Lower bounds on the Deterministic and Quantum Communication Complexity of Hamming-Distance Problems
by Andris Ambainis, William Gasarch, Aravind Srinivasan, Andrey Utis

Abstract

Alice and Bob want to know if two strings of length n are almost equal. That is, do the strings differ on *at most* a bits? Let $0 \leq a \leq n-1$. We show (1) any deterministic protocol – as well as any error-free quantum protocol (C^* version) – for this problem requires at least $n-2$ bits of communication, and (2) a lower bound of $n/2-1$ for error-free Q^* quantum protocols. We also show the same results for determining if two strings differ in *exactly* a bits. 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 $\text{HAM}(x,y)$ 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 $\text{HAM}(x,y)$? The trivial algorithm is to have Alice send x (which takes n bits) and have Bob send $\text{HAM}(x,y)$ (which takes $\lceil \lg(n+1) \rceil$ bits) back to Alice. This takes $n + \lceil \lg(n+1) \rceil$ bits. Pang and El Gamal [15] 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, 5, 13, 14] for more on the communication complexity of HAM. See [7] for how Alice and Bob can approximate HAM without giving away too much information.)

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

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

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

The communication complexity of $HAM_n^{(a)}$ has been studied in various randomized and quantum settings by Yao [17], Gavinsky et al. [8] (Section 6), Gavinsky et al. [9] (Section 3.2), and Huang et al. [10].

How much communication is needed for this problem in the deterministic model? There is the trivial $(n+1)$-bit upper bound. There is an easy reduction from equality on $n-a$ bits to $HAM_n^{(a)}$, hence there is an easy $(n-a)$ lower bound. In this paper we improve the lower bound. Note that this amounts to improving the additive term.

We show the following:

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

4. For $a \leq \frac{\sqrt{n}}{4}$, $HAM_n^{(a)}$ 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 where Alice and Bob share an infinite number of EPR pairs, using a quantum channel, and always obtain the correct answer.

6. For $a \leq \frac{\sqrt{n}}{4}$, $HAM_n^{(a)}$ requires at least $\frac{n}{2}$ 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

$$HAM_n^{(=a)}(x, y) = \begin{cases} 1 & \text{if } \mathrm{HAM}(x, y) = a; \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

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 [11] for more details on classical, and [6] 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. For quantum protocols, we fix the number of qubits communicated in each round (assuming that in the first round Alice always communicates c_1 qubits, in the second round Bob communicates c_2 qubits and so on, where $c_1, c_2, \ldots$ are independent of inputs x and y).

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 [12]; see also [11].

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

We will need the following definition and notation

Definition 2.4 The *Krawtchouk Polynomials* (see [4] and the references therein) are polynomials that are parameterized by $a, n, q \in \mathsf{N}$ with q a prime power and are defined by

$$k_a(n,q;x) = \sum_{k=0}^{a} (-1)^k (q-1)^{a-k} \binom{x}{k} \binom{n-x}{a-k}$$

(In the paper [4] they use N instead of n and order the variables as $k_a(x,q,N)$.)

Definition 2.5 Let

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

Definition 2.6

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

(The equality comes from our convention: if $a < b$, then $\binom{a}{b}$ is taken to be zero.) Note that $G(a,n;x) = k_a(n,2;x)$.

2.1 Lemmas Useful for the Complexity of $HAM_n^{(a)}$

Definition 2.7 Let M_a be $M_{HAM_n^{(a)}}$, the $2^n \times 2^n$ matrix representing $HAM_n^{(a)}$.

Lemma 2.8 *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 2.9 Let $z \in \{0,1\}^n$.

1. $v_z \in \mathsf{R}^{2^n}$ 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 2.10

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 $eig(z) = F(a,n;m)$*

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:

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

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 this to equal $b \times v_z(x)$ for some constant b. We set it equal to $b \times v_z(x)$ and deduce which b's work. Suppose

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

We have

$$\begin{aligned}
b &= \frac{1}{v_z(x)} \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \\
&= v_z(x) \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \\
&= (-1)^{\sum_i x_i z_i} \sum_{y \in L_a} (-1)^{\sum_i y_i z_i} \quad \text{(by the definition of } v_z(x)) \\
&= \sum_{y \in L_a} (-1)^{\sum_i (x_i + y_i) z_i} \\
&= \sum_{y \in L_a} (-1)^{\sum_i |x_i - y_i| z_i} \quad (\text{since } x_i + y_i \equiv |x_i - y_i| \pmod 2) \\
&= \sum_{j=0}^{a} \sum_{y \in E_j} (-1)^{\sum_i |x_i - y_i| z_i} \quad (\text{since } L_a = \textstyle\bigcup_{j=0}^{a} E_j). \qquad (3)
\end{aligned}$$

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 $k \leq m$. Since there are exactly j places where $x_i \neq y_i$ we have $k \leq j$. Hence $k \leq \min\{j, m\}$.

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

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

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

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 (3),

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

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 2.11 Let

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

Lemma 2.12

1. $D(HAM_n^{(a)}) \geq \lg \sum_{m:F(a,n;m)\neq 0} \binom{n}{m}$.
2. $Q^*(HAM_n^{(a)}) \geq \frac{1}{2} \lg \sum_{m:F(a,n;m)\neq 0} \binom{n}{m}$.
3. $C^*(HAM_n^{(a)}) \geq \lg \sum_{m:F(a,n;m)\neq 0} \binom{n}{m}$.

Proof: By Lemma 2.10, 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 2.13 *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. We first show that $F(a,n;m)$ is not identically zero. Plug in $m = n$. Then

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

Since $0 \leq a < n$ this cannot be 0.

We now show that $F(a,n;m)$ has degree a and hence has at most a roots. 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. ∎

2.2 Lemmas Useful for the Complexity of $HAM_n^{(=a)}$

Definition 2.14 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 2.9. We show that v_z is an eigenvector of M. Once that is done we let $eig(z)$ be the eigenvalue of $M_{=a}$ associated to z.

The lemmas needed, and the final theorem, are very similar (in fact easier) to those in the Section 2.1. Hence we just state the needed lemmas and final theorem.

Lemma 2.15

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 $eig(z) = G(a, m; n)$.*

Lemma 2.16

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

3 The Complexity of $HAM_n^{(a)}$ and $HAM_n^{(=a)}$ for $a \leq \frac{\sqrt{n}}{4}$

Theorem 3.1 *If $a \leq \frac{\sqrt{n}}{4}$ 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 2.12 $D(f), Q^*(f) \geq \lg(\sum_{m:F(a,n;m)\neq 0} \binom{n}{m})$ 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 2.13 $|\{m : F(a, n; m) = 0\}| \leq a$. 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,

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

Therefore we have out lower bounds. ∎

The following theorem has a proof that is very similar to the proof of Theorem 3.1; hence we omit it.

Theorem 3.2 *If $a \leq \frac{\sqrt{n}}{4}$ then the following hold.*

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

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

Recall that $G(a, n; x)$ is the Krawtchouk polynomial $k_a(n, 2; x)$.

Lemma 4.1 *For all a, n let $r^n_{a,1} < r^n_{a,2} < \cdots < r^n_{a,a}$ be the roots of the poly $k_a(n, q; x)$. (They need not be integers.)*

1. *For all i there is an integer in the open interval $(r^n_{a,i}, r^n_{a,i+1})$.*
2. *Let m be an integer. If $k_a(n, q; m) = 0$ then $k_a(n, q; m + 1) \neq 0$.*
3. *Let m be an integer. If $G(a, n; m) = 0$ then $G(a, n; m + 1) \neq 0$.*

Proof:
1) This is from [16].
2) Assume, by way of contradiction, that there is an integer such that $k_a(n, q; m) = 0$ and $k_a(n, q; m + 1) = 0$. By part 1 there is an integer in the open interval $(m, m + 1)$. This is a contradiction.
3) This follows from the fact that $G(a, n; x) = k_a(n, 2; x)$. ∎

Theorem 4.2 *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: First suppose $a \leq n/2$. Note that

$$\sum_{m:G(a,n;m)\neq 0} \binom{n}{m} \geq \sum_{m\geq n/2:G(a,n;m)\neq 0} \binom{n}{m}. \tag{4}$$

Lemma 4.1 shows that no two consecutive values of m in the range $a \leq m \leq n$ (and hence in the range $n/2 \leq m \leq n$) satisfy the condition "$G(a,n;m) = 0$". Hence our problem is to minimize the sum of a subset of

$$\left\{ \binom{n}{n/2}, \binom{n}{n/2-1}, \ldots, \binom{n}{0} \right\},$$

where if we omit $\binom{n}{i}$, we must use $\binom{n}{i-1}$. Since $\binom{n}{m}$ decreases in the range $n/2 \leq m \leq n$, this sum is minimized by taking every other term: thus this sum is always at least 2^{n-2}. Our theorem follows from Lemma 2.16.

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 4.3 *Let $0 \leq a < m < n$, and suppose $F(a,n;m) = 0$. Then $F(a,m+1;n) \neq 0$.*

Proof: We will use the terminology and methods of generating functions.
Notation $[x^b]g(x)$ is the coefficient of x^b in the power series expansion of $g(x)$ around $x_0 = 0$.

Lemma 4.4

1. *If $a \in \mathsf{N}$ and $f(x)$ is any power series then*

$$\sum_{j=0}^{a} [x^j]f(x) = [x^a](f(x)\sum_{j=0}^{\infty} x^j) = [x^a]\frac{f(x)}{1-x}.$$

2.
$$G(j,n;m) = (-1)^m[x^j]((x-1)^m(x+1)^{n-m}).$$

3.
$$F(a,n;m) = \sum_{j=0}^{a} G(j,n;m).$$

Proof: Items 1 and 3 are clear. We prove item 2. We show $G(a,n;m) = (-1)^m[x^a]((x-1)^m(x+1)^{n-m})$ for ease of notation; however, the proof clearly holds for j instead of a.

$$(x-1)^m(x+1)^{n-m} = \sum_{k=0}^{m} \binom{m}{k} x^k (-1)^{m-k} \sum_{j=0}^{n-m} \binom{n-m}{j} x^j$$

$$(x-1)^m(x+1)^{n-m} = \sum_{k=0}^{m} \sum_{j=0}^{n-m} \binom{m}{k} x^k (-1)^{m-k} \binom{n-m}{j} x^j$$

$$(x-1)^m(x+1)^{n-m} = \sum_{k=0}^{m}\sum_{j=0}^{n-m}\binom{m}{k}\binom{n-m}{j}x^{k+j}(-1)^{m-k}$$

The coefficient of x^a is
$\sum_{k=0}^{a}\binom{m}{k}\binom{n-m}{a-k}(-1)^{m-k}$ which is $(-1)^m G(a,n;m)$. ∎

Using Lemma 4.4 we obtain the following.

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

Hence $F(a,n;m) = F(a,n;m+1) = 0$ iff $G(a,n-1;m-1) = G(a,n-1;m) = 0$. But the latter is impossible by Lemma 4.1, thus the lemma is proved. ∎

Theorem 4.5 *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 4.2 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 $\mathrm{HAM}_n^{n-a-1}(\overline{x},y) = 1$.

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. ∎

5 Open Problems

We make the following conjectures.

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

The first and third conjecture are just a matter of improving the lower bound by 3 bits. For the second and fourth conjecture, superdense coding [2] provides an upper bound of $\frac{n}{2}+1$ on $Q^*(HAM_n^{(a)})$ and on $Q^*(HAM_n^{(=a)})$ ($\frac{n}{2}$ qubits for Alice to communicate her input x to Bob and 1 bit for Bob to communicate the function value $f(x,y)$ back to Alice). The remaining part is to improve the lower bound by 2 qubits.

6 Acknowledgement

An earlier version of this work appeared in the *Proc. International Symposium on Algorithms and Computation* (ISAAC), 2006; We would like to thank an anonymous referee of the conference version who pointed out the connection to Krawtchouk polynomials. We also thank the journal referees for their helpful suggestions. This work is supported by the National Science Foundation, under grants CCR-01-05413, CCR-02-08005, CCF 14-22569, CNS-1010789, and CCF-1422569.

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] C. Bennett and S. Wiesner. Communication via one- and two-particle operators on einstein-podolsky-rosen states. *Physics Review Letters*, 69:2881–2884, 1992.

[3] H. Buhrman and R. de Wolf. Communication complexity lower bounds by polynomials. In *Proceedings of the 16th IEEE Conference on Complexity Theory,* Chicago IL, pages 120–130, Los Alamitos, CA, 2001. IEEE Computer Society Press. `arxiv.org/abs/cs/9910010`.

[4] L. Chihara and D. Stanton. Zeros of generalized Krawtchouk polynomials. *Journal of Approximation Theory*, 60(1):43–57, 1990. `http://www.sciencedirect.com/science/article/pii/002190459090072X`.

[5] G. Cormode, M. Paterson, S. Sahinalp, and U. Vishkin. Communication complexity of document exchange. In *Eleventh Symposium on Discrete Algorithms: Proceedings of SODA '00*, pages 197–206, New York, NY, USA, 2000. ACM-SIAM. `http://www.research.att.com/people` (Look for Graham Cormode).

[6] R. de Wolf. Quantum communication and complexity. *Theoretical Computer Science*, 12:337–353, 2002. `homepages.cwi.nl/~rdewolf/#PublQC`.

[7] J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright. Secure multiparty computation of approximations. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming ICALP 2001,* Crete, Greece, volume 2076 of *Lecture Notes in Computer Science*, pages 927–938, New York, Heidelberg, Berlin, 2001. Springer-Verlag. `http://www.springerlink.com`.

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

[9] D. Gavinsky, J. Kempe, and R. de Wolf. Strengths and weaknesses of quantum fingerprinting. In *Twenty-First Conference on Computational Complexity: Proceedings of CCC '06*, pages 288–298, Los Alamitos, CA, USA, 2006. IEEE Computer Society Press. `arxiv.org/abs/quant-ph/0603173`.

[10] W. Huang, Y. Shi, S. Zhang, and Y. Zhu. The communication complexity of the Hamming distance problem. *Information Processing Letters*, 99:149–153, 2006. `http://dl.acm.org/citation.cfm?id=1161712`.

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

[12] K. Mehlhorn and E. Schmidt. Las Vegas is better than determinism for VLSI and distributed systems. In *Proceedings of the Fourteenth Annual ACM Symposium on the Theory of Computing,* San Francisco CA, pages 330–337, New York, NY, USA, 1982. ACM Press.

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

[14] A. Orlitsky. Interactive communication: balanced distributions, correlated files. *SIAM Journal of Discrete Mathematics*, 6(4):548–554, 1993. `http://www.fleece.ucsd.edu/~alon`.

[15] K. Pang and A. E. Gamal. Communication complexity of computing the Hamming distance. *SIAM Journal on Computing*, 15:932–947, 1986. `http://www-isl.stanford.edu/~abbas/aegpublications.php`.

[16] G. Szego. *Orthogonal polynomials (fourth edition)*, volume 23 of *Colloquium Publications of the American Math Society.* American Math Society, Providence, 1975.

[17] A. Yao. On the power of quantum fingerprinting. In *Proceedings of the Thirty-fifth Annual ACM Symposium on the Theory of Computing,* San Diego CA, pages 77–81, New York, NY, USA, 2003. ACM Press. `https:///www.cs.princeton.edu/courses/archive/spring04/cos598A/Stoc.pdf`.