On the Hamming Distance Between Two i.i.d. Random n-Tuples over a Finite Set

Fang-Wei Fu, Torleiv Kløve, *Senior Member, IEEE,* and Shi-Yi Shen

***Abstract*— In this paper, we study the Hamming distance $d_{\mathrm{H}}(X, Y)$ between two independent identical distributed (i.i.d.) random n-tuples X and Y over some finite set. Both lower and upper bounds are derived for the expectation $Ed_{\mathrm{H}}(X, Y)$ and the variance $Dd_{\mathrm{H}}(X, Y)$. Also, a generalization of the Grey–Rankin bound is given.**

***Index Terms*— Althöfer–Sillke inequality, codes, expectation, Grey–Rankin bound, Hamming distance, random n-tuple, variance.**

I. Introduction

Let V be some finite set with q elements. For $n \geq 1$, V^n is the set of ordered n-tuples from V. The Hamming distance $d_{\mathrm{H}}(\mathbf{x}, \mathbf{y})$ between two n-tuples $\mathbf{x}, \mathbf{y} \in V^n$ is the number of positions in which $\mathbf{x}$ and $\mathbf{y}$ differ.

Let X and Y be two independent identical distributed (i.i.d.) random n-tuples, and let $P = \{P(\mathbf{x}) | \mathbf{x} \in V^n\}$ be the common probability distribution. The *expectation* of $d_{\mathrm{H}}(X, Y)$ is defined by

$$Ed_{\mathrm{H}}(X, Y) = \sum_{\mathbf{x}, \mathbf{y} \in V^n} P(\mathbf{x}) P(\mathbf{y}) d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}).$$

The *variance* of $d_{\mathrm{H}}(X, Y)$ is defined by

$$Dd_{\mathrm{H}}(X, Y) = E[d_{\mathrm{H}}(X, Y)]^2 - [Ed_{\mathrm{H}}(X, Y)]^2.$$

Let

$$L(P) = q^{n-1} \sum_{\mathbf{x} \in V^n} \left[P(\mathbf{x}) - \frac{1}{q^n} \right]^2$$

where $L(P)$ measures how skewly P is distributed. Note that

$$\begin{aligned} L(P) &= q^{n-1} \sum_{\mathbf{x} \in V^n} P(\mathbf{x})^2 - 2 \frac{q^{n-1}}{q^n} \sum_{\mathbf{x} \in V^n} P(\mathbf{x}) + \frac{q^{n-1} \cdot q^n}{q^{2n}} \\ &= q^{n-1} \sum_{\mathbf{x} \in V^n} P(\mathbf{x})^2 - \frac{2}{q} \cdot 1 + \frac{1}{q}. \end{aligned}$$

Hence

$$qL(P) = -1 + q^n \sum_{\mathbf{x} \in V^n} P(\mathbf{x})^2. \tag{1}$$

This paper is organized as follows. In Section II, we state a number of bounds on the expectation and variance. In Section III, we give applications to coding theory of these results. Furthermore, we give a generalization of the Grey–Rankin bound. Our proofs use group algebras. In Section IV, we quote some basic results on group algebras, in particular, the MacWilliams identity. In Section V, we derive several further lemmas needed to establish our results. In Section VI, we present proofs of the theorems. Finally, in Section VII, we give some examples which illustrate our results.

Manuscript received December 21, 1997; revised September 24, 1998. This work was supported in part by the University Doctor Foundation of China, National Natural Science Foundation of China, and Norwegian Research Council.

F.-W. Fu and S.-Y. Shen are with the Department of Mathematics, Nankai University, Tianjin 300071, China.

T. Kløve is with the Department of Informatics, University of Bergen, N-5020 Bergen, Norway.

Communicated by A. Barg, Associate Editor for Coding Theory.
Publisher Item Identifier S 0018-9448(99)01414-5.

II. Results on Expectation and Variance

Our main results are given in the following theorems.

Theorem 1: We have

$$\frac{n(q-1)}{q} - L(P) \leq Ed_{\mathrm{H}}(X, Y) \leq \frac{n(q-1)}{q}.$$

Theorem 2: If $L(P) \leq (q-2)/q$, then

$$Dd_{\mathrm{H}}(X, Y) \geq \frac{n(q-1)}{q^2}$$

if $L(P) \geq (q-2)/q$, then

$$Dd_{\mathrm{H}}(X, Y) \geq \frac{n(q-1)}{q^2} + \frac{q-2}{q} L(P) - L(P)^2.$$

Theorem 3: If $2 \leq q \leq 4$, then

$$Dd_{\mathrm{H}}(X, Y) \leq \frac{n(q-1)}{q^2} + \frac{2}{q} L(P)$$

if $q > 4$ and $L(P) \leq (q-4)/2q$, then

$$Dd_{\mathrm{H}}(X, Y) \leq \frac{n(q-1)}{q^2} + \frac{q-2}{q} L(P) - L(P)^2$$

if $q > 4$ and $L(P) \geq (q-4)/2q$, then

$$Dd_{\mathrm{H}}(X, Y) \leq \frac{n(q-1)}{q^2} + \left(\frac{q-4}{2q} \right)^2 + \frac{2}{q} L(P).$$

For $q = 2$, the bounds in Theorem 1 coincide with those given by Althöfer and Sillke [3]. They proved the lower bound by induction and the upper bound by a simple direct argument. In this paper, we present an alternative proof of Theorem 1, valid for all q, using the MacWilliams identity for the elements of a group algebra.

Since $L(P) \geq 0$, Theorem 2 in particular implies that for $q = 2$ we have

$$Dd_{\mathrm{H}}(X, Y) \geq \frac{n}{4} - L(P)^2$$

a result that was first given in [4].

Our next theorems consider the situation when equality is obtained in one of the bounds in Theorem 1.

Theorem 4: If $Ed_{\mathrm{H}}(X, Y) = n(q-1)/q$, then

$$\frac{n(q-1)}{q^2} \leq Dd_{\mathrm{H}}(X, Y) \leq \frac{n(q-1)}{q^2} + \frac{2}{q} L(P)$$

and

$$\sum_{\substack{\mathbf{x} \in V^n \\ x_i = j}} P(\mathbf{x}) = \frac{1}{q} \quad \text{for all } i \text{ and } j.$$

Theorem 5: If $Ed_{\mathrm{H}}(X, Y) = (n(q-1)/q) - L(P)$, then

$$Dd_{\mathrm{H}}(X, Y) = \frac{n(q-1)}{q^2} + \frac{q-2}{q} L(P) - L(P)^2.$$

III. Applications to Coding Theory

An $(n, M; q)$ code C is subset of V^n of size M. The average Hamming distance in C is defined by

$$\overline{d}_{\mathrm{H}}(C) = \frac{1}{M^2} \sum_{\mathbf{x}, \mathbf{y} \in C} d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}).$$

Let

$$\alpha_q(n, M) = \min\{\overline{d}_{\mathrm{H}}(C) | C \text{ is an } (n, M; q) \text{ code}\}.$$

Ahlswede and Althöfer [2] studied the asymptotic behavior of $\alpha_2(n, M)$. The variance of the Hamming distance is defined by

$$\mathrm{var}_{\mathrm{H}}(C) = \frac{1}{M^2} \sum_{\mathbf{x}, \mathbf{y} \in C} [d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) - \overline{d}_{\mathrm{H}}(C)]^2.$$

The question to determine

$$\beta_q(n, M) = \min\{\mathrm{var}_{\mathrm{H}}(C) | C \text{ is an } (n, M; q) \text{ code}\}$$

was raised by Fu and Shen [4]. It seems very difficult to determine $\alpha_q(n, M)$ and $\beta_q(n, M)$ in general. Until now, only a few results have been obtained for the binary case (see [2], [3], and [8]).

Let C be an $(n, M; q)$ code. It is easy to see that

$$\overline{d}_{\mathrm{H}}(C) = E d_{\mathrm{H}}(X_C, Y_C) \quad \text{and} \quad \mathrm{var}_{\mathrm{H}}(C) = D d_{\mathrm{H}}(X_C, Y_C)$$

where X_C, Y_C are two i.i.d. random n-tuple with the common distribution

$$P_C(\mathbf{x}) = \begin{cases} \dfrac{1}{M}, & \mathbf{x} \in C \\ 0, & \mathbf{x} \notin C. \end{cases}$$

Furthermore

$$L(P_C) = \frac{q^{n-1}}{M} - \frac{1}{q}.$$

Therefore, Theorems 1–3 give bounds on $\overline{d}_{\mathrm{H}}(C)$ and $\mathrm{var}_{\mathrm{H}}(C)$. For $q = 2$, these coincide with the bounds given in [3] [for $\overline{d}_{\mathrm{H}}(C)$] and [4] [for $\mathrm{var}_{\mathrm{H}}(C)$].

Our method of proof can also be used to give a generalization of the Grey–Rankin bound to nonbinary codes (for the Grey–Rankin bound for binary codes, see, e.g., [6, p. 544]).

Theorem 6: Let C be an $(n, M, d; q)$ code where

$$d \le d_u = \frac{2n(q-1) - (q-2)}{2q} \tag{2}$$

$$d > d_l = \frac{2n(q-1) - (q-2) - \sqrt{(q-2)^2 + 4n(q-1)}}{2q} \tag{3}$$

and

$$d \le d_{\mathrm{H}}(\mathbf{a}, \mathbf{b}) \le \frac{2n(q-1) - (q-2) - qd}{q} \tag{4}$$

for every codeword pair $\mathbf{a}, \mathbf{b} \in C$, $\mathbf{a} \neq \mathbf{b}$. Then

$$M \le \frac{qd(2n(q-1) - (q-2) - qd)}{qd(2n(q-1) - (q-2) - qd) - (q-1)^2 n(n-1)}.$$

Remarks:

1) The condition (4) implies (2), and under this condition, (3) holds if and only if the denominator of the generalized Grey–Rankin bound is positive.
2) Levenshtein [5, p. 90] gave general bounds for packings, and it may be of interest to compare our bound to these bounds. For the d satisfying (2) and (3), this bound on M is

$$M \le \frac{qd((n(q-1)+1)(n(q-1) - qd + 2) - q)}{qd(2n(q-1) - (q-2) - qd) - (q-1)^2 n(n-1)}.$$

We note that this bound has the same denominator as the generalized Grey–Rankin bound. It is not hard to show that the Levenshtein bound is weaker then the generalized Grey–Rankin bound, which is not unexpected because of the additional condition (4) in Theorem 6. We illustrate this with a numerical example: $n = 100$ and $q = 3$. In this case, $d_l \approx 61.8 < d \le d_u = 66.5$. In this range, the bounds have the following values:

d	62	63	64	65	66
Levenshtein	33201	5481	2676	1521	801
Gen. Grey–Rankin	2201	441	276	221	201

IV. Basics on Group Algebras

We note that the mean and variance only depend on P, not on any properties of the set V. Therefore, without loss of generality, we will from now on assume that $V = \mathbb{Z}_q$, that is, the additive group of integers modulo q.

In this section, we quote some known definitions and properties of the group algebra over $\mathbb{Z}_q$. For a more detailed account of group algebras, we refer the reader to the book of Roman [7, pp. 220–237].

In the next section, we give some lemmas which connect this theory with our main topic of expectation and variance, and in the following section we prove the theorems.

Let t be an independent variable, and let $\mathcal{S}$ be the set of all formal sums of the form

$$g = \sum_{\mathbf{x} \in \mathbb{Z}_q^n} \lambda(\mathbf{x}) t^{\mathbf{x}} \tag{5}$$

where the $\lambda(\mathbf{x})$ are complex numbers.

The set $\mathcal{S}$ can be made into an algebra (known as a group algebra) over the complex numbers by defining addition, scalar multiplication, and multiplication in the natural way.

Let $w_{\mathrm{H}}(\mathbf{a})$ be the Hamming weight of the n-tuple $\mathbf{a} \in \mathbb{Z}_q^n$, i.e., the number of nonzero elements.

For an element g, given by (5), its *weight distribution* is given as follows:

$$A_i(g) = \sum_{\substack{\mathbf{x} \in \mathbb{Z}_q^n \\ w_{\mathrm{H}}(\mathbf{x}) = i}} \lambda(\mathbf{x}), \qquad i = 0, 1, \cdots, n.$$

The *weight enumerator* of g is defined by

$$W_g(s) = \sum_{\mathbf{x} \in \mathbb{Z}_q^n} \lambda(\mathbf{x}) s^{w_{\mathrm{H}}(\mathbf{x})} = \sum_{i=0}^{n} A_i(g) s^i.$$

Let ω be a primitive qth root of unity (e.g., $\omega = e^{2\pi i/q}$), and, for a given $g \in \mathcal{S}$, let

$$\hat{\lambda}(\mathbf{y}) = \sum_{\mathbf{x} \in \mathbb{Z}_q^n} \lambda(\mathbf{x}) \omega^{\langle \mathbf{x}, \mathbf{y} \rangle}$$

where

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{j=1}^{n} x_j y_j, \qquad \text{the scalar product.}$$

The *MacWilliams transform* of g is then defined by

$$\hat{g} = \sum_{\mathbf{y} \in \mathbb{Z}_q^n} \hat{\lambda}(\mathbf{y}) t^{\mathbf{y}}.$$

The (generalized) MacWilliams identity connects $W_g(s)$ and $W_{\hat{g}}(s)$

$$W_g(s) = \frac{1}{q^n} [1 + (q-1)s]^n W_{\hat{g}}\left(\frac{1-s}{1+(q-1)s}\right). \tag{6}$$

From (6), we easily get the following (generalized) Pless identities:

$$q^n A_0(g) = \sum_{i=0}^{n} A_i(\hat{g}) \tag{7}$$

$$\sum_{i=0}^{n} A_i(g) = A_0(\hat{g}) \tag{8}$$

$$q\sum_{i=0}^{n} iA_i(g) = n(q-1)A_0(\hat{g}) - A_1(\hat{g}) \tag{9}$$

$$q^2\sum_{i=0}^{n} i^2A_i(g) = n(q-1)(nq-n+1)A_0(\hat{g}) - (2(n-1)(q-1)+q)A_1(\hat{g}) + 2A_2(\hat{g}). \tag{10}$$

These are standard results—we sketch the proofs. For $x \in \mathbb{Z}_q$, let

$$Q(x) = \sum_{y\in\mathbb{Z}_q} \omega^{xy}(1-s)^{w_{\mathrm{H}}(x)}(1+(q-1)s)^{1-w_{\mathrm{H}}(x)}.$$

Then

$$Q(x) = [1+(q-1)s] + (1-s)\sum_{y=1}^{q-1}\omega^{xy} = \begin{cases} q, & \text{if } x=0 \\ qs, & \text{if } x\neq 0. \end{cases}$$

Therefore

$$\begin{aligned}
&[1+(q-1)s]^n W_{\hat{g}}\left(\frac{1-s}{1+(q-1)s}\right) \\
&\quad= \sum_{\mathbf{y}\in\mathbb{Z}_q^n} \hat{\lambda}(\mathbf{y})(1-s)^{w_{\mathrm{H}}(\mathbf{y})}(1+(q-1)s)^{n-w_{\mathrm{H}}(\mathbf{y})} \\
&\quad= \sum_{\mathbf{y},\,\mathbf{x}\in\mathbb{Z}_q^n} \lambda(\mathbf{x})\omega^{\langle \mathbf{x},\,\mathbf{y}\rangle}(1-s)^{w_{\mathrm{H}}(\mathbf{y})}(1+(q-1)s)^{n-w_{\mathrm{H}}(\mathbf{y})} \\
&\quad= \sum_{\mathbf{x}\in\mathbb{Z}_q^n} \lambda(\mathbf{x})Q(x_1)Q(x_2)\cdots Q(x_n) \\
&\quad= \sum_{\mathbf{x}\in\mathbb{Z}_q^n} \lambda(\mathbf{x})q^n s^{w_{\mathrm{H}}(\mathbf{x})} = q^n W_g(s)
\end{aligned}$$

which proves (6). Equation (7) is obtained by putting $s=0$ in (6), (8) is obtained by putting $s=1$ in (6), (9) is obtained by differentiating (6) and putting $s=1$, and (10) is obtained by differentiating (6) twice, putting $s=1$, and combining with (9).

V. Some Lemmas

Let P be a probability distribution on $\mathbb{Z}_q^n$ and let

$$\lambda_P(\mathbf{z}) = \sum_{\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})P(\mathbf{x}-\mathbf{z}).$$

In particular, by (1) we get

$$\lambda_P(\mathbf{0}) = \sum_{\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})^2 = \frac{1}{q^n}(qL(P)+1). \tag{11}$$

Furthermore

$$\begin{aligned}
Ed_{\mathrm{H}}(X,Y) &= \sum_{\mathbf{z},\,\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})P(\mathbf{x}-\mathbf{z})d_{\mathrm{H}}(\mathbf{x},\,\mathbf{x}-\mathbf{z}) \\
&= \sum_{\mathbf{z}\in\mathbb{Z}_q^n} w_{\mathrm{H}}(\mathbf{z})\lambda_P(\mathbf{z}).
\end{aligned} \tag{12}$$

Let f_P be the corresponding element of $\mathcal{S}$

$$f_P = \sum_{\mathbf{x}\in\mathbb{Z}_q^n} \lambda_P(\mathbf{x})t^{\mathbf{x}}.$$

By definition of the transform, we get

$$\begin{aligned}
\hat{\lambda}_P(\mathbf{u}) &= \sum_{\mathbf{z}\in\mathbb{Z}_q^n} \lambda_P(\mathbf{z})\omega^{\langle \mathbf{u},\,\mathbf{z}\rangle} \\
&= \sum_{\mathbf{x},\,\mathbf{z}\in\mathbb{Z}_q^n} P(\mathbf{x})P(\mathbf{x}-\mathbf{z})\omega^{\langle \mathbf{u},\,\mathbf{z}\rangle} && (13) \\
&= \sum_{\mathbf{x},\,\mathbf{z}\in\mathbb{Z}_q^n} P(\mathbf{x})P(\mathbf{x}-\mathbf{z})\omega^{\langle \mathbf{u},\,\mathbf{x}\rangle - \langle \mathbf{u},\,\mathbf{x}-\mathbf{z}\rangle} && (14) \\
&= \sum_{\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})\omega^{\langle \mathbf{u},\,\mathbf{x}\rangle}\overline{\sum_{\mathbf{y}\in\mathbb{Z}_q^n} P(\mathbf{y})\omega^{\langle \mathbf{u},\,\mathbf{y}\rangle}} \\
&= \left|\sum_{\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})\omega^{\langle \mathbf{u},\,\mathbf{x}\rangle}\right|^2. && (15)
\end{aligned}$$

Lemma 1: We have

i) $A_0(\hat{f}_P) = 1$;
ii) $A_i(\hat{f}_P) \geq 0$ for $1 \leq i \leq n$;
iii) $\sum_{i=1}^{n} A_i(\hat{f}_P) = qL(P)$.

Proof: By (15) we have

$$A_0(\hat{f}_P) = \hat{\lambda}_P(\mathbf{0}) = \left|\sum_{\mathbf{x}\in\mathbb{Z}_q^n} P(\mathbf{x})\right|^2 = 1$$

and

$$A_i(\hat{f}_P) = \sum_{\substack{\mathbf{u}\in\mathbb{Z}_q^n \\ w_{\mathrm{H}}(\mathbf{u})=i}} \hat{\lambda}_P(\mathbf{u}) \geq 0.$$

Finally, since $A_0(f_P) = \lambda_P(\mathbf{0})$, iii) follows from i), (7), and (11). ∎

Lemma 2: Let X, Y be two i.i.d. random n-tuples with the common probability distribution P. Then

$$Ed_{\mathrm{H}}(X,Y) = \frac{n(q-1)}{q} - \frac{A_1(\hat{f}_P)}{q} \tag{16}$$

$$Dd_{\mathrm{H}}(X,Y) = \frac{n(q-1)}{q^2} + \frac{(q-2)}{q^2}A_1(\hat{f}_P) - \frac{1}{q^2}A_1(\hat{f}_P)^2 + \frac{2}{q^2}A_2(\hat{f}_P). \tag{17}$$

Proof: By (12)

$$Ed_{\mathrm{H}}(X,Y) = \sum_{i=0}^{n} iA_i(f_P)$$

and similarly

$$Dd_{\mathrm{H}}(X,Y) = \sum_{i=0}^{n} i^2A_i(f_P) - \left(\sum_{i=0}^{n} iA_i(f_P)\right)^2.$$

Therefore, the lemma follows from (9), (10), and Lemma 1-i). ∎

VI. Proofs of the Theorems

Proof of Theorem 1: From Lemma 1-iii), we get

$$0 \leq A_1(\hat{f}_P) \leq qL(P). \tag{18}$$

Combining this with (16), Theorem 1 follows. ∎

Proof of Theorem 2: Let $r = A_1(\hat{f}_P)/q$. By (18), $0 \leq r \leq L(P)$. Since $A_2(\hat{f}_P) \geq 0$, (17) implies

$$Dd_{\mathrm{H}}(X,Y) \geq \frac{n(q-1)}{q^2} + F(r)$$

where

$$F(x) = \frac{q-2}{q}x - x^2.$$

If $L(P) \le (q-2)/q$, then $0 \le r \le (q-2)/q$ and so

$$F(r) \ge 0.$$

If $L(P) \ge (q-2)/q$, then

$$F(r) \ge F(L(P)).$$

■

Proof of Theorem 3: First we note that Lemma 1-iii) implies

$$0 \le A_2(\hat{f}_P) \le qL(P) - A_1(\hat{f}_P). \tag{19}$$

Combining this with (17), we see that $Dd_{\mathrm{H}}(X, Y)$ is upper bounded by

$$\frac{n(q-1)}{q^2} + \frac{2}{q} L(P) + \frac{(q-4)}{q^2} A_1(\hat{f}_P) - \frac{1}{q^2} A_1(\hat{f}_P)^2.$$

The bounds in Theorem 3 can be derived from this in the same way as was done above for the bounds in Theorem 2. We omit the details. ■

Proof of Theorem 4: If $Ed_{\mathrm{H}}(X, Y) = n(q-1)/q$, then, by (16), $A_1(\hat{f}_P) = 0$. Therefore, the bounds for $Dd_{\mathrm{H}}(X, Y)$ follow from (17) and (19).

The second part of the theorem is most easily proven by a generalization of the proof given by Althöfer and Sillke [3] for $q = 2$. Let

$$h_i(j) = \sum_{\substack{\mathbf{x} \in V^n \\ x_i = j}} P(\mathbf{x}). \tag{20}$$

Clearly

$$\sum_{j=0}^{q-1} h_i(j) = 1 \quad \text{for } i = 1, 2, \cdots, n. \tag{21}$$

Furthermore

$$\begin{aligned} Ed_{\mathrm{H}}(X, Y) &= \sum_{\mathbf{x}, \mathbf{y} \in V^n} P(\mathbf{x}) P(\mathbf{y}) d_{\mathrm{H}}(\mathbf{x}, \mathbf{y}) \\ &= \sum_{i=1}^{n} \sum_{\substack{\mathbf{x}, \mathbf{y} \in V^n \\ x_i \ne y_i}} P(\mathbf{x}) P(\mathbf{y}) \\ &= \sum_{i=1}^{n} \sum_{j=0}^{q-1} h_i(j) \sum_{\substack{l=0 \\ l \ne j}}^{q-1} h_i(l) \\ &= \sum_{i=1}^{n} \left(\left(\sum_{j=0}^{q-1} h_i(j) \right)^2 - \sum_{j=0}^{q-1} h_i(j)^2 \right) \\ &= \sum_{i=1}^{n} \left(1 - \sum_{j=0}^{q-1} h_i(j)^2 \right) \le n\left(1 - \frac{1}{q}\right) \end{aligned}$$

by (21), and we have equality if and only if $h_i(j) = 1/q$ for all i and j. ■

Proof of Theorem 5: If $Ed_{\mathrm{H}}(X, Y) = n(q-1)/q - L(P)$, then, by (16), $A_1(\hat{f}_P) = qL(P)$, so, by (19), $A_2(\hat{f}_P) = 0$, and the result follows from (17). ■

Proof of Theorem 6: Let X and Y be two i.i.d. random n-tuples with the common distribution P_C. Then

$$E[d_{\mathrm{H}}(X, Y)]^j = \frac{1}{M^2} \sum_{\mathbf{a}, \mathbf{b} \in C} d_{\mathrm{H}}(\mathbf{a}, \mathbf{b})^j.$$

By assumption, if $d(\mathbf{a}, \mathbf{b}) \ne 0$, then (4) is satisfied. Hence, by Lemma 2, we get

$$\begin{aligned} &\frac{(q-1)^2 n(n-1)}{q^2} \\ &\quad \ge \frac{(q-1)^2 n(n-1)}{q^2} - \frac{2}{q^2} A_2(\hat{f}_{P_C}) \\ &\quad = \frac{2n(q-1) - (q-2)}{q} Ed_{\mathrm{H}}(X, Y) - E[d_{\mathrm{H}}(X, Y)]^2 \\ &\quad = \frac{1}{M^2} \sum_{\mathbf{a}, \mathbf{b} \in C} d_{\mathrm{H}}(\mathbf{a}, \mathbf{b}) \frac{2n(q-1) - (q-2) - qd_{\mathrm{H}}(\mathbf{a}, \mathbf{b})}{q} \\ &\quad \ge \frac{M-1}{M} \cdot d \cdot \frac{2n(q-1) - (q-2) - qd}{q}. \end{aligned}$$

Solving this inequality for M, Theorem 6 follows. ■

VII. Some Examples

In this section, we give some examples which illustrate the theorems above.

Example 1: Let P be uniform, that is, $P(\mathbf{x}) = 1/q^n$ for all $\mathbf{x}$. Then $L(P) = 0$. Hence, Theorems 1 and 2 give $Ed_{\mathrm{H}}(X, Y) = n(q-1)/q$ and $Dd_{\mathrm{H}}(X, Y) = n(q-1)/q^2$. We note that $P = P_C$ where $C = GF(q)^n$.

Example 2: Now, consider the other extreme: let $P(\mathbf{0}) = 1$ and $P(\mathbf{x}) = 0$ for all $\mathbf{x} \ne \mathbf{0}$. Then $L(P) = q^{n-1} - (1/q)$. By direct computation, we find that $Ed_{\mathrm{H}}(X, Y) = Dd_{\mathrm{H}}(X, Y) = 0$. We see that the bounds given by the theorems are quite crude in this case.

Example 3: Consider $n = 1$. Then, by Lemma 1-iii), $A_1(\hat{f}_P) = qL(P)$ [and $A_2(\hat{f}_P) = 0$] and so by (16) and (17),

$$Ed_{\mathrm{H}}(X, Y) = \frac{q-1}{q} - L(P)$$

and

$$Dd_{\mathrm{H}}(X, Y) = \frac{(q-1)}{q^2} + \frac{q-2}{q} L(P) - L(P)^2.$$

Example 4: Consider $n = 2$. Let $r = A_1(\hat{f}_P)/q$, where then $0 \le r \le L(P)$. By Lemma 1-iii), $A_2(\hat{f}_P)/q = L(P) - r$. Hence, by (16) and (17)

$$Ed_{\mathrm{H}}(X, Y) = 2\frac{q-1}{q} - r$$

and

$$Dd_{\mathrm{H}}(X, Y) = 2\frac{(q-1)}{q^2} + \frac{2}{q} L(P) + \frac{q-4}{q} r - r^2.$$

In particular, for $r = 0$ we get $Ed_{\mathrm{H}}(X, Y) = 2(q-1)/q$ and $Dd_{\mathrm{H}}(X, Y) = 2(q-1)/q^2 + (2/q)L(P)$. For $q = 2$, this occurs if $P(00) = P(11) = 1/2$ and $P(01) = P(10) = 0$.

For $r = L(P)$, we get $Ed_{\mathrm{H}}(X, Y) = 2(q-1)/q - L(P)$ and $Dd_{\mathrm{H}}(X, Y) = 2(q-1)/q^2 + ((q-2)/q)L(P) - L(P)^2$. For $q = 2$, this occurs if $P(00) + P(11) = 1/2$ and $P(01) + P(10) = 1/2$.

Example 5: Consider $n = 3$ and $q = 2$. If

$$P(000) = P(110) = P(101) = P(011) = r$$
$$P(001) = P(010) = P(100) = P(111) = \tfrac{1}{4} - r$$

where $0 \le r \le \frac{1}{4}$, and then some calculations show that $A_1(\hat{f}_P) = A_2(\hat{f}_P) = 0$, so $Ed_{\mathrm{H}}(X, Y) = \frac{3}{2}$ and $Dd_{\mathrm{H}}(X, Y) = \frac{3}{4}$ (which could also easily be shown directly). Examples 4 and 5 illustrate that both the lower and the upper bounds on $Dd_{\mathrm{H}}(X, Y)$ in Theorem 4 may be obtained for nonzero $L(P)$.

ACKNOWLEDGMENT

The authors thanks the anonymous referees and Associate Editor A. Barg who helped the authors improve their results and the exposition. In particular, one referee pointed out that the results are valid for all q; in the original manuscript, the authors only considered prime powers q and used $V = GF(q)$.

REFERENCES

[1] R. Ahlswede and G. Katona, "Contributions to the geometry of Hamming spaces," *Discrete Math.*, vol. 17, pp. 1–22, 1977.
[2] R. Ahlswede and I. Althöfer, "The asymptotic behaviour of diameters in the average," *J. Comb. Theory, Series B*, vol. 61, pp. 167–177, 1994.
[3] I. Althöfer and T. Sillke, "An 'average distance' inequality for large subsets of the cube," *J. Comb. Theory, Series B*, vol. 56, pp. 296–301, 1992.
[4] F.-W. Fu and S.-Y. Shen, "On the expectation and variance of Hamming distance between two binary i.i.d. random n-tuple," *Acta Mathematicae Applicatae Sinica*, vol. 13, no. 3, pp. 243–250, 1997.
[5] V. I. Levenshtein, "Bounds for packings of metric spaces and some of their applications," *Prob. Cyber.*, vol. 40, pp. 43–110, 1983 (in Russian).
[6] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes.* New York: North-Holland, 1977.
[7] S. Roman, *Coding and Information Theory.* Berlin, Germany: Spring-Verlag, 1991.
[8] S.-T. Xia and F.-W. Fu, "On the average Hamming distance for binary codes," *Discrete Applied Mathematics*, to be published.

Note on Taking Square-Roots Modulo N

Eric Bach and Klaus Huber, *Member, IEEE*

***Abstract*— In this contribution it is shown how Gauss' famous cyclotomic sum formula can be used for extracting square-roots modulo N.**

***Index Terms*—Cryptography, factoring, Gauss sums, square-roots modulo N.**

I. INTRODUCTION

The task of taking square-roots modulo an integer N is a problem of considerable importance, in particular, as some well-known cryptographic schemes are based on this operation ([7], [25]). In this correspondence, we modify and analyze a famous formula due to Gauss (see [8], [12], and [15])

$$\sum_{s=0}^{n-1} e^{2\pi i s^2/n} = \begin{cases} (1+i)\sqrt{n}, & \text{for } n \equiv 0 \bmod 4 \\ \sqrt{n}, & \text{for } n \equiv 1 \bmod 4 \\ 0, & \text{for } n \equiv 2 \bmod 4 \\ i \cdot \sqrt{n}, & \text{for } n \equiv 3 \bmod 4 \end{cases} \tag{1}$$

to compute square-roots modulo N. We start with prime numbers and then comment on composite numbers. For simplicity without loss of generality the N considered are odd numbers. First, however, we give a short overview of other square-root methods.

Manuscript received December 21, 1997; revised July 6, 1998.

E. Bach is with the Department of Computer Sciences, University of Wisconsin-Madison, Madison, WI 53706 USA.

K. Huber is with FE31a, Deutsche Telekom AG, Technologiezentrum, Am Kavalleriesand, Darmstadt, Germany.

Communicated by D. Stinson, Associate Editor on Complexity and Cryptography.

Publisher Item Identifier S 0018-9448(99)01389-9.

II. KNOWN METHODS FOR EXTRACTING SQUARE-ROOTS MODULO p

We consider the problem of finding the values of x in

$$x^2 \equiv n \bmod p \tag{2}$$

where n and p are given. Equation (2) has two solutions x_1 and $x_2 = -x_1$, which lie in $GF(p)$ if n is a quadratic residue, i.e., if the Legendre symbol $(n/p) = n^{(p-1)/2} \bmod p$ equals one.

Clearly, we have

$$n^{(p+1)/2} \equiv n \bmod p$$

and, thus, the determination of the roots $x_{1/2}$ is particularly easy if $(p+1)/2$ is even, i.e., for $p \equiv 3 \bmod 4$. In this case (see [3] and [11]), we immediately obtain the two solutions

$$x_{1/2} = \pm n^{(p+1)/4} \bmod p \quad \text{for } p \equiv 3 \bmod 4 \tag{3}$$

which can be determined efficiently using the square and multiply algorithm (see, e.g., [3], [10], and [20]).

For $(p+1)/2$ odd, i.e., $p \equiv 1 \bmod 4$, the situation is more difficult. In this case, all known efficient methods (see, e.g., [3], [6], [10], [16], [18], [20], [23], and [24]), are either probabilistic or make use of a quadratic nonresidue (QNR). Most of these are equivalent to or use the same basic ideas as either the Tonelli–Shanks or the Cipolla–Lehmer procedure.

The Tonelli–Shanks method [23], [24] relies on exponentiation in $GF(p)$. The method of Cipolla–Lehmer [6], [13], [20, pp. 287–288] uses exponentiation in $GF(p^2)$, which can be done efficiently using Lucas numbers. Compared to the Tonelli–Shanks method, the Cipolla–Lehmer method has the disadvantage that one has to determine a QNR which depends both on p and n, whereas the method of Tonelli–Shanks can reuse the same QNR for different n. From an aesthetic point of view this is an advantage if many square roots are computed with the same prime p, a case that often occurs in practice. From a computational point of view this advantage is not great as the Jacobi symbol algorithm is much faster than an exponentiation (see, e.g., [3, p. 113]). For further details on the computing time of Shanks' algorithms see [14].

In addition to the above two techniques, one can also use polynomial factoring methods such as the algorithms of Ben-Or/Rabin or Berlekamp (see [4], [5], and [19]). One probabilistic square-root method of Peralta [17, first algorithm] is a disguised version of Rabin/Ben-Or's algorithm (see [2, Remark 1, p. 1496]). It is interesting to note that a "shortcut" version of the Cipolla–Lehmer procedure does practically the same computation. In its polynomial version [3, p. 157] this procedure computes $\sqrt{n}$ as $X^{(p+1)/2}$ modulo the irreducible polynomial $X^2 - tX + n$. If we stop at the $(p-1)/2$ power, we obtain

$$X^{(p-1)/2} \equiv \sqrt{n} X^{-1} \equiv uX + v.$$

We have $X - t + nX^{-1} \equiv 0$, so $\sqrt{n} = -u^{-1}$.

Schoof's deterministic algorithm [22]—as the algorithm treated here—is efficient for small input values.

III. METHOD BASED ON GAUSS' FORMULA

Gauss' formula (1) can be modified in a rather straightforward way for use in finite fields $GF(p)$ by simply replacing the number