SIAM J. DISCRETE MATH.
Vol. 14, No. 3, pp. 423–431

A HYPERGRAPH APPROACH TO THE IDENTIFYING PARENT PROPERTY: THE CASE OF MULTIPLE PARENTS*

ALEXANDER BARG†, GÉRARD COHEN‡, SYLVIA ENCHEVA§, GREGORY KABATIANSKY¶, AND GILLES ZÉMOR‖

Abstract. Let C be a code of length n over an alphabet of q letters. An n-word y is called a descendant of a set of t codewords $x^1, \dots, x^t$ if $y_i \in \{x_i^1, \dots, x_i^t\}$ for all $i = 1, \dots, n$. A code is said to have the t-identifying parent property if for any n-word that is a descendant of at most t parents it is possible to identify at least one of them. We prove that for any $t \le q - 1$ there exist sequences of such codes with asymptotically nonvanishing rate.

Key words. Helly property, error-correcting codes, identifying parent property

AMS subject classifications. 94A60, 05C65

PII. S0895480100376848

1. Introduction.

Let Q be an alphabet of size q, and let us call any subset C of Q^n an (n, M)-*code* when $|C| = M$. Elements $x = (x_1, \dots, x_n)$ of C will be called *codewords*.

Let C be an (n, M)-code. Suppose $X \subseteq C$. For any coordinate i define the *projection*

$$P_i(X) = \bigcup_{x \in X} x_i.$$

Define the *envelope* $e(X)$ of X by

$$e(X) = \{x \in Q^n : \forall i, x_i \in P_i(X)\}.$$

Elements of the envelope $e(X)$ will be called *descendants* of X. Observe that $X \subseteq e(X)$ for all X, and $e(X) = X$ if $|X| = 1$.

Given a word $s \in Q^n$ (a son) which is a descendant of X, we would like to identify without ambiguity at least one member of X (a parent). When this is always possible for any descendant s of an X of size two, the code C is said to have the *identifiable parent property* [9]. More generally, we have the following definition.

DEFINITION 1.1. *For any $s \in Q^n$ let $\mathcal{H}_t(s)$ be the set of subsets $X \subset C$ of size at most t such that $s \in e(X)$. We shall say that C has the* identifiable parent property

*Received by the editors August 11, 2000; accepted for publication June 4, 2001; published electronically August 29, 2001. These results were presented at the International Workshop on Coding and Cryptography, Paris, France, 2001.
http://www.siam.org/journals/sidma/14-3/37684.html

†Bell Labs, Lucent Technologies, 600 Mountain Ave., Rm. 2C-375, Murray Hill, NJ 07974 and IPPI RAN, Moscow, Russia (abarg@research.bell-labs.com). The research of this author was supported in part by Binational Science Foundation (USA-Israel) under grant 1999099.

‡ENST, 46 rue Barrault, 75013 Paris, France (cohen@infres.enst.fr).

§HSH, Bjørnsonsg. 45, 5528 Haugesund, Norway (sbe@hsh.no).

¶IPPI RAN, Bol'shoj Karetnyj 19, Moscow 101447, Russia (kaba@iitp.ru). This research was done while the author was visiting DIMACS Center, Rutgers University, Piscataway, NJ, in March, 2000. The research of this author was supported in part by the Russian Foundation for Fundamental Research grant 99-01-00828.

‖ENST, 46 rue Barrault, 75013 Paris, France (zemor@infres.enst.fr).

of order t *(or is a* t*-identifying code, or is* t *i.p.p. for short) if for any* $s \in Q^n$*, either* $\mathcal{H}_t(s) = \emptyset$ *or*

$$\bigcap_{X \in \mathcal{H}_t(s)} X \neq \emptyset.$$

It will be convenient to view $\mathcal{H}_t(s)$ as the set of edges of a hypergraph. Its vertices are codewords of C.

Example. Let $C \subset \{0,1,2,3\}^4$ be the code defined by $C = \{u, v, w, x, y, z\}$ where

$$\begin{aligned} u &= \begin{bmatrix} 0 & 1 & 2 & 3 \end{bmatrix}, \\ v &= \begin{bmatrix} 1 & 2 & 3 & 0 \end{bmatrix}, \\ w &= \begin{bmatrix} 2 & 3 & 0 & 1 \end{bmatrix}, \\ x &= \begin{bmatrix} 3 & 0 & 1 & 2 \end{bmatrix}, \\ y &= \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, \\ z &= \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}. \end{aligned}$$

The triple $\{w, y, z\}$ can produce the son $s = (2010)$. The hypergraph $\mathcal{H}_3(s)$ contains three edges, namely, $X = \{v, w, x\}$, $X' = \{w, x, y\}$, and $X'' = \{w, y, z\}$. Their intersection is $X \cap X' \cap X'' = \{w\}$, and w is therefore identified as a parent of s. The code C, however, is not 3-identifying; it is not even 2-identifying since $\mathcal{H}_2(0101) = \{u, w\} \cup \{y, z\}$ and $\{u, w\} \cap \{y, z\} = \emptyset$.

The concept of t-identification originates with the work of Chor, Fiat, and Naor [5] on broadcast encryption. It is also related to the problem of fingerprinting numerical data [4].

It is not difficult to prove that if the minimum Hamming distance of C is big enough, then C must be t-identifying. We have [5] the following proposition.

PROPOSITION 1.2. *If* C *has minimum Hamming distance* d *satisfying*

$$d > (1 - 1/t^2)n,$$

then C *is a* t*-identifying code.*

Actually, this condition implies a stronger property, namely, t-traceability; see [15].

As usual, let $R = R(C) = \log_q M/n$ denote the rate of the (n, M)-code C. Let $R_q(t) = \liminf_{n\to\infty} \max R(C_n)$, where the maximum is computed over all t-identifying codes C_n of length n.

Note that for alphabet sizes $q \leq t^2$, Proposition 1.2 does not prove that $R_q(t) > 0$ (because, for example, (n, M)-codes that satisfy the distance condition must have $M \leq qd$; see Plotkin's bound, e.g., in [12]).

In fact, nontrivial t-identifying codes do not always exist if the alphabet size q is not big enough. Hollman et al. [9] give constructions of 2-identifying codes and existence bounds on $R_q(2)$ for any alphabet size $q \geq 3$. They prove

$$R_q(2) \geq \log_q(q/(4q^2 - 6q + 3)^{1/3}). \tag{1.1}$$

The case of arbitrary t was discussed in a recent paper [15], where it is shown that nontrivial t-identifying codes do not exist when $t > q - 1$ and do exist when $q \geq \lfloor (t+2)^2/4 \rfloor$. Consequently, it was asked in [15] whether $R_q(t) > 0$ for any $t \leq q - 1$. In this paper we shall answer this question and prove the following theorem.

THEOREM 1.3. *$R_q(t) > 0$ if and only if $t \le q - 1$.*

We shall also give a lower bound on $R_q(t)$. We shall give particular attention to the case $t = 3$ and in the case $t = 2$, $q = 3$ strengthen (1.1) by showing that it can be achieved by a sequence of linear ternary i.p.p. codes.

2. Decomposing the t-identifying property with the Berge–Duchet theorem. Let us call a subset of edges of a hypergraph a *star* if it has a nonempty intersection. A code C is t-identifying if all the nonempty hypergraphs $\mathcal{H}_t(s)$ are stars. In this section we give necessary and sufficient conditions for $\mathcal{H}_t(s)$ to be a star.

Let us say that a family of sets, any t (or less) of which have nonempty intersection, is *t-wise* intersecting.

Recall that a family of sets has the *t-Helly property* if every t-wise intersecting finite subfamily is a star.

Let us quote a Helly-type result due to Berge and Duchet [3]; see also [6, p. 393].

THEOREM 2.1. *A hypergraph has the t-Helly property if and only if, for every set A of $t + 1$ vertices, all the edges E such that $|E \cap A| \ge t$ share a common vertex.*

Let us reword this result for our purposes. Recall that a hypergraph on $t + 1$ vertices whose edges are all the t-subsets is called a t-simplex, denoted $K_t(t + 1)$.

COROLLARY 2.2. *The hypergraph $\mathcal{H}_t(s)$ has the t-Helly property if and only if it does not contain $K_t(t + 1)$ as a subhypergraph.*

Proof. Consider any set A of size $t + 1$ vertices of $\mathcal{H}_t(s)$. Let $E_1, E_2, \dots, E_m$ be all the edges of $\mathcal{H}_t(s)$ that have at least t vertices in A. Since the edges of $\mathcal{H}_t(s)$ have at most t vertices we have $|E_i| = t$ for all i and

$$|E_1 \cap E_2 \cap \cdots \cap E_m| = t + 1 - m.$$

Therefore this intersection is nonempty if and only if $m < t + 1$; i.e., $E_1, E_2, \dots, E_m$ do not make up the t-simplex with vertex set A. □

Reworded again, we get the following corollary.

COROLLARY 2.3. *Suppose the hypergraph $\mathcal{H}_t(s)$ has at least $t + 1$ vertices. Then it is a star if and only if*

1. *any t (or less) of its edges have a nonempty intersection;*
2. *it does not contain $K_t(t + 1)$.*

3. Ensuring t-identification for any $t \le q + 1$.

3.1. Hashing families. A subset C of Q^n is said to be *t-hashing* (or t-separating; see, e.g., [10]) if any t of its members have t distinct entries in some common coordinate $i \in \{1, \dots, n\}$.

LEMMA 3.1. *$C \subset Q^n$ is $(t + 1)$-hashing if and only if $\mathcal{H}_t(s)$ has the t-Helly property for every $s \in Q^n$.*

Proof. Suppose C is $(t + 1)$-hashing. Let A be any set of $t + 1$ codewords, and let $s \in Q^n$. Since there is a coordinate where the codewords of A are all different, there exists at least one subset $X \subset A$, $|X| = t$, such that $s \notin e(X)$. Therefore $\mathcal{H}_t(s)$ cannot contain a t-simplex.

Conversely, suppose C is not $(t + 1)$-hashing, so that there exists a subset A of $t + 1$ codewords such that for every coordinate i, there exist at least two distinct codewords a, b of A such that $a_i = b_i$. Then define $s \in Q^n$ by choosing, for every coordinate $i \in \{1, \dots, n\}$, a value that occurs at least twice among the a_i, $a \in A$. Then we have $s \in e(X)$ for every subset X of size t of A, which means that $\mathcal{H}_t(s)$ contains the t-simplex with vertex set A. □

Remark. A consequence of Lemma 3.1 is that when $q \le t$, there are no q-ary t-identifying codes C of size $|C| \ge t+1$ (see Lemma 1.6 of [15]).

We now have a condition on C that ensures that all the hypergraphs $\mathcal{H}_t(s)$ do not contain t-simplexes. To apply Corollary 2.3 we now need a condition to ensure that any t edges of any $\mathcal{H}_t(s)$ have a nonempty intersection.

3.2. Partially hashing families.

DEFINITION 3.2. *Let us say that a subset $C \subset Q^n$ is (t,u)* partially hashing *if for any two subsets T, U of C such that $T \subset U \subset C$, $|T| = t$, $|U| = u$, there is some coordinate $i \in \{1, \dots, n\}$ such that for any $x \in T$ and any $y \in U, y \ne x$, we have $x_i \ne y_i$.*

Remark. If $u = t+1$, then (t,u) partial hashing is the same as $(t+1)$-hashing.

The motivation for this last definition is the following lemma.

LEMMA 3.3. *Let $\mathcal{X}$ be a subset of edges of $\mathcal{H}_t(s)$, and let u be an upper bound on the number of vertices spanned by the edges of $\mathcal{X}$. If C is (t,u) partially hashing, then $\mathcal{X}$ is a star, i.e., $\cap_{X \in \mathcal{X}} X \ne \emptyset$.*

Proof. Let $U = \cup_{X \in \mathcal{X}} X$, so that $|U| \le u$ by the hypothesis. Let T be some edge of $\mathcal{X}$. Because C is (t,u) partially hashing, there is some coordinate i satisfying the condition of Definition 3.2 for T and U. Then $s_i = x_i$ for some $x \in T$ because $s \in e(T)$. However, then the definition implies that for all $y \ne x$, $y \in U$, we have $y_i \ne s_i$. Since all edges X of $\mathcal{X}$ are in $\mathcal{H}_t(s)$ we conclude that they must all contain x. □

Lemma 3.1 means that to enforce t-identification it is sufficient to have $(t+1)$-hashing and any property which forces any t edges of $\mathcal{H}_t(s)$ to intersect. Since any t edges of $\mathcal{H}_t(s)$ span at most t^2 vertices, Lemma 3.3, together with the remark after Definition 3.2, now implies the following corollary.

COROLLARY 3.4. *If C is (t,t^2) partially hashing, then C is a t-identifying code.*

The (t,u) hashing property is easier to handle than t-identification; in particular, it will give us a lower bound on $R_q(t)$ through the probabilistic method.

3.3. A lower bound on the size of (t,u) partially hashing codes. Fix $t \le q-1$ and let $u \ge t+1$. We apply the probabilistic method with expurgation (see, e.g., [2]) to (t,u) partially hashing codes. This means that we take a random (n,M)-code C and compute the expectation $\mathbf{E}$ of the number of pairs of subsets T, U, $T \subset U \subset C$, $|T| = t$, $|U| = u$, that contradict the (t,u) partially hashing property. Whenever $\mathbf{E} \le M/2$, then $(n, M/2)$-codes with the (t,u) partially hashing property exist.

The probability that a given T and U violate the partially hashing property is

$$P_{t,u,n} = \left(1 - \frac{q(q-1)\cdots(q-t+1)(q-t)^{u-t}}{q^u}\right)^n = \left(1 - \frac{q!(q-t)^{u-t}}{(q-t)!q^u}\right)^n.$$

The expectation of the number of pairs T, U that violate the partially hashing property is

$$\mathbf{E} = \binom{M}{u}\binom{u}{t} P_{t,u,n}.$$

Writing $M = q^{Rn}$ and letting n go to infinity we get that infinite sequences of (t,u) partially hashing codes exist for all rates R such that $\log_q \mathbf{E} < Rn$, i.e., such that

$$uR + \frac{1}{n}\log_q P_{t,u,n} < R.$$

Hence we get the following lemma.

LEMMA 3.5. *Let $u \geq t+1$: infinite sequences of (t,u) partially hashing codes exist for all rates R such that*

$$R < \frac{1}{u-1} \log_q \frac{(q-t)!q^u}{(q-t)!q^u - q!(q-t)^{u-t}}.$$

As a consequence, applying Corollary 3.4, we obtain $R_q(t) > 0$ for any $q \geq t+1$ which proves Theorem 1.3.

3.4. Improvements: Forbidding minimal configurations. We shall now show that the quantity t^2 in Corollary 3.4 can be lowered; namely, we shall obtain the following lemma.

LEMMA 3.6. *Let $u = \lfloor (t/2+1)^2) \rfloor$. If C is (t,u) partially hashing, then C is a t-identifying code.*

Before proving Lemma 3.6 it will be convenient to decompose all subsets of edges with empty intersection into "minimal forbidden configurations."

Let $\mathcal{X} = (X_1, \ldots, X_m)$ be a collection of subsets of codewords with $|X_i| \leq t$, $i = 1, \ldots, m$. We shall call $\mathcal{X}$ a *configuration* if it has an empty intersection, $\cap_{i=1}^m X_i = \emptyset$, and we shall say that $\mathcal{X}$ is a *minimal configuration* if it is minimal under inclusion, i.e., if $\cap_{i \neq j} X_i \neq \emptyset$ for any $j = 1, \ldots, m$.

Let $\mathcal{X}$ be a minimal configuration of size m. A set $B(\mathcal{X}) = (b^1, \ldots, b^m)$ will be called a *frame* of $\mathcal{X}$ if

$$b^j \in \bigcap_{i \neq j} X_i.$$

By minimality, frames of minimal configurations always exist. One useful property of frames gives rise to the following lemma, which follows somewhat along the lines of [15].

LEMMA 3.7. *Let $\mathcal{X}$ be a minimal configuration. Then*

$$\left| \bigcup_{i=1}^m X_i \right| \leq \sum |X_i| - m(m-2).$$

Proof. All the points b^j of a frame $B(\mathcal{X})$ are different since otherwise $\cap_i X_i \neq \emptyset$. Furthermore, by definition of $B(\mathcal{X})$ we have $(B(\mathcal{X}) \setminus b^j) \subset X_j$ for any $j = 1, \ldots, m$, and hence $m - 1 \leq t$. Then

$$\begin{aligned} \left| \bigcup_{i=1}^m X_i \right| &\leq \sum (|X_i \setminus (B(\mathcal{X}) \setminus b^i)|) + |B(\mathcal{X})| \\ &= \sum_{i=1}^m (|X_i| - (m-1)) + m. \qquad \square \end{aligned}$$

Since $|X_i| \leq t$, we obtain $|\cup X_i| \leq m(t-m+2)$. The maximum on m of this expression for $m = 1, \ldots, t$ is $u = \lfloor (t/2+1)^2) \rfloor$, which is also an upper bound on the cardinality of a minimal configuration (see [15]).

Note that the maximum value of m which gives a positive upper bound is $t+1$. Also note that the only minimal configurations with $m = t+1$ are t-simplexes. In particular this gives an alternative proof of Corollary 2.3.

We observe that any configuration must contain some minimal configuration. Now apply Lemma 3.3 as before to obtain Lemma 3.6. Lemmas 3.5 and 3.6 imply the following improved lower bound on $R_q(t)$.

THEOREM 3.8. *Let* $u = \lfloor (t/2+1)^2) \rfloor$. *We have*

$$R_q(t) \geq \frac{1}{u-1} \log_q \frac{(q-t)!q^u}{(q-t)!q^u - q!(q-t)^{u-t}}.$$

4. Small t. The (t,u) partially hashing property is only a sufficient condition for a code to be t-identifying. In the case of small t we can obtain precise necessary and sufficient conditions. A code C is t-identifying if and only if, for every s such that $\mathcal{H}_t(s) \neq \emptyset$, the hypergraph $\mathcal{H}_t(s)$ has the t-Helly property and any m edges of $\mathcal{H}_t(s)$ have a nonempty intersection for any m, $2 \leq m \leq t$. Lemma 3.1, together with Corollary 2.3, tells us therefore that C is t-identifying if and only if it is $(t+1)$-hashing and any m edges of $\mathcal{H}_t(s)$ have a nonempty intersection for any m, $2 \leq m \leq t$, and for any $s \in Q^n$. For $t = 2$, the latter property means that $\mathcal{H}_2(s)$ does not contain disjoint edges. Equivalently, for any $X = \{a, b\}$, $Y = \{c, d\}$, with a, b, c, d distinct codewords, $e(X) \cap e(Y) = \emptyset$. Equivalently again, this means that there exists a coordinate i such that

$$\{a_i, b_i\} \cap \{c_i, d_i\} = \emptyset. \tag{4.1}$$

This property of C was named IPP2 by Hollman et al. in [9]. In other contexts it has often been called $(2,2)$-separation and has been investigated by a number of authors [7, 8, 11, 13, 14]. The 3-hashing property was called IPP1 in [9].

Let us now characterize the 3 i.p.p. property.

4.1. The case $t = 3$. This time Lemma 3.1 and Corollary 2.3 tell us that C is 3-identifying if and only if

(i) it is 4-hashing;

(ii) for any $s \in Q^n$, any two edges of $\mathcal{H}_3(s)$ have a nonempty intersection and any three edges of $\mathcal{H}_3(s)$ have a nonempty intersection.

Condition (ii) is equivalent to saying that $\mathcal{H}_3(s)$ does not contain minimal configurations of size $m = 2$ and of size $m = 3$.

That $\mathcal{H}_3(s)$ does not contain minimal configurations of size two is equivalent to saying that for any $X = \{a, b, c\}$, $\{d, e, f\}$, with a, b, c, d, e, f distinct codewords, $e(X) \cap e(Y) = \emptyset$, which means that there exists a coordinate i such that

$$\{a_i, b_i, c_i\} \cap \{d_i, e_i, f_i\} = \emptyset. \tag{4.2}$$

Property (4.2) is usually called $(3,3)$-separation [7, 8, 11, 14].

Remark. As proved in [15] and follows from Corollary 2.3, the $(t+1)$-hashing and (t,t) separation properties are necessary for a code to have the t i.p.p. property. For $t = 2$, they are also sufficient [9]. For $t = 3$, we show how to complement them to form a set of sufficient conditions.

There remains to characterize the condition that $\mathcal{H}_3(s)$ does not contain a minimal configuration of three edges, i.e., $\mathcal{X} = (X, Y, Z)$ such that $X \cap Y \cap Z = \emptyset$, but any two edges of $\mathcal{X}$ intersect. Clearly, the only cases that we need to consider are when $|X| = |Y| = |Z| = 3$. We have two situations to forbid:

(a) for any $X, Y, Z \subset \binom{C}{3}$ such that $X \cap Y \cap Z = \emptyset$ and $|X \cap Y| = |Y \cap Z| = |Z \cap X| = 1$;

(b) for any $X, Y, Z \subset \binom{C}{3}$ such that $X \cap Y \cap Z = \emptyset$ and $|X \cap Y| = 2$ and $|Y \cap Z| = |Z \cap X| = 1$.

Forbidding these configurations means that we must have, in each case,

$$e(X) \cap e(Y) \cap e(Z) = \emptyset.$$

Those two cases involve, respectively, six and five codewords. After a straightforward examination we finally obtain the following proposition.

PROPOSITION 4.1. *C is 3-identifying if and only if the four following conditions hold:*

1. *C is 4-hashing.*
2. *C is $(3,3)$-separating.*
3. *For any six distinct codewords a, b, c, d, e, f there exists a coordinate i such that*
   - *a_i, b_i, c_i are all different,*
   - *$d_i \neq a_i$, $e_i \neq b_i$, $f_i \neq c_i$, and*
   - *d_i, e_i, f_i are not all equal.*
4. *For any five distinct codewords a, b, c, d, e there exists a coordinate i such that*
   - *$c_i \neq e_i$ and $\{a_i, b_i\} \cap \{c_i, e_i\} = \emptyset$, and*
   - *$d_i \neq a_i$ and $d_i \neq b_i$.*

Example. q=4. By repeatedly applying the probabilistic expurgation method, we get lower bounds on the rate of codes satisfying the previous four conditions. Namely,

$$\begin{aligned} R_1 &\geq \frac{1}{3}\log_4(32/29), \\ R_2 &\geq \frac{1}{5}\log_4(2^{10}/919), \\ R_3 &\geq \frac{1}{5}\log_4(256/217), \\ R_4 &\geq \frac{1}{4}\log_4(256/226). \end{aligned}$$

Taking the smallest of the R_i's gives a 3-identifying quaternary code with rate R_2, showing that

$$R_4(3) \geq \frac{1}{5}\log_4(1024/919) \approx 0.0156.$$

The lower bound of Theorem 3.8 gives only

$$R_4(3) \geq \frac{1}{5}\log_4(1024/1018) \approx 0.000848.$$

4.2. Linear i.p.p. codes. Bound (1.1) for $q = 3$ implies that

$$R_3(2) \geq (1/3)\log_3(9/7).$$

We strengthen this result by proving that the same bound holds for linear codes as well. The result in [9] implies only the existence of unrestricted codes with the same rate.

THEOREM 4.2. *There exists a sequence of linear ternary 2-identifying codes C_n with $R(C_n) \geq (1/3)\log_3(9/7)$.*

Proof. We again apply the probabilistic method and prove Theorem 4.2 by averaging this time over the ensemble of linear ternary codes. Let C be a linear subspace of F_3^n. The code C is i.p.p. if any triple of vectors in it is 3-hashing and any quadruple satisfies the separation condition (4.1).

Linear $(2,2)$-separating codes were first studied in [13], where most of the calculations below were essentially carried out. We include them here for completeness, showing that there exist such codes of rate $R \geq (1/3)\log_3(9/7)$.

Consider condition (4.1). Suppose that $\dim C = k$ and let G be a generator matrix of C, i.e., a $k \times n$ matrix whose rows form a basis of C as an F_3-linear space. Let $g^1, g^2, \ldots, g^n$ be the columns of G. Any vector $c \in C$ has the form aG for some $a \in F_3^k$. Let $c^1, \ldots, c^4$ be some vectors in C. Since the i.p.p. property is translation invariant, suppose that $c^4 = 0$. Suppose that $c^i = a^i G$ for $i = 1, 2, 3$.

Case (a). a^1, a^2, a^3 are linearly independent. Choose a basis $f_1, \ldots, f_k$ in F_3^k such that $a^i \cdot f_j = \delta_{ij}$ for $i = 1, 2, 3$; $j = 1, \ldots, k$. (Complement a^1, a^2, a^3 to a basis and take the dual basis.) Observe $\{c_m^1, c_m^2\} \cap \{c_m^3, c_m^4\} = \emptyset$ (for any given $m = 1, 2, \ldots, n$) if and only if the first three coordinates of the column g^m in the basis (f) have one of the following forms:

$$\begin{array}{ccc} \pm 1 & \pm 1 & 0 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \end{array}.$$

Hence the total number of favorable choices is 6 out of 27. This implies that the probability for a matrix G to be bad for a given linearly independent triple is $(21/27)^n = (7/9)^n$. The number of triples is less than 3^{3k}, so the probability that a given matrix spans a quadruple of vectors that violate condition (4.1) is bounded above by $3^{3k}(7/9)^n$. Hence if $R = (1/3)\log_3(9/7) - \epsilon$, for any $\epsilon > 0$, there exists a favorable choice.

Case (b). Some of the vectors a^1, a^2, a^3 are linearly dependent. For instance, suppose that a^3 is spanned by a^1, a^2, and these two are not collinear. Let $a^3 = a^1 + a^2$. Take the basis dual to a basis that includes a^1, a^2. As above, we count the number of unfavorable choices for the column g^m. Good choices for (g_1^m, g_2^m) are $(\pm 1, \pm 1)$. Hence the fraction of bad choices of G is at most $3^{2k}(5/9)^n$, and this is less than $3^{3k}(7/9)^n$. Other cases of dependence are dealt with analogously; none accounts for a fraction of bad matrices larger than in Case (a).

Now let us give a lower bound on the rate of linear 3-hashing codes. (This is a special case of a result announced in [1, Thm. 2].) Again let c^1, c^2, c^3 be some vectors in C. Since the 3-hash property is translation invariant we can assume that $c^3 = 0$. Suppose that $c^i = a^i G$ for $i = 1, 2$. There are two cases.

Case (a). a^1 and a^2 are linearly independent. Choose a basis $f_1, \ldots, f_k$ in F_3^k such that $a^i \cdot f_j = \delta_{ij}$ for $i = 1, 2$; $j = 1, \ldots, k$. Observe that $c_m^1 \neq c_m^2 \neq 0$ if and only if the first two coordinates of the column g^m in the basis (f) equal either $(1, -1)$ or $(-1, 1)$. Hence the total number of favorable choices is 2 out of 9. This implies that the probability for a matrix G to be bad for a given linearly independent pair is $(7/9)^n$. The number of pairs is less than 3^{2k}, so the probability that a given matrix spans a triple of vectors such that a^1 and a^2 are linearly independent, and such that they violate the 3-hash condition, is bounded above by $3^{2k}(7/9)^n$. Hence if $R = (1/2)\log_3(9/7) - \epsilon$, for any $\epsilon > 0$, there exists a favorable choice.

Case (b). a^1 and a^2 are collinear, i.e., $a^1 = \lambda a^2$. As above, take the basis dual to a basis that includes a^1. Good choices for g_1^m are ± 1. Hence the number of bad choices of G is at most $3^{2k}(1/3)^n$, and this is less than $3^{2k}(7/9)^n$.

It remains to find the minimum of the achievable rates for $(2,2)$-separating linear codes and for linear 3-hashing codes. This minimum is $(1/3)\log_3(9/7)$ as was to be proved. □

The argument in this section is generalized directly to prove existence of linear 2 i.p.p. codes that reach bound (1.1) over any finite field alphabet.

REFERENCES

[1] L. A. Bassalygo, M. Burmester, A. Dyachkov, and G. Kabatianski, *Hash codes*, in Proceedings of the IEEE International Symposium on Information Theory, Ulm, Germany, 1997, p. 174.

[2] L. A. Bassalygo, S. I. Gelfand, and M. S. Pinsker, *Simple methods for obtaining lower bounds in coding theory*, Problems Inform. Transmission, 27 (1991), pp. 277–281.

[3] C. Berge and P. Duchet, *A generalisation of Gilmore's theorem*, in Recent Advances in Graph Theory, M. Fiedler, ed., Academia, Prague, 1975, pp. 49-55.

[4] D. Boneh and J. Shaw, *Collusion-secure fingerprinting for digital data*, IEEE Trans. Inform. Theory, 44 (1998), pp. 480–491.

[5] B. Chor, A. Fiat, and M. Naor, *Tracing traitors*, in Advances in Cryptology–CRYPTO'94, Lecture Notes in Comput. Sci. 839, Springer-Verlag, Berlin, 1994, pp. 257–270.

[6] P. Duchet, *Hypergraphs*, in Handbook of Combinatorics, Vol. 1, R. L. Graham, M. Grötschel and L. Lovász, eds., North-Holland, Amsterdam, 1995, pp. 381–432.

[7] M. L. Fredman and J. Komlós, *On the size of separating systems and families of perfect hash functions*, SIAM J. Algebraic Discrete Methods, 5 (1984), pp. 61–68.

[8] A. D. Friedman, R. L. Graham, and J. D. Ullman, *Universal single transition time asynchronous state assignments*, IEEE Trans. Comput., 18 (1969), pp. 541–547.

[9] H. D. L. Hollmann, J. H. van Lint, J.-P. Linnartz, and L. M. G. M. Tolhuizen, *On codes with the identifiable parent property*, J. Combin. Theory Ser. A, 82 (1998), pp. 121–133.

[10] J. Körner and A. Orlitski, *Zero-error information theory,* IEEE Trans. Inform. Theory, 44 (1998), pp. 2207–2229.

[11] J. Körner and G. Simonyi, *Separating partition systems and locally different sequences*, SIAM J. Discrete Math., 1 (1988), pp. 355–359.

[12] J. H. van Lint, *Introduction to Coding Theory*, Springer-Verlag, Berlin, 1982.

[13] M. S. Pinsker and Yu. L. Sagalovich, *Lower bound for the power of an automaton state code*, Problems Inform. Transmission, 8 (1972), pp. 224–230.

[14] Yu. L. Sagalovich, *Separating systems*, Problems Inform. Transmission, 30 (1994), pp. 105–123.

[15] J. N. Staddon, D. R. Stinson, and R. Wei, *Combinatorial properties of frameproof and traceability codes*, IEEE Trans. Inform. Theory, 47 (2001), pp. 1042–1049.