $t = 1, 2, \dots, n$ *, then*

$$C(f^n; 1 \leftrightarrow 2) \le n \log D^*(f) + \theta \ .$$

Eventhough $n \log\big(D^*(f)\big)$ may be much larger than $\log D^*(f^n)$ we show in Section 10 that quite surprisingly this is not the case if one of the alphabets $\mathcal{X}, \mathcal{Y}$ is smaller than 5.

6. Application of upper bound to a probabilistic model.

Let (X, Y) be a pair of random variables with X taking values in $\mathcal{X}$, Y taking values in $\mathcal{Y}$ and joint distribution P_{XY} . For $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ and a protocol Q $\ell_Q(x, y)$ is the length of the protocol in the input (x, y) . We are interested in the expected 2–way communication complexity

$$\widetilde{C}(f; 1 \leftrightarrow 2) \triangleq \min_{Q \in \mathcal{Q}_f} \mathbb{E} \ \ell_Q(X, Y) \ .$$

For $\mathcal{P} = \{A_1, \dots, A_m\} \in \mathcal{A}_f$ consider the entropy

$$H_{\mathcal{P}} \triangleq H\big(P_{XY}(A_1), \ldots, P_{XY}(A_m)\big) \ .$$

Theorem 4.

$$\widetilde{C}(f; 1 \leftrightarrow 2) \le H_{\mathcal{P}} + d(\mathcal{P})$$

for all $\mathcal{P} \in \mathcal{A}_f$.

Proof: The derivation is almost the same as that of Theorem 2. The difference is this: When we employ the induction hypothesis for $\{A_i \times \mathcal{Y} : i = 1, 2, \ldots, m\}$, we give a protocol Q_i for the restriction of f on $A_i \times \mathcal{Y}$. This happens with probability $P_{XY}(A_i \times \mathcal{Y})$. Conditional on the event $A_i \times \mathcal{Y}$ for $\mathcal{P}_i$ (defined as in the proof of Theorem 2) by induction hypothesis

$$\mathbb{E}\, \ell_{Q_i} \le H_{\mathcal{P}_i | A_i \times \mathcal{Y}} + d(\mathcal{P}) - 1 \ .$$

We have to add the prefix free coding of $\{A_i \times \mathcal{Y}\}_i$ with an average length not greater than $H\big(P_{XY}(A_1 \times \mathcal{Y}), \ldots, P_{XY}(A_m \times \mathcal{Y})\big)$. Now just use that by the grouping axiom of entropy

$$\begin{aligned} H_{\mathcal{P}} = &\sum_i P_{XY}(A_i \times \mathcal{Y}) H_{\mathcal{P}_i | A_i \times \mathcal{Y}} \\ &+ H\big(P_{XY}(A_1 \times \mathcal{Y}), \ldots, P_{XY}(A_m \times \mathcal{Y})\big) \ . \end{aligned}$$

7. An asymptotic result for correlated sources.

We consider the vector–valued case with identical components, that is,

$$f^n = (f, \ldots, f) \ .$$

Furthermore, we assume that $(X_t, Y_t)_{t=1}^{\infty}$ is a sequence of independent, identically distributed pairs of random variables.

Theorem 5. *Under the preceding probabilistic assumptions*

$$\lim_{n \to \infty} \frac{1}{n} \widetilde{C}(f^n; 1 \leftrightarrow 2) = \lim_{n \to \infty} \frac{1}{n} \min_{\mathcal{P} \in \mathcal{A}_{f^n}} H_{\mathcal{P}} \ .$$

Proof: For any $\varepsilon > 0$ choose first $n(\varepsilon)$ such that

$$\frac{1}{n(\varepsilon)} \min_{\mathcal{P} \in \mathcal{A}_{f^{n(\varepsilon)}}} H_{\mathcal{P}} \le \frac{\varepsilon}{2} + \underline{\lim}_{n \to \infty} \frac{1}{n} \min_{\mathcal{P} \in \mathcal{A}_{f^n}} H_{\mathcal{P}}$$

and then choose $m^*(\varepsilon)$ and $\mathcal{P}^* = \{A_1, \ldots, A_{m^*(\varepsilon)}\} \in \mathcal{A}_{f^{n(\varepsilon)}}$ such that

$$\frac{1}{n(\varepsilon)} H_{\mathcal{P}^*} \le \varepsilon + \underline{\lim}_{n\to\infty} \frac{1}{n} \min_{\mathcal{P}\in\mathcal{A}_{f^n}} H_{\mathcal{P}} \ . \tag{7.1}$$

Use now for any r the partition $\mathcal{P}^{*(r)} = \left\{ A_{j_1} \times \cdots \times A_{j_r} : 1 \le j_1, \dots, j_r \le m^{*(\varepsilon)} \right\}$ obtained as product from $\mathcal{P}^*$. Then $\mathcal{P}^{*(r)} \in \mathcal{A}_{f^{n(\varepsilon)r}}$ and

$$H_{\mathcal{P}^{*(r)}} = r H_{\mathcal{P}^*} \ . \tag{7.2}$$

By Theorem 4 $\widetilde{C}(f^{n(\varepsilon)r}; 1 \leftrightarrow 2) \le H_{\mathcal{P}^{*(r)}} + d(\mathcal{P}^{*(r)})$ and thus by (5.2) and (7.2)

$$\frac{1}{n(\varepsilon)r} \widetilde{C}(f^{n(\varepsilon)r}; 1 \leftrightarrow 2) \le \frac{1}{n(\varepsilon)} H_{\mathcal{P}^*} + \frac{\theta}{n(\varepsilon)r} \ . \tag{7.3}$$

Using the inequality (7.1) we continue with

$$\frac{1}{n(\varepsilon)r} \widetilde{C}(f^{n(\varepsilon)r}; 1 \leftrightarrow 2) \le \varepsilon + \underline{\lim}_{n\to\infty} \frac{1}{n} \min_{\mathcal{P}\in\mathcal{A}_{f^n}} H_{\mathcal{P}} + \frac{\theta}{n(\varepsilon)r} \tag{7.4}$$

and

$$\overline{\lim}_{r\to\infty} \frac{1}{n(\varepsilon)r} \widetilde{C}(f^{n(\varepsilon)r}; 1 \leftrightarrow 2) \le \varepsilon + \underline{\lim}_{n\to\infty} \frac{1}{n} \min_{\mathcal{P}\in\mathcal{A}_{f^n}} H_{\mathcal{P}} \ . \tag{7.6}$$

Since for $n_1 \le n_2$

$$\widetilde{C}(f^{n_1}; 1 \leftrightarrow 2) \le \widetilde{C}(f^{n_2}; 1 \leftrightarrow 2) \tag{7.7}$$

we also have for $n \in \left[n(\varepsilon)(r-1), n(\varepsilon)r \right]$

$$\frac{1}{n} \widetilde{C}(f^n; 1 \leftrightarrow 2) \le \frac{r}{r-1} \frac{1}{n(\varepsilon)r} \widetilde{C}(f^{n(\varepsilon)r}; 1 \leftrightarrow 2) \tag{7.8}$$

and thus by (7.6), and since ε was arbitrary,

$$\overline{\lim}_{n\to\infty} \frac{1}{n} \widetilde{C}(f^n; 1 \leftrightarrow 2) \le \underline{\lim}_{n\to\infty} \frac{1}{n} \min_{\mathcal{P}\in\mathcal{A}_{f^n}} H_{\mathcal{P}} \ . \tag{7.9}$$

The lower bound is immediate from the Noiseless Coding Theorem.

8. Lower bounds for vector–valued functions.

Using Yao's bound $C(f^n; 1 \leftrightarrow 2) \geq \log D(f^n)$ and

$$D(f^n) \geq I(f^n) \geq \prod_{t=1}^{n} I(f_t) \tag{8.1}$$

we get

Lemma 2. $C(f^n; 1 \leftrightarrow 2) \geq \sum_{t=1}^{n} \log I(f_t)$.

Recall now the definition of Δ_z in (1.7) and set

$$r(f_t) = \sum_{z \in \mathcal{Z}_t} \operatorname{rank}(\Delta_{z_t}) \ . \tag{8.2}$$

Next observe that, in analogous notation,

$$\Delta_{z^n} = \bigotimes_{t=1}^{n} \Delta_{z_t}, \tag{8.3}$$

where $\bigotimes$ denotes the Kronecker product of matrices, i.e.

$$\Delta_{z^n}(x^n, y^n) = \Delta_{z_1}(x_1, y_1)\Delta_{z_2}(x_2, y_2) \dots \Delta_{z_n}(x_n, y_n) \ .$$

Therefore $\sum_{z^n \in \mathcal{Z}^n} z^n \Delta_{z^n} = \sum_{z^n \in \mathcal{Z}^n} z^n \bigotimes_{t=1}^{n} \Delta_{z_t}$, and

$$\begin{aligned} r(f^n) &= \sum_{z^n \in \mathcal{Z}^n} \operatorname{rank}(\Delta_{z^n}) = \sum_{z^n \in \mathcal{Z}^n} \operatorname{rank}\left(\bigotimes_{t=1}^{n} \Delta_{z_t}\right) \\ &= \sum_{z^n \in \mathcal{Z}^n} \prod_{t=1}^{n} \operatorname{rank}(\Delta_{z_t}) = \prod_{t=1}^{n} \sum_{z_t \in \mathcal{Z}_t} \operatorname{rank}(\Delta_{z_t}) \\ &= \prod_{t=1}^{n} r(f_t) \ . \end{aligned}$$

We have derived an identity.

Lemma 3. $r(f^n) = \prod_{t=1}^{n} r(f_t)$.

From (1.6), (1.8) and this identity follows a useful lower bound.

Lemma 4. $C(f^n; 1 \leftrightarrow 2) \geq \sum_{t=1}^n \log r(f_t)$.

Especially for the stationary case $f^n = (f, \dots, f)$ the bound depends only on one component f . In Information Theory this is called a “single–letter characterisation”.

The bound in Lemma 2 and Lemma 4 can be combined to one bound

$$C(f^n; 1 \leftrightarrow 2) \geq \max \left(\sum_{t=1}^n \log I(f_t), \sum_{t=1}^n \log r(f_t) \right) . \tag{8.4}$$

Whereas this bound is canonical and perhaps known, now comes as novelty an improvement of this bound.

Using the independence numbers $\mathrm{ind}(\Delta_z)$ we can write

$$I(f_t) = \sum_{z_t \in \mathcal{Z}_t} \mathrm{ind}(\Delta_{z_t}) . \tag{8.5}$$

We need the local maxima

$$j(\Delta_{z_t}) = \max\bigl(\mathrm{rank}(\Delta_{z_t}), \mathrm{ind}(\Delta_{z_t})\bigr) \tag{8.6}$$

and their sums

$$J(f_t) = \sum_{z_t \in \mathcal{Z}_t} j(\Delta_{z_t}) .$$

Theorem 6.

$$C(f^n; 1 \leftrightarrow 2) \geq \log D(f^n) \geq \sum_{t=1}^n \log J(f_t) . \tag{8.7}$$

Proof: Denote Yao's decomposition number of Δ_{z^n} by $\delta(\Delta_{z^n})$. Since

$$D(f^n) = \sum_{z^n} \delta(\Delta_{z^n}) \tag{8.8}$$

and since $\prod_{t=1}^n J(f_t) = \sum_{t=1}^n \sum_{z_t \in \mathcal{Z}_t} j(\Delta_{z_t}) = \sum_{z^n} \prod_{t=1}^n j(\Delta_{z_t})$, (8.7) follows by summation over z^n , if we know that

$$\delta(\Delta_{z^n}) \geq \prod_{t=1}^n j(\Delta_{z_t}) . \tag{8.9}$$

In order to verify this inequality, we introduce first an auxiliary matrix Δ_z^* as follows. Let Ω_z be an independent set of Δ_z with $|\Omega_z| = \mathrm{ind}(\Delta_z)$. Then set $\Delta_z^* = \Delta_z$ if $\mathrm{rank}(\Delta_z) \geq \mathrm{ind}(\Delta_z)$, and otherwise define

$$\Delta_z^*(x,y) = \begin{cases} 1, & \text{if } (x,y) \in \Omega_z \\ 0, & \text{otherwise.} \end{cases}$$

If we can prove that

$$\delta(\Delta_{z^n}) \geq \delta\left(\bigotimes_{t=1}^{n} \Delta_{z_t}^*\right) , \tag{8.10}$$

thus we are done, because

$$\begin{aligned} \delta\left(\bigotimes_{t=1}^{n} \Delta_{z_t}^*\right) &\geq \prod_{t=1}^{n} \operatorname{rank}(\Delta_{z_t}^*) \\ &= \prod_{t=1}^{n} \max\bigl(\operatorname{rank}(\Delta_{z_t}), \operatorname{ind}(\Delta_{z_t})\bigr) \\ &= \prod_{t=1}^{n} j(\Delta_{z_t}) \ . \end{aligned}$$

Now just observe that a monochromatic partition of $\Delta_{z_1} \times \Delta_{z_2}$ restricted to $\Delta_{z_1}^* \times \Delta_{z_2}$ (in case $\Delta_{z_1} \neq \Delta_{z_1}^*$) gives again a monochromatic partition and thus $\delta(\Delta_{z_1} \bigotimes \Delta_{z_2}) \geq \delta(\Delta_{z_1}^* \bigotimes \Delta_{z_2})$. Finally, (8.10) follows by repetition of this argument.

9. Generalisations of the result on the set–intersection problem to meets in a lattice.

Consider first the case of chains, that is, $\mathcal{X}_t = \mathcal{Y}_t = \mathcal{Z}_t = \{0, 1, 2, \ldots, \alpha_t - 1\}$ and

$$\Lambda^n(x^n, y^n) = (x_1 \wedge y_1, \ldots, x_n \wedge y_n) \ . \tag{9.1}$$

Theorem 7.

$$\sum_t \log(2\alpha_t - 1) \leq C(\Lambda^n; 1 \leftrightarrow 2) \leq \sum_t \log(2\alpha_t - 1) + 2 \max_t \alpha_t - 3 \ .$$

Remark: When $\alpha_t = 2$ for $t = 1, 2, \ldots, n$ the result is only slightly weaker than that of Theorem 1, the difference being 1 bit in the upper bound.

Proof: The lower bound follows from either Lemma 2 or Theorem 6. The upper bound is one bit better than what follows from Theorem 3. This one bit can be saved, because parts of the second last partition have sizes which are powers of 2 (c.f. proof of Theorem 1).

Suppose now that $\mathcal{L}$ is a finite lattice.

Let $(\ell] = \{x : x \leq \ell\}, [\ell) = \{x : \ell \leq x\}$ for all $\ell \in \mathcal{L}$, and $S = \{(\ell, \ell') : \ell \leq \ell'\} \subset \mathcal{L}^2$. Clearly, $\sum_{\ell \in \mathcal{L}} |[\ell)| = \log |S|$.

Considering the family of monochromatic rectangles $\{(\ell, x) : x \in [\ell)\}, \ell \in \mathcal{L}$, Theorem 2 yields

$$C(\Lambda; 1 \leftrightarrow 2) \leq \log \sum_{\ell \in \mathcal{L}} |[\ell)| + 1 = \log |S| + 1 \ . \tag{9.2}$$

We use an additional structure to get also a lower bound.

The element $\ell^* \in \mathcal{L}$ is called a pseudocomplement of $\ell \in \mathcal{L}$ iff $\ell^* \wedge \ell = 0$ and $\ell \wedge x = 0$ imply $x \leq \ell^*$.

$\mathcal{L}$ is called pseudocomplemented, if every element has a pseudocomplement. Notice that for pseudocomplemented $\mathcal{L}$ for any $\ell \in \mathcal{L}$ $[\ell)$ is also pseudocomplemented ($a \in [\ell)$ has a relative pseudocomplement $a^* \vee \ell$, where a^* is pseudocomplement of a in $\mathcal{L}$).

Now the proof of Theorem 1 generalizes (see also [12]) to yield

Theorem 8. *For a finite pseudocomplemented lattice*

$$\lceil \log |S| \rceil \leq C(\Lambda; 1 \leftrightarrow 2) \leq \lfloor \log |S| \rfloor + 1 \ .$$

For general (not necessarily pseudocomplemented) lattices we derive now a lower bound on $C(\Lambda; 1 \leftrightarrow 2)$ along the lines initiated by Hajnal, Maass and Turan [3], who introduced the Möbius function in the study of 2–way communication complexity for estimating the rank of the function table.

In particular we rely upon an idea of Lovasz and Saks [4], who used the following result.

Theorem: (Lindström and Dowling & Wilson, [5], [6])

For the disjointness indicator $T_{\mathcal{L}}$ of a lattice $\mathcal{L}$, that is a matrix whose rows and columns are labelled by the elements of $\mathcal{L}$ and $T_{\mathcal{L}}(\ell, k) = 1$ if ℓ and k are disjoint, and $T_{\mathcal{L}}(\ell, k) = 0$ otherwise, $\operatorname{rank}(T_{\mathcal{L}}) = |\{\ell \in \mathcal{L} : \mu(\phi, \ell) \neq 0\}|$, where μ is the Möbius inverse.

Consider now $\Lambda : \mathcal{L}^2 \to \mathcal{L}$ and for all $\ell \in \mathcal{L}$ $\Delta_\ell = T_{[\ell)}$. From $C(\Lambda; 1 \leftrightarrow 2) \geq \log r(\Lambda)$ it follows with the previous Theorem that for $S^* \triangleq \{(\ell, \ell') : \mu(\ell, \ell') \neq 0\} \subset S$

$$C(\Lambda; 1 \leftrightarrow 2) \geq \log r(\Lambda) \geq \lceil \log |S^*| \rceil \ . \tag{9.3}$$

In summary we thus have

Theorem 9.

$$\lceil \log |S^*| \rceil \leq C(\Lambda; 1 \leftrightarrow 2) \leq \lfloor \log |S| \rfloor + 1 \ .$$

For Boolean lattices, subspaces of finite fields and the partition lattice $|S^*| = |S|$. For multi–sets $|S^*| \neq |S|$, but Theorem 7 shows that the lower bound in (9.3) is "almost tight".

This investigation has been continued in [12].

10. Sharp results for vector–valued functions over small alphabets on one side.

For small alphabets Theorem 6 in Section 8 can now be used to derive a "single–letter" lower bound for the complexity $C(f^n; 1 \leftrightarrow 2)$, which meets asymptotically the upper bound of Theorem 3. In these case we have the surprising fact that equality holds in

$$D(f^n) \leq \prod_{t=1}^{n} D(f_t) \ . \tag{10.1}$$

Theorem 10. *Suppose that* $|\mathcal{X}_t| \leq \min(4, |\mathcal{Y}_t|)$ *for* $t \in \mathbb{N}$ *, then*

$$(i) \quad D^*(f_t) = D(f_t) \qquad (ii) \quad J(f_t) = D(f_t) \tag{10.2}$$

and

$$\sum_t \log D(f_t) + \theta + 1 \geq C(f^n; 1 \leftrightarrow 2) \geq \sum_{t=1}^{n} \log D(f_t) \ . \tag{10.3}$$

Particularly, for $f^n = (f, \dots, f)$

$$n \log D(f) + \theta \geq C(f^n; 1 \leftrightarrow 2) \geq n \log D(f) \ . \tag{10.4}$$

Proof: The upper bounds follow from Theorem 3 and identity (i) in (10.2).

The lower bounds in (10.3) and (10.4) follow from Theorem 6 and the identity (ii) in (10.2), which we now prove.

Clearly, it suffices to show that

$$\delta(\Delta_z) = j(\Delta_z) \quad \text{for} \quad z \in \mathcal{Z} \ . \tag{10.5}$$

If $\mathrm{rank}(\Delta_z) = |\mathcal{X}|$, then $\delta(\Delta_z) = |\mathcal{X}| = j(\Delta_z)$. We need to consider only those z for which $\mathrm{rank}(\Delta_z) < |\mathcal{X}|$. For then we actually show even that

$$\delta(\Delta_z) = \mathrm{ind}(\Delta_z) \tag{10.6}$$

and thus a fortiori (10.5) holds.

We can assume that there is no all 0 row or column and that there are no two identical rows or columns in Δ_z , because otherwise rows or columns can be removed without changing anyone of the parameters of interest.

<u>Case $|\mathcal{X}| = 3$:</u> Since now $\mathrm{rank}(\Delta_z) \leq 2$, w.l.o.g. the first row is a linear combination of the two others. Therefore the matrix is of the form

$$\Delta_z = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and (10.6) holds.}$$

<u>Case $|\mathcal{X}| = 4$:</u> The case $\mathrm{rank}(\Delta_z) \leq 2$ being obvious again, we assume that $\mathrm{rank}(\Delta_z) = 3$. We also assume that the last 3 rows in Δ_z are lineary independent. As subcases arise:

a) The first row vector $\vec{V_1}$ is a linear combination of two others, say $\vec{V_2}$ and $\vec{V_3}$. Thus, necessarily
$$\vec{V_1} = \vec{V_2} + \vec{V_3} \tag{10.7}$$

b) $\vec{V_1} = \lambda_2\vec{V_2} + \lambda_3\vec{V_3} + \lambda_4\vec{V_4}, \lambda_i > 0$

c) $\lambda_1\vec{V_1} + \lambda_2\vec{V_2} = \lambda_3\vec{V_3} + \lambda_4\vec{V_4}, \lambda_i > 0$.

In case a) Δ_z has one of the forms

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ ? & ? & 0 & 1 & 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ ? & ? & ? & ? & 1 \end{pmatrix}$$

and in both cases (10.6) holds.

In case b) there are the forms

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

and again (10.6) holds.

Finally, in case b) we have the forms

$$\begin{pmatrix} 1&1&0&0 \\ 0&0&1&1 \\ 1&0&1&0 \\ 0&1&0&1 \end{pmatrix}, \begin{pmatrix} 1&1&0&1 \\ 0&0&1&1 \\ 1&0&0&1 \\ 0&1&1&1 \end{pmatrix}, \begin{pmatrix} 1&1&0&0&1 \\ 0&0&1&1&1 \\ 1&0&1&0&1 \\ 0&1&0&1&1 \end{pmatrix}$$

and (10.6) holds in all cases!

References.

[1] A. Yao, Some complexity questions related to distributive computing, in "Proceedings 11th Ann. ACM Sympos. Theory of Computing, 1979", 209–213.

[2] K. Mehlhorn and E.M. Schmidt, Las Vegas is better than determinism in VLSI and distributed computing, in "Proceedings 14th ACM STOC, 1982", 330–337.

[3] A. Hajnal, W. Maass, and G. Turán, On the communication complexity of graph properties, in "Proceedings 20th ACM STOC, 1988", 186–191.

[4] L. Lovász and M. Saks, Lattices, Möbius functions and communication complexity, in "Proceedings 29th IEEE FOCS, 1988", 81–90.

[5] B. Lindström, Determinants on semilattices, Proc. AMS 20 (1969), 207–208.

[6] T. Dowling and R. Wilson, Whitney number inequalities for geometric lattices, Proc. AMS 47 (1975), 504–512.

[7] R. Brualdi, N. Cai, and V. Pless, Orphan structure of the first order Reed–Muller codes, Discrete Math. 102 (1992), 239–247.

[8] U. Tamm, On the communication complexity of sum–type functions invariant under translation, Preprint 91–016 SFB 343 "Diskrete Strukturen in der Mathematik", to appear in Information and Computation.

[9] C.H. Papadimitriou and M. Sipser, Communication complexity, in "Proceedings 14th Ann. ACM Sympos. on Theory of Computing, 1982", 201–214.

[10] R. Ahlswede, N. Cai, and Z. Zhang, A general 4–words inequality with consequences for 2–way communication complexity, Advances in Applied Mathematics 10 (1989), 75–94.

[11] J.D. Ullman, Computational aspects of VLSI, Computer Science Press.

[12] R. Ahlswede, N. Cai, and U. Tamm, Communication complexity in lattices, to appear in Applied Math. Letters.