Electronic Colloquium on Computational Complexity, Report No. 120 (2004)

Lower bounds on the Deterministic and Quantum Communication Complexity of $HAM_n^{(a)}$

Andris Ambainis *
Univ. of Waterloo

William Gasarch †
Univ. of MD at College Park

Aravind Srinivasan ‡
Univ. of MD at College Park

Andrey Utis §
Univ. of MD at College Park

Abstract

Alice and Bob want to know if two strings of length n are almost equal. That is, do they differ on *at most* a bits? Let $0 \le a \le n-1$. We show that any deterministic protocol, as well as any error-free quantum protocol (C^* version), for this problem requires at least $n-2$ bits of communication. We show the same bounds for the problem of determining if two strings differ in *exactly* a bits. We also prove a lower bound of $n/2-1$ for error-free Q^* quantum protocols. Our results are obtained by lower-bounding the ranks of the appropriate matrices.

1 Introduction

Given $x, y \in \{0,1\}^n$ one way to measure how much they differ is the Hamming distance.

Definition 1.1 If $x, y \in \{0,1\}^n$ then $\mathrm{HAM}(x,y)$ is the number of bits on which x and y differ.

If Alice has x and Bob has y then how many bits do they need to communicate such that they both know $\mathrm{HAM}(x,y)$? The trivial algorithm is to have Alice send x (which takes n bits) and have Bob send $\mathrm{HAM}(x,y)$ (which takes $\lceil \lg(n+1) \rceil$ bits) back to Alice. This takes $n + \lceil \lg(n+1) \rceil$ bits. Pang and El Gamal [12] showed that this is essentially optimal. In particular they showed that HAM requires at least $n + \lg(n+1-\sqrt{n})$ bits to be communicated. (See [1, 3, 9, 11] for more on the communication complexity of HAM. See [5] for how Alice and Bob can approximate HAM without giving away too much information.)

What if Alice and Bob just want to know if $\mathrm{HAM}(x,y) \le a$?

Definition 1.2 Let $n \in \mathsf{N}$. Let a be such that $0 \le a \le n-1$. $HAM_n^{(a)} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is the function

$$HAM_n^{(a)}(x,y) = \begin{cases} 1 & \text{if } \mathrm{HAM}(x,y) \le a \\ 0 & \text{otherwise.} \end{cases}$$

*University of Waterloo, Dept. of Combinatorics and Optimization and Instuitute for Quantum Computing, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1 ambainis@uwaterloo.ca, Partially supported by IQC University Professorship and CIAR.

†University of Maryland, Dept. of Computer Science and Institute for Advanced Computer Studies, College Park, MD 20742. gasarch@cs.umd.edu, Partially supported by NSF grant CCR-01-05413.

‡University of Maryland, Dept. of Computer Science and Institute for Advanced Computer Studies, College Park, MD 20742. srin@cs.umd.edu, Partially supported by NSF grant CCR-020-8005.

§University of Maryland, Dept. of Computer Science, College Park, MD 20742. utis@cs.umd.edu

ISSN 1433-8092

The problem $HAM_n^{(a)}$ has been studied by Yao [14] and Gavinsky et al [6]. Yao showed that there is an $O(a^2)$ public coin simultaneous protocol for $HAM_n^{(a)}$ which yields (by Newman [10], see also [7]) an $O(a^2 + \log n)$ private coin protocol and also an $O(2^{a^2} \log n)$ quantum simulataneous message protocol with bounded error [14]. Gavinsky et al. give an $O(a \log n)$ public coin simultaneous protocol, which yields an $O(a \log n)$ private coin protocol. For $a \gg \log n$ this is better than Yao's protocol.

All of the protocols mentioned have a small probability of error. How much communication is needed for this problem if we demand no error? There is, of course, the trivial $(n+1)$-bit protocol. Is there a better one?

In this paper we show the following; in the list of results below, the "c" (in the "$c\sqrt{n}$" terms) is some positive absolute constant.

1. For any $0 \leq a \leq n-1$, $HAM_n^{(a)}$ requires at least $n-2$ bits in the deterministic model.
2. For $a \leq c\sqrt{n}$, $HAM_n^{(a)}$ requires at least n bits in the deterministic model.
3. For any $0 \leq a \leq n-1$, $HAM_n^{(a)}$ requires at least $n-2$ bits in the quantum model with Alice and Bob share an infinite number of EPR pairs, using a classical channel, and always obtain the correct answer.
4. For $a \leq c\sqrt{n}$, $HAM_n^{(a)}$ requires at least n bits in the quantum model in item 3.
5. For any $0 \leq a \leq n-1$, $HAM_n^{(a)}$ requires at least $\frac{n}{2} - 1$ bits in the quantum model with Alice and Bob share an infinite number of EPR pairs, using a quantum channel, and always obtain the correct answer.
6. For $a \leq c\sqrt{n}$, $HAM_n^{(a)}$ requires at least $n/2$ bits in the quantum model in item 5.

Note that if $a = n$ then $(\forall x, y)[HAM_n^{(a)}(x,y) = 1$, hence we do not include that case.

What if Alice and Bob need to determine if $\mathrm{HAM}(x,y) = a$ or not?

Definition 1.3 Let $n \in \mathsf{N}$. Let a be such that $0 \leq a \leq n$. $HAM_n^{(=a)} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is the function

$$HAM_n^{(=a)}(x,y) = \begin{cases} 1 & \text{if } \mathrm{HAM}(x,y) \leq a \\ 0 & \text{otherwise.} \end{cases}$$

We show the exact same results for $HAM_n^{(=a)}$ as we do for $HAM_n^{(a)}$. There is one minor difference: for $HAM_n^{(a)}$ the $a = n$ case had complexity 0 since all pairs of strings differ on at most n bits; however, for $HAM_n^{(=a)}$ the $a = n$ case has complexity $n+1$ as it is equivalent to equality.

All our results use the known "log rank" lower bounds on classical and quantum communication complexity: Lemmas 2.2 and 2.3. Our approach is to lower-bound the ranks of the appropriate matrices, and then to invoke these known lower bounds.

2 Definitions, Notations, and Useful Lemmas

We give brief definitions of both classical and quantum communication complexity. See [7] for more details on classical, and [4] for more details on quantum.

Definition 2.1 Let f be any function from $\{0,1\}^n \times \{0,1\}^n$ to $\{0,1\}$.

1. A *protocol* for computing $f(x,y)$, where Alice has x and Bob has y, is defined in the usual way (formally using decision trees). At the end of the protocol both Alice and Bob know $f(x,y)$.
2. $D(f)$ is the number of bits transmitted in the optimal deterministic protocol for f.
3. $Q^*(f)$ is the number of bits transmitted in the optimal quantum protocol where we allow Alice and Bob to share an infinite number of EPR pairs and communicate over a quantum channel.
4. $C^*(f)$ is the number of bits transmitted in the optimal quantum protocol where we allow Alice and Bob to share an infinite number of EPR pairs and communicate over a classical channel.
5. M_f is the $2^n \times 2^n$ matrix where the rows and columns are indexed by $\{0,1\}^n$ and the (x,y)-entry is $f(x,y)$.

Let lg denote the logarithm to the base two. Also, as usual, if $x < y$, then $\binom{x}{y}$ is taken to be zero.

The following theorem is due to Mehlhorn and Schmidt [8]; see also [7].

Lemma 2.2 *If* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *then* $D(f) \geq \lg(\text{rank}(M_f))$.

Buhrman and de Wolf [2] proved a similar theorem for quantum communication complexity.

Lemma 2.3 *If* $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ *then the following hold.*

1. $Q^*(f) \geq \frac{1}{2}\lg(\text{rank}(M_f))$.

2. $C^*(f) \geq \lg(\text{rank}(M_f))$.

3 The Complexity $HAM_n^{(a)}$ for $a \leq O(\sqrt{n})$

We start by presenting results for general a, and then specialize to the case where $a \leq c\sqrt{n}$.

Definition 3.1 Let M_a be $M_{HAM_n^{(a)}}$, the $2^n \times 2^n$ matrix representing $HAM_n^{(a)}$.

Lemma 3.2 *M_a has 2^n orthogonal eigenvectors.*

Proof: This follows from M_a being symmetric. ▮

We know that M_a has 2^n eigenvalues; however, some of them may be 0. We prove that M_a has few 0-eigenvalues. This leads to a lower bound on $D(HAM_n^{(a)})$ by Lemma 2.2.

Definition 3.3 Let $z \in \{0,1\}^n$.

1. $v_z \in R^{2^n}$ is defined by, for all $x \in \{0,1\}^n$, $v_z(x) = (-1)^{\sum_i x_i z_i}$. The entries $v_z(x)$ of v_z are ordered in the natural way: in the same order as the order of the index x in the rows (and columns) of M_a.

2. We show that v_z is an eigenvector of M_a. Once that is done we let $eig(z)$ be the eigenvalue of M_a associated with v_z.

Lemma 3.4

1. *The vectors* $\{v_z : z \in \{0,1\}^n\}$ *are orthogonal.*
2. *For all* $z \in \{0,1\}^n$, v_z *is an eigenvector of* M_a.
3. *If* z *has exactly* m *1's in it, then*

$$eig(z) = \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} \binom{m}{k}\binom{n-m}{j-k}(-1)^k.$$

Proof: The first assertion (orthogonality) follows by simple counting. We now prove the final two assertions together. Let $z \in \{0,1\}^n$ have exactly m ones in it.

Fix a row in M_a that is indexed by $x \in \{0,1\}^n$. Denote this row by R_x. We need the following notation:

$$\begin{aligned} L_a = & \ \{y \mid \mathrm{HAM}(x,y) \le a\} \\ E_j = & \ \{y \mid \mathrm{HAM}(x,y) = j\} \end{aligned}$$

We will show that $R_x \cdot v_z$ is a constant multiple (independent of x) times $v_z(x)$. Now,

$$R_x \cdot v_z = \sum_{y \in \{0,1\}^n} HAM_n^{(a)}(x,y) v_z(y) = \sum_{y \in L_a} v_z(y) = \sum_{y \in L_a} (-1)^{\sum_i y_i z_i}.$$

We would like to have this equal $b \times v_z(x)$ for some constant b. We set it equal to $b \times v_z(x)$ and deduce what b works. So, suppose

$$b \times v_z(x) = \sum_{y \in L_a} (-1)^{\sum_i y_i z_i}.$$

We have

$$\begin{aligned}
b &= \frac{1}{v_z(x)} \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \\
&= v_z(x) \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \\
&= (-1)^{\sum_i x_i z_i} \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \quad \text{(by the definition of } v_z(x)) \\
&= \sum_{y \in L_a} (-1)^{\sum_i (x_i+y_i) z_i} \\
&= \sum_{y \in L_a} (-1)^{\sum_i |x_i-y_i| z_i} \quad \text{(since } x_i + y_i \equiv |x_i - y_i| \pmod 2) \\
&= \sum_{j=0}^{a} \sum_{y \in E_j} (-1)^{\sum_i |x_i-y_i| z_i} \quad \text{(since } L_a = \textstyle\bigcup_{j=0}^{a} E_j). \qquad (1)
\end{aligned}$$

We partition E_j. If $y \in E_j$ then x and y differ in exactly j places. Some of those places i are such that $z_i = 1$. Let k be such that the number of places where $x_i \neq y_i$ and $z_i = 1$.

Upper Bound on k: Since there are exactly m places where $z_i = 1$ we have $k \le m$. Since there are exactly j places where $x_i \ne y_i$ we have $k \le j$. Hence $k \le \min\{j, m\}$.

Lower Bound on k: Since there are exactly $n - m$ places where $z_i = 0$, we have $j - k \le n - m$. Hence $k \ge \max\{0, j + m - n\}$.

In summary, the only relevant k are $\max\{0, j+m-n\} \le k \le \min\{j, m\}$. Fix j. For $\max\{0, j+m-n\} \le k \le \min\{j, m\}$, let $D_{j,k}$ be defined as follows:

$$D_{j,k} = \{y \mid ((y \in E_j) \wedge (\text{on exactly } k \text{ of the coordinates where } x_i \ne y_i, \text{ we have } z_i = 1))\}.$$

Note that

$$E_j = \bigcup_{k=0}^{\min\{j,m\}} D_{j,k}$$

and $|D_{j,k}| = \binom{m}{k}\binom{n-m}{j-k}$. So, by (1),

$$b = \sum_{j=0}^{a} \sum_{y \in E_j} (-1)^{\sum_i |x_i - y_i| z_i} = \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} \sum_{y \in D_{j,k}} (-1)^{\sum_i |x_i - y_i| z_i}.$$

By the definition of $D_{j,k}$ we know that for exactly k of the values of i we have both $|x_i - y_i| = 1$ and $z_i = 1$. On all other values one of the two quantities is 0. Hence we have the following:

$$\begin{aligned} b &= \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} \sum_{y \in D_{j,k}} (-1)^k \\ &= \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} |D_{j,k}|(-1)^k \\ &= \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} \binom{m}{k}\binom{n-m}{j-k}(-1)^k. \end{aligned}$$

Notice that b is independent of x and is of the form required. ∎

Definition 3.5 Let

$$F(a, n, m) = \sum_{j=0}^{a} \sum_{k=\max\{0,j+m-n\}}^{\min\{j,m\}} \binom{m}{k}\binom{n-m}{j-k}(-1)^k.$$

The following lemma will be used in this section to obtain a lower bound when $a = O(\sqrt{n})$, and in Section 5 to obtain a lower bound for general a.

Lemma 3.6

1. $D(HAM_n^{(a)}) \ge \lg \sum_{m:F(a,n,m)\ne 0} \binom{n}{m}$.
2. $Q^*(HAM_n^{(a)}) \ge \frac{1}{2} \lg \sum_{m:F(a,n,m)\ne 0} \binom{n}{m}$.

3. $C^*(HAM_n^{(a)}) \geq \lg \sum_{m:F(a,n,m)\neq 0} \binom{n}{m}$.

Proof: By Lemma 3.4, the eigenvector v_z has a nonzero eigenvalue if v_z has m 1's and $F(a,n,m) \neq 0$. The rank of M_a is the number of nonzero eigenvalues that correspond to linearly independent eigenvectors. This is $\sum_{m:F(a,n,m)\neq 0} \binom{n}{m}$. The theorem follows from Lemmas 2.2 and 2.3. ▮

Lemma 3.7 *The number of values of m for which $F(a,n,m) = 0$ is $\leq a$.*

Proof: View the double summation $F(a,n,m)$ as a polynomial in m. The jth summand has degree $k + (j-k) = j$. Since $j \leq a$ the entire sum can be written as a polynomial in m of degree a. This has at most a roots. ▮

Theorem 3.8 *There is a constant $c > 0$ such that if $a \leq c\sqrt{n}$ then the following hold.*

1. $D(HAM_n^{(a)}) \geq n$.
2. $Q^*(HAM_n^{(a)}) \geq n/2$.
3. $C^*(HAM_n^{(a)}) \geq n$.

Proof: By Lemma 3.6 $D(f), Q^*(f) \geq \lg(\sum_{m:F(a,n,m)\neq 0} \binom{n}{m})$ and $C^*(f) \geq \frac{1}{2}\lg(\sum_{m:F(a,n,m)\neq 0} \binom{n}{m})$. Note that

$$2^n = \sum_{m:F(a,n,m)\neq 0} \binom{n}{m} + \sum_{m:F(a,n,m)=0} \binom{n}{m}.$$

By Lemma 3.7 $|\{m : F(a,n,m) = 0\}| \leq a$. Hence,

$$\sum_{m:F(a,n,m)=0} \binom{n}{m} \leq |\{m : F(a,n,m) = 0\}| \cdot \max_{0\leq m\leq n} \binom{n}{m} \leq a\binom{n}{n/2} \leq \frac{a2^n}{\sqrt{n}}.$$

So, if $a \leq \frac{1}{4}\sqrt{n}$, then

$$\sum_{m:F(a,n,m)\neq 0} \binom{n}{m} \geq 2^n - \frac{a2^n}{\sqrt{n}} \geq 2^n - 2^{n-2}.$$

Hence,

$$\lg\left(\sum_{m:F(a,n,m)\neq 0} \binom{n}{m}\right) \geq \lg(2^n - 2^{n-2}); \text{ i.e., } \left\lceil \lg\left(\sum_{m:F(a,n,m)\neq 0} \binom{n}{m}\right)\right\rceil \geq n.$$

▮

4 The Complexity of $HAM_n^{(=a)}$ for $a \le O(\sqrt{n})$

We again start by deducing results for general a, and then specialize to the case where $a \le c\sqrt{n}$.

Definition 4.1 Let $M_{=a}$ be $M_{HAM_n^{(=a)}}$, the $2^n \times 2^n$ matrix representing $HAM_n^{(=a)}$.

The vectors v_z are the same ones defined in Definition 3.3. We show that v_z is an eigenvector of M. Once that is done we let $eig(z)$ be the eigenvalue of M associated to z.

The lemmas needed, and the final theorem, are very similar (in fact easier) to those in the prior section. Hence we just state the needed lemmas and final theorem.

Lemma 4.2

1. *For all $z \in \{0,1\}^n$ v_z is an eigenvector of $M_{=a}$.*
2. *If z has exactly m 1's in it then*

$$eig(z) = \sum_{k=\max\{0,a+m-n\}}^{\min\{a,m\}} \binom{m}{k}\binom{n-m}{a-k}(-1)^k.$$

Definition 4.3

$$f(a,n,m) = \sum_{k=\max\{0,a+m-n\}}^{\min\{a,m\}} \binom{m}{k}\binom{n-m}{a-k}(-1)^k.$$

Note, from our convention that "if $x < y$, then $\binom{x}{y}$ is taken to be zero", that we can also write

$$f(a,n,m) = \sum_{k=0}^{a} \binom{m}{k}\binom{n-m}{a-k}(-1)^k.$$

The following lemma will be used in this section to obtain a lower bound when $a = O(\sqrt{n})$, and in Section 5 to obtain a lower bound for general a.

Lemma 4.4

1. $D(HAM_n^{(=a)}) \ge \lg \sum_{m:f(a,n,m)\neq 0} \binom{n}{m}$.
2. $Q^*(HAM_n^{(=a)}) \ge \lg \sum_{m:f(a,n,m)\neq 0} \binom{n}{m}$.
3. $C^*(HAM_n^{(=a)}) \ge \frac{1}{2} \cdot \lg \sum_{m:f(a,n,m)\neq 0} \binom{n}{m}$.

Lemma 4.5 *The number of values of m for which $f(a,n,m) = 0$ is $\le a$.*

Theorem 4.6 *There is a constant $c > 0$ such that if $a \le c\sqrt{n}$ then the following hold.*

1. $D(HAM_n^{(=a)}) \ge n$.
2. $Q^*(HAM_n^{(=a)}) \ge n/2$.
3. $C^*(HAM_n^{(=a)}) \ge n$.

5 The Complexity of $HAM_n^{(a)}$ and $HAM_n^{(=a)}$ for General a

We now consider the case of general a. As above, we will show that $F(a, m, n)$ and $f(a, m, n)$ are nonzero for many values of m. This will imply that the matrices M_a and $M_{=a}$ have high rank, hence $HAM_n^{(a)}$ and $HAM_n^{(=a)}$ have high communication complexity. We will use general generating-function methods to derive facts about these sums. A good source on generating functions is [13].

One of our main results will be Lemma 5.11, which states that if $0 \leq a \leq m < n$, then "$f(a, m, n) = 0$" implies "$f(a, m+1, n) \neq 0$". The idea behind our proof of Lemma 5.11 will be the following: we will show a relationship between the sum $f(a, m, n)$ and a certain new sum $h(a, m, n)$. Then we will derive generating functions for f and h, and translate this relationship into a relation between their generating functions. Finally, we will show that this relation cannot hold under the assumption that $f(a, m, n) = f(a, m + 1, n) = 0$, thus reaching a contradiction. Some auxiliary results needed for this are now developed in Section 5.1.

5.1 Auxiliary Notation and Results

Notation 5.1 $[x^b]g(x)$ is the coefficient of x^b in the power series expansion of $g(x)$ around $x_0 = 0$.

Notation 5.2 $t^{(i)}(x)$ is the i'th derivative of $t(x)$.

We will make use of the following lemma, which follows by an easy induction on i:

Lemma 5.3 *Let $t(x)$ be an infinitely differentiable function. Let $T_1(x) = (x - 1)t(x)$, and $T_2(x) = (x + 1)t(x)$. Then for any $i \geq 1$:*
$T_1^{(i)}(x) = (x - 1)t^{(i)} + i \cdot t^{(i-1)}(x)$
$T_2^{(i)}(x) = (x + 1)t^{(i)} + i \cdot t^{(i-1)}(x)$

For the rest of Section 5.1, the integers a, m, n are arbitrary subject to the constraint $0 \leq a \leq m \leq n$, unless specified otherwise.

Definition 5.4

1. $h(a, m, n) = \sum_{i=0}^{a} \binom{m}{i}\binom{n-m}{a-i}\frac{(-1)^i}{m-i+1}$.
2. $g(x) = \frac{x^{m+1}-(x-1)^{m+1}}{m+1} \cdot (x + 1)^{n-m}$.

We will show an interesting connection between h and f.

Claim 5.5 *Suppose $f(a, m, n) = 0$. Then $f(a, m + 1, n) = 0$ iff $h(a, m, n) = 0$.*

Proof:

$$\begin{aligned}
f(a, m+1, n) &= \sum_{i=0}^{a} \binom{m+1}{i}\binom{n-m-1}{a-i}(-1)^i \\
&= \frac{m+1}{n-m}\sum_{i=0}^{a} \binom{m}{i}\binom{n-m}{a-i}(-1)^i \cdot \frac{n-m-a+i}{m-i+1} \\
&= \frac{m+1}{n-m}\left((n+1-a)\sum_{i=0}^{a} \binom{m}{i}\binom{n-m}{a-i}\frac{(-1)^i}{m-i+1}\right) - \sum_{i=0}^{a} \binom{m}{i}\binom{n-m}{a-i}(-1)^i) \\
&= \frac{m+1}{n-m}((n+1-a)h(a, m, n) - f(a, m, n))
\end{aligned}$$

Thus, if $f(a, m, n) = 0$, then $f(a, m + 1, n) = 0$ iff $h(a, m, n) = 0$. ∎

We next show a connection between $g(x)$ and h.

Claim 5.6 $h(a, m, n) = (-1)^m \cdot [x^a]g(x)$.

Proof:

$$\begin{aligned} g(x) &= \frac{x^{m+1}-(x-1)^{m+1}}{m+1} \cdot (x+1)^{n-m} \\ &= \frac{x^{m+1}-\sum_{i=0}^{m+1}\binom{m+1}{i}x^i(-1)^{m+1-i}}{m+1} \cdot (x+1)^{n-m} \\ &= (-1)^m \sum_{i=0}^{m} \binom{m}{i} x^i \frac{(-1)^i}{m+1-i} \cdot (x+1)^{n-m} \\ &= (-1)^m \sum_{i=0}^{m} \binom{m}{i} x^i \frac{(-1)^i}{m+1-i} \cdot \sum_{j=0}^{n-m} \binom{n-m}{j} x^j \end{aligned}$$

Therefore, $h(a, m, n) = (-1)^m \cdot [x^a]g(x)$. ▮

Next, define an auxiliary function $\phi(u, v, w)$ as the w'th derivative of the function $(x+1)^u(x-1)^v$ evaluated at $x = 0$. We now relate ϕ and h.

Claim 5.7 $h(a, m, n) = 0$ *iff* $\phi(n - m, m + 1, a) = 0$.

Proof:
By Claim 5.6

$$\begin{aligned} h(a, m, n) &= (-1)^m \cdot [x^a]g(x) \\ &= \frac{(-1)^m}{m+1}([x^a](x^{m+1} \cdot (x+1)^{n-m}) - [x^a]((x-1)^{m+1} \cdot (x+1)^{n-m})). \end{aligned}$$

But $[x^a](x^{m+1} \cdot (x+1)^{n-m}) = 0$, since $a < m + 1$. So

$$\begin{aligned} h(a, m, n) &= \frac{(-1)^{m+1}}{m+1}[x^a]((x-1)^{m+1} \cdot (x+1)^{n-m}) \\ &= \frac{(-1)^{m+1}}{m+1} \cdot \frac{\phi(n-m,m+1,a)}{a!}. \end{aligned}$$

Thus, $h(a, m, n) = 0$ iff $\phi(n - m, m + 1, a) = 0$. ▮

Now we can relate the zeroes of f with those of ϕ:

Claim 5.8 $f(a, m, n) = 0$ *iff* $\phi(n - m, m, a) = 0$.

Proof:

$$\begin{aligned} (x-1)^m(x+1)^{n-m} &= \sum_{i=0}^{m} \binom{m}{i} x^i (-1)^{m-i} \cdot \sum_{j=0}^{n-m} \binom{n-m}{j} x^j \\ &= (-1)^m \sum_{i=0}^{m} \binom{m}{i} x^i (-1)^i \cdot \sum_{j=0}^{n-m} \binom{n-m}{j} x^j \\ &= (-1)^m \sum_{b=0}^{n} \sum_{k=0}^{b} \binom{m}{k}\binom{n-m}{b-k}(-1)^k x^b \\ &= (-1)^m \sum_{b=0}^{n} f(b, m, n) \cdot x^b. \end{aligned}$$

So $f(a, m, n) = \frac{(-1)^m}{a!} \cdot \phi(n - m, m, a)$, thus $f(a, m, n) = 0$ iff $\phi(n - m, m, a) = 0$. ▮

Claim 5.9 *Suppose* $m < n$ *and* $\phi(n - m, m, a) = 0$. *Then*

$$\phi(n - m - 1, m + 1, a) = 0 \textit{ iff } \phi(n - m, m + 1, a) = 0.$$

Proof: This claim follows from Claims 5.5, 5.7, and 5.8. ∎

We are now able to prove a recursive relation between values of ϕ:

Claim 5.10 *If $k > 0$, $a > 0$, and $\phi(k, m, a) = \phi(k, m, a-1) = 0$, then $\phi(k-1, m, a) = \phi(k-1, m, a-1) = 0$.*

Proof: Suppose $\phi(k, m, a) = \phi(k, m, a-1) = 0$. By Lemma 5.3,

$$\phi(k, m+1, a) = -\phi(k, m, a) + a \cdot \phi(k, m, a-1) = 0. \tag{2}$$

By Claim 5.9, since $\phi(k, m, a) = 0$, we know that

$$\phi(k-1, m+1, a) = 0 \text{ iff } \phi(k, m+1, a) = 0.$$

Now, (2) yields $\phi(k-1, m+1, a) = 0$. Applying Lemma 5.3 again, we obtain:

$$\begin{aligned} 0 = &\quad \phi(k-1, m+1, a) = -\phi(k-1, m, a) + a \cdot \phi(k-1, m, a-1); \\ 0 = &\quad \phi(k, m, a) = \phi(k-1, m, a) + a \cdot \phi(k-1, m, a-1) \end{aligned}$$

Solving the equations, we get

$$\phi(k-1, m, a) = \phi(k-1, m, a-1) = 0.$$

Thus the claim is proved. ∎

5.2 The main results

We are now ready to prove our main lemma.

Lemma 5.11 *Let $0 \leq a \leq m < n$, and suppose $f(a, m, n) = 0$. Then $f(a, m+1, n) \neq 0$.*

Proof: The lemma holds trivially for $a = 0$, since both $f(a, m, n)$ and $f(a, m+1, n)$ are nonzero if $a = 0$. So suppose $a \geq 1$. Suppose $f(a, m, n) = f(a, m+1, n) = 0$. Then by Claims 5.8 and 5.9, we know that

$$\phi(n-m, m, a) = \phi(n-m-1, m+1, a) = \phi(n-m, m+1, a) = 0.$$

By Lemma 5.3,

$$\phi(n-m, m+1, a) = -\phi(n-m, m, a) + a \cdot \phi(n-m, m, a-1),$$

i.e., $\phi(n-m, m, a-1) = 0$. Hence $\phi(n-m, m, a-1) = \phi(n-m, m, a) = 0$. Now, an iterative application of Claim 5.10 eventually yields $\phi(0, m, a) = \phi(0, m, a-1) = 0$. By definition, $\phi(0, m, a)$ is the a'th derivative of

$$(x-1)^m = \sum_{i=0}^{m} \binom{m}{i} x^i (-1)^{m-i}$$

evaluated at $x = 0$. But $m \geq a$, so this is clearly not zero. Thus we have reached a contradiction, and Lemma 5.11 is proved. ∎

Theorem 5.12 *For large enough n and all $0 \leq a \leq n$ the following hold.*

1. $D(HAM_n^{(=a)}) \geq n - 2.$

2. $Q^*(HAM_n^{(=a)}) \geq \frac{n}{2} - 1.$

3. $C^*(HAM_n^{(=a)}) \geq n - 2.$

Proof: By Lemma 4.4,

$$D(f), C^*(f) \geq \lg(\sum_{m:f(a,m,n)\neq 0} \binom{n}{m})$$

and

$$Q^*(f) \geq \frac{1}{2}\lg(\sum_{m:f(a,m,n)\neq 0} \binom{n}{m}).$$

First suppose $a \leq n/2$. We have

$$\sum_{m:f(a,m,n)\neq 0} \binom{n}{m} \geq \sum_{m\geq n/2:f(a,m,n)\neq 0} \binom{n}{m}. \tag{3}$$

Let us lower-bound the r.h.s. of (3). First of all, since the r.h.s. of (3) works in the regime where $m \geq n/2 \geq a$, Lemma 5.11 shows that no two consecutive values of m in this range satisfy the condition "$f(a,m,n) = 0$". Also, for $m \geq n/2$, $\binom{n}{m}$ is a non-increasing function of m. Thus, if we imagine an adversary whose task is to keep the r.h.s. of (3) as small as possible, the adversary's best strategy, in our regime where $m \geq n/2$, is to make $f(a,m,n) = 0$ exactly when $m \in S$, where

$$S \doteq \{\lceil n/2\rceil, \lceil n/2\rceil + 2, \lceil n/2\rceil + 4, \ldots\}. \tag{4}$$

Now,

$$2^{n-1} \leq \sum_{m\geq n/2} \binom{n}{m} \leq 2^{n-1} + O(2^n/\sqrt{n}). \tag{5}$$

(We need the second inequality to handle the case where n is even.) Also, recall that an $(1 - o(1))$ fraction of the sum $\sum_{m\geq n/2} \binom{n}{m}$ is obtained from the range $n/2 \leq m \leq n/2 + \sqrt{n \log n}$, for instance. (Here and in what follows, "$o(1)$" denotes a function of n that goes to zero as n increases.) In this range, the values of $\binom{n}{m}$ for any two consecutive values of m are within $(1 + o(1))$ of each other. In conjunction with (5), this shows that

$$\sum_{m\geq n/2:f(a,m,n)\neq 0} \binom{n}{m} \geq \sum_{m\geq n/2:m\notin S} \binom{n}{m} \geq (1/2 - o(1))2^{n-1}.$$

Thus,

$$\left\lceil \lg\left(\sum_{m\geq n/2:f(a,m,n)\neq 0} \binom{n}{m}\right)\right\rceil \geq n - 2,$$

completing the proof for the case where $a \leq n/2$.

Now we apply symmetry to the case $a > n/2$: note that Alice can reduce the problem with parameter a to the problem with parameter $n - a$, simply by complementing each bit of her input x. Thus, the same communication complexity results hold for the case $a > n/2$. ∎

Lemma 5.13 *Let $0 \le a < m < n$, and suppose $F(a,m,n) = 0$. Then $F(a,m+1,n) \neq 0$.*

Proof: We have $f(j,m,n) = (-1)^m [x^j]((x-1)^m(x+1)^{n-m})$. By definition,

$$\begin{aligned} F(a,m,n) &= \sum_{j=0}^{a} f(j,m,n) \\ &= (-1)^m \sum_{j=0}^{a} [x^j]((x-1)^m(x+1)^{n-m}) \\ &= (-1)^m [x^a]((x-1)^m(x+1)^{n-m} \cdot \sum_{j=0}^{\infty} x^j) \\ &= (-1)^m [x^a]((x-1)^m(x+1)^{n-m} \cdot \tfrac{1}{1-x}) \\ &= (-1)^{m-1}[x^a]((x-1)^{m-1}(x+1)^{n-m}) = f(a,m-1,n-1). \end{aligned}$$

So $F(a,m,n) = F(a,m+1,n) = 0$ iff $f(a,m-1,n-1) = f(a,m,n-1) = 0$. But the latter is impossible by Lemma 5.11, thus the lemma is proved. ∎

Theorem 5.14 *For large enough n and all $0 \le a \le n-1$, the following hold.*

1. $D(HAM_n^{(a)}) \ge n-2$.
2. $Q^*(HAM_n^{(a)}) \ge \frac{n}{2} - 1$.
3. $C^*(HAM_n^{(a)}) \ge n-2$.

Proof: The proof is identical to that of Theorem 5.12 except for one point. In that proof we obtained the $a > n/2$ case easily from the $a \le n/2$ case. Here it is also easy but needs a different proof. Let $a > n/2$ and, for all $x \in \{0,1\}^n$, let $\overline{x}$ be obtained from x by flipping every single bit. Note that

$HAM_n^{(a)}(x,y) = 1$ iff $\mathrm{HAM}(x,y) \le a$ iff $\mathrm{HAM}(\overline{x},y) \ge n-a$ iff $\mathrm{NOT}(\mathrm{HAM}(\overline{x},y) \le (n-a)-1$ iff $\mathrm{HAM}_{n-a-1}(\overline{x},y) = 1$.

Since $n-a-1 \le n/2$ we have that a lower bound for the $a \le n/2$ case implies a lower bound for the $a > n/2$ case. ∎

6 Open Problems

We make the following conjectures.

1. For all n, for all a, $0 \le a \le n-1$, $D(HAM_n^{(a)}), C^*(HAM_n^{(a)}), Q^*(HAM_n^{(a)}) \ge n+1$.
2. For all n, for all a, $0 \le a \le n$, $D(HAM_n^{(=a)}), C^*(HAM_n^{(=a)}), Q^*(HAM_n^{(=a)}) \ge n+1$.

References

[1] K. Abdel-Ghaffar and A. E. Ababdi. An optimal strategy for comparing file copies. *IEEE Transactions on Parallel and Distributed Systems*, 5:87–93, 1994.

[2] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proc. of the 16th IEEE Conf on Complexity Theory.* IEEE Computer Society Press, 2001.

[3] G. Cormode, M. Paterson, S. Sahinalp, and U. Vishkin. Communication complexity of document exchange. In *Proc. of the 11th ACM Sym. on Discrete Algorithms*, 2000.

[4] R. de Wolf. Quantum communication and complexity. *Theoretical Comput. Sci.*, 12:337–353, 2002.

[5] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright. Secure multiparty computation of approximations. In *Proc. of the 28th ICALP (LNCS 2076)*, volume 2076 of *Lecture Notes in Computer Science*, pages 927–938, Berlin, 2001. Springer-Verlag.

[6] D. Gavinsky, J. Kempe, and R. de Wolf. Quantum communication cannot simulate a public coin, 2004. arxiv.org/abs/quant-ph/0411051.

[7] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[8] K. Mehlhorn and E. Schmidt. Las Vegas is better than determinism for VLSI and distributed systems. In *Proc. of the 14th ACM Sym. on Theory of Computing*, pages 330–337, 1982.

[9] J. Metzner. Efficient replicated remote file comparison. *IEEE Transactions on Computers*, 40:651–659, 1991.

[10] I. Newman. Private vs. common random bits in communication complexity. *Inf. Process. Lett.*, 39:67–71, 1991.

[11] A. Orlitsky. Interactive communication: balanced distributions, correlated files, and average-case complexity. In *Proc. of the 32st IEEE Sym. on Found. of Comp. Sci.*, pages 228–238, 1991.

[12] K. Pang and A. E. Gamal. Communication complexity of computing the Hamming distance. *SIAM Journal of Computing*, 15, 1986.

[13] H. Wilf. *Generatingfunctionology*. Academic Press, 1994.

[14] A. Yao. On the power of quantum fingerprinting. In *Proc. of the 35th ACM Sym. on Theory of Computing*, 2003.

ECCC ISSN 1433-8092
http://www.eccc.uni-trier.de/eccc
ftp://ftp.eccc.uni-trier.de/pub/eccc
ftpmail@ftp.eccc.uni-trier.de, subject 'help eccc'