Separating and Intersecting Properties of BCH and Kasami Codes

Hans Georg Schaathun and Tor Helleseth*

Dept. Informatics,
University of Bergen
Pb. 7800
N-5020 Bergen
Norway

Abstract Separating codes have recently been applied in the construction of collusion secure fingerprinting schemes. They are related to other combinatorial concepts like intersecting codes, superimposed codes, hashing families, and group testing. In this paper we study some good, binary asymptotic constructions of such codes.

Keywords: *separating systems, fingerprinting, BCH codes, Kasami codes*

1 Introduction

Copyright violations are of increasing concern to artists and distributors, as production of copies get simpler and cheaper for common people. Digital fingerprinting and traitor tracing is a technique to trace guilty users or pirates when illegal copies are found.

A digital fingerprinting scheme [4] marks every copy sold with an individual mark, such that if one pirate reproduces his copy, the illegal copies may be traced back to him. In traitor tracing schemes [6,7], a similar technique is applied to the decryption keys of a broadcast encryption system. The fingerprint must be hidden such that a user cannot change it by inspecting only his own copy.

If several pirates collude, they can compare their copies, and identify portions where they differ, which must then be part of the fingerprint. Thus having identified parts of the fingerprint, they can also change it, producing a hybrid copy which cannot trivially be traced. It is generally assumed that in each mark or symbol of the fingerprint, the pirate coalition can choose the symbol from either of their copies, but nothing else.

* Both authors were supported by the Norwegian Research Council under grant number 146874/420 and by the Aurora programme.

Collusion secure fingerprinting schemes are designed to trace at least one pirate when a coalition is guilty.

We view the fingerprints as codewords over some alphabet Q. The fingerprints the pirates are able to forge form the so-called feasible set, defined as

$$F(T) := \{(v_1, \ldots, v_n) \in Q^n \mid \forall i, 1 \leq i \leq n, \exists (a_1, \ldots, a_n) \in T, a_i = v_i\},$$

where T is the set of fingerprints held by the pirates, Q is the alphabet, and n is the length of a fingerprint. If the code of valid fingerprints still makes it possible to trace at least one guilty pirate out of a coalition of size t or less, we say that the code has the t-identifiable parent property (t-IPP). If the pirates are able to forge the fingerprint of an innocent user, we say that this user is framed. Codes which prevent framing by t pirates are called t-frameproof or $(t, 1)$-separating codes. A code is (t, t)-separating, or t-secure frameproof, if no two disjoint coalitions of size t or less can produce the same fingerprint.

Unfortunately (combinatorial) t-IPP codes are possible only for huge alphabets. Therefore it is interesting to study probabilistic t-IPP, where we permit a small non-zero probability ϵ of incorrect tracing. Recently, (t, t)-separating codes were used to construct probabilistic t-IPP codes [3,2]. In [22] it was proved that the best known asymptotic $(2, 2)$-separating codes is also probabilistic 2-IPP.

The case of $(2, 2)$-separation was introduced by Sagalovich in the context of automata: two such systems transiting simultaneously from state a to a' and from b to b' respectively should be forbidden to pass through a common intermediate state. A state of the system in this case is an n-bit binary string, and the moving from one state to another is obtained by flipping bits one by one. Only shortest paths from the old to the new state are allowed, so moving from a to a' will only involve flipping bits where a and a' differ. The set of valid states Γ forms a $(2, 2)$-separating system, if for any four distinct states, a, a', b, and b' from Γ, the transitions $a \to a'$ and $b \to b'$ cannot pass through any common state. Sagalovich's contribution on this topic is substantial and has been surveyed in [21].

The design of self-checking asynchronous networks has been a challenging problem. Friedmann et al. [15] have shown that the unicode single-transition-time asynchronous state assignment corresponds to $(2, 2)$- and $(2, 1)$-separating systems. The coding problem for automata states also motivated research on $(3, 3)$-SS [27]. In [22] it was proved that the best known asymptotic $(2, 2)$-separating codes is also 2-IPP.

Separating codes have also been studied in a set-theoretic framework, e.g. [18], and Körner [19] gives a series of problems equivalent to $(2,1)$-separation.

In this paper we present new binary, asymptotic constructions of (t,u)-separating codes. We compute the rates for (t,u)-SS when $2 \le t,u \le 5$ find that our constructions improve on previous ones. The constructions work for arbitrary t and u, but for $(t,1)$-SS previous constructions based on designs [9] are still the best.

2 Preliminary definitions and bounds

Let Q be an additive group (often a field) called the alphabet, and denote by q its order. Let $\mathbb{V}$ be the set of n-tuples over Q. An $(n,M)_q$ code Γ is an M-subset $\Gamma \subseteq \mathbb{V}$. If Q is a field of q elements and C is a k-dimensional subspace $C \leqq \mathbb{V}$, then we say that C is a $[n,k]_q$ (linear) code. We will refer to the elements of $\mathbb{V}$ as words. Let d_1 and m_1 denote respectively the minimum and maximum (Hamming) distance of the code.

Definition 1. *A pair (T,U) of disjoint sets of words is called a (t,u)-configuration if $\#T = t$ and $\#U = u$. The separating weight $\theta(T,U)$ is the number of positions i, where every word of T is different from any word of U on position i.*

The (t,u)-separating weight $\theta_{t,u}(C)$ of a code C, is the least separating weight of any (t,u)-configuration of the code. If $\theta_{t,u}(C) > 0$, then we say that C is a (t,u)-separating code or a (t,u)-SS (separating system).

In earlier works on watermarking and fingerprinting, (t,t)-separating codes have been called t-SFP (secure frameproof) [24,25,23]. The current terminology appears to be older though [21]. It is well known that codes with sufficiently large minimum distance are separating [21].

Lemma 1. *If Γ is a code with minimum distance d_1 and maximum distance m_1, then $2\theta_{2,1} \ge 2d_1 - m_1$.*

Proof. Let $(\mathbf{c};\mathbf{a},\mathbf{b})$ be a $(2,1)$-configuration. Letting the three words be rows of a matrix, we have essentially four types of columns: Type 0 where all the elements are equal, Type I where $\mathbf{a}$ or $\mathbf{b}$ differs from the two others, Type A where $\mathbf{c}$ differs from the two others, and Type B with three different elements. Let v_i denote the number of elements of Type i.

Consider the sum

$$\Sigma := w(\mathbf{c}-\mathbf{a}) + w(\mathbf{c}-\mathbf{b}) \ge 2d_1.$$

Observe that $\Sigma = 2(v_{\mathrm{A}} + v_{\mathrm{B}}) + v_{\mathrm{I}}$. Clearly we have $\theta(\mathbf{c};\mathbf{a},\mathbf{b}) = v_{\mathrm{A}} + v_{\mathrm{B}}$, and $w(\mathbf{a}-\mathbf{b}) = v_{\mathrm{B}} + v_{\mathrm{I}}$, so $v_{\mathrm{I}} \leq m_1$, and the theorem follows.

Remark 1. In the binary case, there are no columns of Type B, and therefore

$$2\theta(\mathbf{c};\mathbf{a},\mathbf{b}) = w(\mathbf{c}-\mathbf{a}) + w(\mathbf{c}-\mathbf{b}) - w(\mathbf{a}-\mathbf{b}),$$

and consequently we get equality $\theta_{2,1} = d_1 - m_1/2$ if and only if there are three codewords $\mathbf{a},\mathbf{b},\mathbf{c}$ such that $w(\mathbf{c}-\mathbf{a}) = w(\mathbf{c}-\mathbf{b}) = d_1$ and $w(\mathbf{b}-\mathbf{a}) = m_1$.

A similar argument also gives the following result [21].

Theorem 1. *Let Γ be a code with minimum distance d_1 and maximum distance m_1. Then $4\theta_{2,2} \geq 4d_1 - 2m_1 - n$. If Γ is linear, then $4\theta_{2,2} \geq 4d_1 - 3m_1$.*

Proposition 1. *Any $(n,M,d_1)_q$ code Γ has $\theta_{t,u} \geq n - tu(n-d_1)$.*

Corollary 1. *An $(n,M,d_1)_q$ code Γ is (t,u)-separating if $d_1/n > 1 - 1/(tu)$.*

Proof. Consider any (t,u)-configuration (T,U) from Γ, and define the sum

$$\Sigma := \sum_{(x,y)\in T\times U} d(x,y).$$

This is the sum of (T,U) distances in the code, so $\Sigma \geq tud_1$. Each coordinate can contribute at most tu to the sum Σ, but if any coordinate does contribute that much, then the configuration is separated on this coordinate. Hence we get that $\Sigma \leq n(tu-1) + \theta_{t,u}$. The proposition follows by combining the upper and lower bounds and simplifying.

It must be noted that, to get infinite families of separating codes with good rate, the alphabet size q grows extremely rapidly in the t and u, due to the Plotkin bound. On the other hand, for sufficiently large alphabets, we can use the following lemma by Tsfasman [26].

Theorem 2 (The Tsfasman Codes). *For any $\alpha > 0$ there are constructible, infinite families of codes $\mathfrak{A}(N)$ with parameters $[N, NR, N\delta]_q$ for $N \geq N_0(\alpha)$ and*

$$R + \delta \geq 1 - \left(\sqrt{q} - 1\right)^{-1} - \alpha.$$

Infinite families of separating codes over small alphabets can be built by concatenation [1]. The outer codes for concatenation will very often be Tsfasman codes.

Definition 2 (Concatenation). *Let C_1 be a $(n_1, Q)_q$ and let C_2 be an $(n_2, M)_Q$ code. Then the concatenated code $C_1 \circ C_2$ is the $(n_1 n_2, M)_q$ code obtained by taking the words of C_2 and mapping every symbol on a word from C_1.*

Proposition 2. *Let Γ_1 be a $(n_1, M)_{M'}$ code with minimum (t,u)-separating weight $\theta^{(1)}_{t,u}$, and let Γ_2 be a $(n_2, M')_q$ code with separating weight $\theta^{(1)}_{t,u}$. Then the concatenated code $\Gamma := \Gamma_2 \circ \Gamma_1$ has minimum separating weight $\theta_{t,u} = \theta^{(1)}_{t,u} \cdot \theta^{(2)}_{t,u}$.*

Note that Γ will usually not satisfy the requirements of Proposition 1.

3 Intersection gives separation

The first relationship between intersecting codes and separating codes appeared in [5], and further links have been explored in [11,10] (see also [13]).

Definition 3. *A linear code C of dimension $k \geq t$ is said to be t-wise intersecting if any t linearly independent codewords have intersecting supports. If $t > k$, we say that C is t-wise intersecting if and only if it is k-wise intersecting.*

It is easy to verify that any t-wise intersecting code is also $(t-1)$-wise intersecting. The following relation between intersection and separation is well known [5,11].

Proposition 3. *For a linear, binary code, is*

1. *2-wise intersecting if and only if it is $(2,1)$-separating, and*
2. *3-wise intersecting if and only if it is $(2,2)$-separating.*

Due to this proposition, we can use many bounds on separating codes as bounds on intersecting codes. For instance, by Theorem 1, every code with $4d > 3m$ is 3-wise intersecting.

It was shown in [13], that if C is a (t,u)-SS, then any $\bar{\iota}(t,u)$ codewords must be linearly independent, where

$$\bar{\iota}(t,u) := \begin{cases} t+u, & \text{when } t \equiv u \equiv 1 \pmod 2, \\ t+u-1, & \text{when } t \not\equiv u \pmod 2, \\ t+u-2, & \text{when } t \equiv u \equiv 0 \pmod 2. \end{cases}$$

If the codewords are taken as a non-linear subcode of a $(t + u - 1)$-wise intersecting code, this condition is also sufficient. The following theorem is from [10], but we include a proof for completeness.

Theorem 3. *Let $i, j \geq 1$ be integers such that $t := i+j-1 \geq 2$. Consider a t-wise intersecting, binary, linear code C, and a non-linear subcode $\Gamma \subseteq C$. The code Γ is (i, j)-separating if any $\bar{\iota}(i, j)$ non-zero codewords are linearly independent.*

Proof. We start by proving that any $t + 1$ codewords being linearly independent is sufficient for Γ to be (i, j)-separating. This holds as the theorem states irrespectively of the parities of i and j. Afterward we will strengthen the result in the cases where i and j are not both odd.

Choose any (two-part) sequence Y' of $t + 1$ codewords from Γ,

$$Y' := (\mathbf{a}'_1, \ldots, \mathbf{a}'_j; \mathbf{c}'_1, \ldots, \mathbf{c}'_{t+1-j}).$$

We have that Y' is $(j, t + 1 - j)$-separated if and only if $Y := Y' - \mathbf{c}'_{t+1-j}$ is. Hence it suffices to show that

$$Y = (\mathbf{a}_1, \ldots, \mathbf{a}_j; \mathbf{c}_1, \ldots, \mathbf{c}_{t-j}, \mathbf{0})$$

is $(j, t + 1 - j)$-separated.

Since the $t + 1$ codewords of Y' are linearly independent, so are the t first codewords of Y. Now, consider

$$X := \{\mathbf{a}_1 + \mathbf{c}_1, \ldots, \mathbf{a}_1 + \mathbf{c}_{t-j}; \mathbf{a}_1, \ldots, \mathbf{a}_j\},$$

which is a set of linearly independent codewords from C, and hence all non-zero on some coordinate i. Since $\mathbf{a}_1 + \mathbf{c}_l$ is non-zero on coordinate i, $\mathbf{c}_l$ must be zero for all l. Hence Y, and consequently Y', is separated on coordinate i.

This completes the first step. In the case where $i \not\equiv j \pmod 2$, we get that t is even, and consequently the t first codewords of Y are linearly independent whenever any t words of Y' are. Therefore it is sufficient that any t codewords of Γ be linearly independent.

Finally, we consider the case where i and j are both even. We shall again show that Y' is separated. If all the $t + 1$ words of Y' are linearly independent, then we are done by the first part of the proof. By assumption, we know that any $t - 1$ words are linearly independent. This gives two cases to consider:

1. $\mathbf{c}'_{t+1-j}$ is the sum of the t first words, which are linearly independent.

2. $\mathbf{c}'_{t-j}$ is the sum of the $t-1$ first words and $\mathbf{c}'_{t+1-j}$ is independent of the others.

Let Y', Y, and X be defined as before. Consider the first case first. Any $t-1$ non-zero words of Y are linearly independent, while all the t non-zero words sum to $\mathbf{0}$. Hence, the only linear independence found between the elements of X is that

$$\mathbf{0} = \mathbf{b}_1 + \ldots + \mathbf{b}_{t-j} + \mathbf{a}_2 + \ldots + \mathbf{a}_j, \tag{1}$$

where $\mathbf{b}_i = \mathbf{c}_i + \mathbf{a}_1$. It follows that the $t-1$ first words of X intersect, since C is t-wise intersecting. Thus there is a position l, where $\mathbf{a}_i$ is 1 for $i = 1, \ldots, j-1$ and $\mathbf{c}_{i'}$ is zero for $i' = 1, \ldots, t-j$. Furthermore, $\mathbf{a}_j$ is one in position l by (1). Hence Y is separated.

In the second case, we get that the t non-zero words of Y are linearly independent. Thus the result follows like the first part of the proof.

It is perhaps not obvious how these propositions may be used to construct non-linear separating codes with a reasonable rate. The following lemma [11] does the trick.

Lemma 2. *Given an $[n, rm]$ linear, binary code C, we can extract a non-linear subcode Γ of size 2^r such that any $2m$ non-zero codewords are linearly independent.*

Proof. Let C' be the $[2^r-1, 2^r-1-rm, 2m+1]$ BCH code. The columns of the parity check matrix of C' make a set Γ' of 2^r-1 vectors from $\mathsf{GF}(2)^{rm}$, such that any $2m$ of them are linearly independent. Now there is an isomorphism $\phi : \mathsf{GF}(2)^{rm} \to C$, so let $\Gamma = \phi(\Gamma') \cup \{\mathbf{0}\}$.

There is a sufficient condition for intersecting codes, resembling the results we have for separating codes in Proposition 1 and Theorems 1 and 1 [8].

Proposition 4. *Let C be a binary linear code. Any t independent codewords intersect in at least $d_1 - m_1(1-2^{1-t})$ coordinate positions.*

Remark 2. The code C has t-wise intersection weight exactly $d_1 - m_1(1-2^{1-t})$ if and only if there are subcodes $D_0 \subseteq D_1 \subseteq C$ such that D_0 has dimension $t-1$ and contains $2^{t-1}-1$ words of maximum weight, and D_1 has dimension t containing 2^{t-1} words of minimum weight.

4 Kasami codes

Let T_m denote the Froebenius trace from $\mathsf{GF}(q^m)$ to $\mathsf{GF}(q)$, defined as

$$T_m(x) = \sum_{i=0}^{m-1} x^{q^i}.$$

It is well-known that

$$\begin{aligned} T_m(x+y) &= T_m(x) + T_m(y), \\ T_m(x) &= T_m(x^q), \end{aligned}$$

and if x runs through $\mathsf{GF}(q^m)$, then $T_m(x)$ takes each value in $\mathsf{GF}(q)$ exactly q^{m-1} times. The original Kasami code is a binary code, so let $q = 2$ and write $Q = 2^m$.

Definition 4 (The Kasami Codes). *The* $[2^{2m} - 1, 3m, 2^{2m-1} - 2^{m-1}]$ *Kasami code is the set*

$$\mathcal{K}_m = \left\{\mathbf{c}(a,b) : a \in \mathsf{GF}(Q^2), b \in \mathsf{GF}(Q)\right\},$$

where

$$\mathbf{c}(a,b) = \left(T_{2m}(ax) + T_m(bx^{Q+1}) : x \in \mathsf{GF}(Q^2)^*\right).$$

The Kasami codes have three different non-zero weights, given by the following lemma [16].

Lemma 3. *The weight of a codeword* $\mathbf{c}(a,b) \in \mathcal{K}_m$ *is given by*

$$w(\mathbf{c}(a,b)) = \begin{cases} 2^{2m-1} - 2^{m-1}, & \text{if } b \neq 0 \text{ and } T_m(a^{Q+1}/b) = 1, \\ 2^{2m-1} + 2^{m-1}, & \text{if } b \neq 0 \text{ and } T_m(a^{Q+1}/b) = 0, \\ 2^{2m-1}, & \text{if } b = 0 \text{ and } a \neq 0, \\ 0, & \text{if } b = 0 \text{ and } a = 0. \end{cases} \tag{2}$$

Using Proposition 4, we get the following result.

Proposition 5. *The Kasami code* $\mathcal{K}_m$ *is* m*-wise intersecting, and its* t*-wise intersection weight is at least*

$$\ell_t(\mathcal{K}_m) \geq 2^m(2^{m-t} - 1) + 2^{m-t}.$$

This implies that the $\mathcal{K}_m$ is a $(2,1)$-SS for $m \geq 2$ and a $(2,2)$-SS for $m \geq 3$. For $t = 2$, the above bound is tight as the following proposition shows. It can be shown by exhaustive search that the bound is tight for $t = m = 3$ as well, but it is an interesting open problem whether the bound is tight in general.

Proposition 6. *The Kasami code $\mathcal{K}_m$ has $(2,1)$-separating weight*

$$\theta_{2,1} = \max\{0, (2^m - 3)2^{m-2}\}.$$

Proof. Recall that $\theta_{2,1} \geq d_1 - m_1/2 = (2^m - 3)2^{m-2}$. This is negative if and only if $m = 1$. Observe that $\mathcal{K}_1$ contains all words of length 3, and thus has $\theta_{2,1} = 0$, as required.

If $m \geq 2$, we get that $\theta_{2,1} > 0$, and by Remark 1 it remains to prove that there are are two codewords $\mathbf{a}$ and $\mathbf{b}$ of minimum weight such that $\mathbf{a} + \mathbf{b}$ has maximum weight. This is fulfilled for $\mathbf{a} = \mathbf{c}(\gamma b, b^2)$ and $\mathbf{b} = \mathbf{c}(\gamma b, f^2 b^2)$ if $T_m(\gamma^{Q+1}) = T_m(\gamma^{Q+1}/f^2) = 1$. Such an f exists as long as $m \geq 2$.

5 BCH codes

Several good (t,u)-separating codes may be constructed from intersecting codes and columns from the parity check matrices of BCH codes. In the tables at the end of this section, we use non-linear subcodes of dual BCH codes as inner codes.

5.1 Finite constructions of intersecting codes

The intersecting properties of the duals of 2-error-correcting BCH codes were first pointed out in [8]. In the sequel, we describe the intersecting properties of arbitrary dual BCH codes.

In MacWilliams and Sloane [20], we find the following lemma.

Lemma 4. *Let C be a BCH code of length $2^m - 1$ and designed distance $d' = 2e + 1$, where $2e - 1 < 2^{\lceil m/2 \rceil} + 1$. For any non-zero words in $C^{\perp}$, the weight w lies in the range*

$$2^{m-1} - (e-1)2^{m/2} \leq w \leq 2^{m-1} + (e-1)2^{m/2}.$$

By using Proposition 4, we get the following result.

Proposition 7. *The dual of a* $[2^m - 1, me]$ *BCH code with designed distance* $d' = 2e + 1$ *has* t*-wise intersection weight*

$$\begin{aligned}\ell_t &\geq 2^{m-t} + (e-1)2^{m/2+1-t} - (e-1)2^{m/2+1} \\ &= 2^{m/2+1}\big(2^{m/2-t-1} - (e-1)(1-2^{-t})\big).\end{aligned}$$

Corollary 2. *The dual of the* e*-error-correcting BCH code with parameters* $[2^m - 1, me]$, *is* t*-wise intersecting if*

$$m > 2\big(1 + \log(e-1) + \log(2^t - 1)\big).$$

The bounds in Lemma 4 are not necessarily tight, and for $e = 2, 3$, the exact maximum and minimum weights are known [17].

Lemma 5. *Let* C *be a 2-error-correcting BCH code of length* $2^m - 1$. *Then*

$$\begin{aligned}d_1 &= 2^{m-1} - 2^{\lfloor m/2 \rfloor}, \\ m_1 &= 2^{m-1} + 2^{\lfloor m/2 \rfloor}.\end{aligned}$$

Proposition 8. *The dual of the* 2*-error-correcting BCH code with parameters* $[2^{2t+1} - 1, 4t + 2, 2^{2t} - 2^t]$, *is* t*-wise intersecting, with intersecting weight* $\ell_t \geq 2$.

This proposition is a direct consequence of the preceding lemma [8].

Lemma 6. *Let* C *be a 3-error-correcting BCH code of length* $2^m - 1$ *for* $m \geq 4$. *Then*

$$\begin{aligned}d_1 &= 2^{m-1} - 2^{\lceil m/2 \rceil}, \\ m_1 &= 2^{m-1} + 2^{\lceil m/2 \rceil}.\end{aligned}$$

Proposition 9. *The punctured dual of the* 3*-error-correcting BCH code with parameters* $[2^{2t+2} - 1, 6t + 6]$, *is* t*-wise intersecting, with intersecting weight* $\ell_t \geq 4$.

5.2 Infinite families of intersecting codes

The following lemma was found in [8].

Lemma 7. *Let* C_1 *be an* $[n_1, k_1, d_1]_q$ *code with* $q = 2^{k_2}$ *and minimum distance* $d_1 > n_1(1 - 2^{1-t})$. *Let* C_2 *be an* $[n_2, k_2, d_2]$ *binary* t*-wise intersecting code. Then the concatenation* $C_1 \circ C_2$ *is a binary* t*-wise intersecting* $[n_1 n_2, k_1 k_2, d_1 d_2]$ *code.*

Lemma 8. *There are constructive infinite sequences of t-wise intersecting binary codes with rates arbitrarily close to*

$$R_t^{(2)} = \left(2^{1-t} - \frac{1}{2^{2t+1}-1}\right)\frac{2t+1}{2^{2t}-1} = 2^{2-3t}(t+o(t)),$$
$$R_t^{(3)} = \left(2^{1-t} - \frac{1}{2^{3t+3}-1}\right)\frac{3t+3}{2^{2t+1}-2} = 2^{-3t}(3t+o(t)).$$

Proof. By concatenating geometric $[N,K,D]_q$ codes from Theorem 2 satisfying $D > N(1-2^{1-t})$ with $q = 2^{4t+2}$, and with a rate arbitrarily close to $2^{1-t} - 1/(\sqrt{q}-1)$, with the $[2^{2t+1}-2, 4t+2, 2^{2t}-2^t-1]$ code of Proposition 8, we obtain the result.

5.3 Constructions of separating codes

There are two basic techniques for constructing asymptotic separating codes from intersecting codes.

Technique I uses a finite intersecting $[n,k]$ code C' as a seed. Then a non-linear subcode is extracted from C' to form an separating inner code C_I. Finally, C_I is concatenated with a separating Tsfasman code C_O. The rate is given by

$$R^{\mathrm{I}} = \frac{\log Q}{n}\left(\frac{1}{uv} - \frac{1}{\sqrt{Q}-1}\right), \tag{3}$$

where $Q = q^2 \leq 2^{2k/\bar{\iota}(u,v)}$ is as large as possible with q a prime power.

Technique II uses a finite intersecting code C_I as a seed, which is concatenated with a Tsfasman code with minimum distance at least $(1-2^{1-t})$ in concordance with Lemma 7, to form an asymptotic intersecting code C'. The asymptotic separating code is a non-linear subcode of C'. The resulting rate is

$$R^{\mathrm{II}} = \frac{k}{n}\left(2^{2-u-v} - \frac{1}{2^{k/2}-1}\right)\frac{2}{\bar{\iota}(u,v)}, \tag{4}$$

provided k is even. Otherwise 2^k is replaced by $Q = q^2 \leq 2^k$ where q is the largest possible prime power.

Comparing (3) and (4), we see that the difference is in the parenthesised expression. Except when $t \leq 3$, we have $2^{2-u-v} > 1/uv$ which tend to give Technique I a better rate. However Technique II uses a larger alphabet for the outer code, which tends to decrease the penalty factor and hence improve the rate of the outer code.

The following proposition gives the rates obtained when Technique II is applied on the duals of BCH(2) and BCH(3). It is easy to check that $R_2^{\mathrm{II}} > R_3^{\mathrm{II}}$ (except for the degenerate case $u = v = 1$).

Proposition 10. *There are constructive infinite sequences of binary (u, v)-separating codes of rate*

$$\begin{aligned} R_2^{II}(u,v) &= \frac{4(u+v)-2}{\bar{\iota}(u,v)\left(2^{2(u+v-1)}-1\right)}\left(2^{2-u-v}-\frac{1}{2^{2(u+v)-1}-1}\right) \\ &\geq 2^{-3(u+v-2)}(1+o(1)), \\ R_3^{II}(u,v) &= \frac{3(u+v)}{\bar{\iota}(u,v)(2^{2(u+v-1)}-1)}\left(2^{2-u-v}-\frac{1}{2^{3(u+v)}-1}\right). \end{aligned}$$

(u,v)	m	$[n,k]$	K	inner rate	outer rate	total rate
$(2,2)$	7	$[126,14]$	14	$1.111 \cdot 10^{-1}$	$2.421 \cdot 10^{-1}$	$2.690 \cdot 10^{-2}$
$(2,3)$	9	$[510,18]$	9	$1.666 \cdot 10^{-2}$	$1.111 \cdot 10^{-1}$	$1.851 \cdot 10^{-3}$
$(2,4)$	11	$[2046,22]$	11	$5.304 \cdot 10^{-3}$	$1.012 \cdot 10^{-1}$	$5.367 \cdot 10^{-4}$
$(2,5)$	13	$[8190,26]$	8	$9.768 \cdot 10^{-4}$	$3.333 \cdot 10^{-2}$	$3.256 \cdot 10^{-5}$
$(3,3)$	11	$[2046,22]$	7	$3.382 \cdot 10^{-3}$	$1.111 \cdot 10^{-2}$	$3.757 \cdot 10^{-5}$
$(3,4)$	13	$[8190,26]$	8	$9.768 \cdot 10^{-4}$	$1.667 \cdot 10^{-2}$	$1.628 \cdot 10^{-5}$
$(4,4)$	15	$[32766,30]$	10	$3.052 \cdot 10^{-4}$	$3.024 \cdot 10^{-2}$	$9.230 \cdot 10^{-6}$

Table 1. Good (u, v)-SS from BCH(2).

(u,v)	m	$[n,k]$	K	inner rate	outer rate	total rate
$(2,2)$	8	$[252,24]$	24	$9.524 \cdot 10^{-2}$	$2.498 \cdot 10^{-1}$	$2.379 \cdot 10^{-2}$
$(2,3)$	10	$[1020,30]$	15	$1.471 \cdot 10^{-2}$	$1.611 \cdot 10^{-1}$	$2.369 \cdot 10^{-3}$
$(2,4)$	12	$[4092,36]$	18	$4.399 \cdot 10^{-3}$	$1.230 \cdot 10^{-1}$	$5.412 \cdot 10^{-4}$
$(2,5)$	14	$[16380,42]$	14	$8.547 \cdot 10^{-4}$	$9.213 \cdot 10^{-2}$	$7.874 \cdot 10^{-5}$
$(3,3)$	12	$[4092,36]$	12	$2.933 \cdot 10^{-3}$	$9.524 \cdot 10^{-2}$	$2.793 \cdot 10^{-4}$
$(3,4)$	14	$[16380,42]$	14	$8.547 \cdot 10^{-4}$	$7.546 \cdot 10^{-2}$	$6.450 \cdot 10^{-5}$
$(3,5)$	16	$[65532,48]$	12	$1.831 \cdot 10^{-4}$	$5.079 \cdot 10^{-2}$	$9.301 \cdot 10^{-6}$
$(4,4)$	16	$[65532,48]$	16	$2.442 \cdot 10^{-4}$	$5.858 \cdot 10^{-2}$	$1.430 \cdot 10^{-5}$
$(4,5)$	18	$[262140,54]$	13	$4.941 \cdot 10^{-5}$	$3.864 \cdot 10^{-2}$	$1.909 \cdot 10^{-6}$
$(5,5)$	20	$[1048572,60]$	12	$1.144 \cdot 10^{-5}$	$2.413 \cdot 10^{-2}$	$2.761 \cdot 10^{-7}$

Table 2. Good (u, v)-SS from BCH(3).

For Technique I, we do not obtain such nice closed form formulæ, because we do not have a nice expression for the alphabet size Q of the outer code.

In Table 1 and 2, we present some good separating codes from duals of BCH(2) and BCH(3) with Technique I. The constructions with BCH(2) are known from [12,13], while the BCH(3)-constructions are new. The symbol K denotes the log-cardinality of the inner code. For big u and v, the inner code resulting from BCH(2) are so small that we do not get a positive rate for the outer code. This could have been amended by increasing m, but better results are obtained by increasing e. Therefore these codes are omitted from Table 1.

	Technique II		Technique I		
(u,v)	BCH(2)	BCH(3)	BCH(2)	BCH(3)	BCH(5)
$(2,2)$	$2.690 \cdot 10^{-2}$	$2.379 \cdot 10^{-2}$	$2.690 \cdot 10^{-2}$	$2.379 \cdot 10^{-2}$	
$(2,3)$	$2.171 \cdot 10^{-3}$	$1.838 \cdot 10^{-3}$	$1.851 \cdot 10^{-3}$	$2.369 \cdot 10^{-3}$	$2.026 \cdot 10^{-4}$
$(2,4)$	$3.334 \cdot 10^{-4}$	$2.749 \cdot 10^{-4}$	$5.367 \cdot 10^{-4}$	$5.412 \cdot 10^{-4}$	$4.045 \cdot 10^{-5}$
$(2,5)$	$3.294 \cdot 10^{-5}$	$2.671 \cdot 10^{-5}$	$3.256 \cdot 10^{-5}$	$7.874 \cdot 10^{-5}$	$6.324 \cdot 10^{-6}$
$(3,3)$	$2.223 \cdot 10^{-4}$	$1.833 \cdot 10^{-4}$	$3.757 \cdot 10^{-5}$	$2.793 \cdot 10^{-4}$	$2.396 \cdot 10^{-5}$
$(3,4)$	$3.294 \cdot 10^{-5}$	$2.671 \cdot 10^{-5}$	$1.628 \cdot 10^{-5}$	$6.450 \cdot 10^{-5}$	$5.270 \cdot 10^{-6}$
$(3,5)$	$3.570 \cdot 10^{-6}$	$2.861 \cdot 10^{-6}$	0	$9.301 \cdot 10^{-6}$	$8.269 \cdot 10^{-7}$
$(4,4)$	$4.759 \cdot 10^{-6}$	$3.815 \cdot 10^{-6}$	$9.230 \cdot 10^{-6}$	$1.430 \cdot 10^{-5}$	$1.105 \cdot 10^{-6}$
$(4,5)$	$5.062 \cdot 10^{-7}$	$4.023 \cdot 10^{-7}$	0	$1.909 \cdot 10^{-6}$	$1.669 \cdot 10^{-7}$
$(5,5)$	$5.660 \cdot 10^{-8}$	$4.470 \cdot 10^{-8}$	0	$2.761 \cdot 10^{-7}$	$2.969 \cdot 10^{-8}$

Table 3. Some good (u,v)-separating codes from duals of BCH codes.

In Table 3, we can compare the constructions for Techniques I and II. Technique I using BCH(3) gives the best rate in all cases studied except for $(2,2)$-SS. However, it is interesting to note that Technique II gives a non-zero rates for some seed codes which give zero rate with Technique I.

We have not presented any results using BCH(4). It is easy to check that when $t \geq 5$, the minimum required value of m, according to Corollary 2, is the same for $e = 4$ and $e = 5$. Consequently, there is no reason for using duals of BCH(4) when $t \geq 5$; using BCH(5) instead can only improve the rate. It can also be checked that BCH(4) is inferior to BCH(5) in the other cases.

The minimum value of m is $2t + 1$ for BCH(2). It increases only by 1 to $2t + 2$ for BCH(3). Moving to BCH(4), m must make a jump by 3 or more, depending on the value of t. This is of course because the bounds on d_1 and m_1 are much worse for BCH(e) when $e > 3$. It explains why the rates for the inner codes as well as for the outer codes in Tables 1 and 2 are so close together. The big increase needed in m from BCH(3) to BCH(4) is only worthwhile when the rate of the outer code is only small fraction of the ideal rate $1/uv$. For $t, u \leq 5$, BCH(3) performs very well, and BCH(5) cannot improve the overall rate. However, for $(7, 9)$-SS we would not get a positive rate using BCH(3), but BCH(5) does the trick.

6 Conclusion

We have shown that Kasami codes and BCH codes have certain separating properties, and that they can be used to construct record breaking families of separating codes. We only have lower bounds on the t-wise intersection weights for $t > 2$. It would be interesting to find the exact intersection weights, and if the bounds are not tight, the constructed rates may be slightly improved.

The fingerprinting schemes of [2] uses (t, t)-SS as components. The present constructions with improved rates for (t, t)-SS, will thus make it possible to build fingerprinting schemes with better rates as well.

7 Acknowledgement

The authors wish to thank prof. Gérard Cohen for many a useful conversation on fingerprinting and separation.

References

1. N. Alon. Explicit construction of exponential sized families of k-independent sets. *Discrete Math.*, 58(2):191–193, 1986.
2. A. Barg, G. R. Blakley, and G. A. Kabatiansky. Digital fingerprinting codes: Problem statements, constructions, identification of traitors. *IEEE Trans. Inform. Theory*, 49(4):852–865, 2003.
3. A. Barg, G. Cohen, S. Encheva, G. Kabatiansky, and G. Zémor. A hypergraph approach to the identifying parent property. *SIAM J. Disc. Math.*, 14(3):423–431, 2001.
4. Dan Boneh and James Shaw. Collusion-secure fingerprinting for digital data. *IEEE Trans. Inform. Theory*, 44(5):1897–1905, 1998. Presented in part 1995, see Springer LNCS.
5. Bella Bose and T. R. N. Rao. Separating and completely separating systems and linear codes. *IEEE Trans. Comput.*, 29(7):665–668, 1980.
6. B. Chor, A. Fiat, and M. Naor. Tracing traitors. In *Advances in Cryptology - CRYPTO '94*, volume 839 of *Springer Lecture Notes in Computer Science*, pages 257–270. Springer-Verlag, 1994.
7. B. Chor, A. Fiat, M. Naor, and B. Pinkas. Tracing traitors. *IEEE Trans. Inform. Theory*, 46(3):893–910, May 2000.

8. Gérard Cohen and Gilles Zémor. Intersecting codes and independent families. *IEEE Trans. Inform. Theory*, 40:1872–1881, 1994.
9. Gérard D. Cohen and Sylvia B. Encheva. Efficient constructions of frameproof codes. *Electronics Letters*, 36(22), 2000.
10. Gérard D. Cohen, Sylvia B. Encheva, Simon Litsyn, and Hans Georg Schaathun. Erratum to 'intersecting codes and separating codes'. *Discrete Applied Mathematics*, 2003. Submitted.
11. Gérard D. Cohen, Sylvia B. Encheva, Simon Litsyn, and Hans Georg Schaathun. Intersecting codes and separating codes. *Discrete Applied Mathematics*, 128(1):75–83, 2003.
12. Gérard D. Cohen, Sylvia B. Encheva, and Hans Georg Schaathun. On separating codes. Technical report, Ecole Nationale Supérieure des Télécommunications, 2001.
13. Gérard D. Cohen and Hans Georg Schaathun. Asymptotic overview on separating codes. Available at `http://www.ii.uib.no/publikasjoner/texrap/index.shtml`, May 2003.
14. P. Delsarte and J.-M. Goethals. Tri-weight codes and generalized Hadamard matrices. *Information and Control*, 15:196–206, 1969.
15. A. D. Friedman, R. L. Graham, and J. D. Ullman. Universal single transition time asynchronous state assignments. *IEEE Trans. Comput.*, 18:541–547, 1969.
16. Tor Helleseth and P. Vijay Kumar. The weight hierarchy of the Kasami codes. *Discrete Math.*, 145(1-3):133–143, 1995.
17. Tadao Kasami. Weight distributions of Bose-Chaudhuri-Hocquenghem codes. In *Combinatorial Mathematics and its Applications (Proc. Conf., Univ. North Carolina, Chapel Hill, N.C., 1967)*, pages 335–357. Univ. North Carolina Press, Chapel Hill, N.C., 1969.
18. J. Körner and G. Simonyi. Separating partition systems and locally different sequences. *SIAM J. Discrete Math.*, 1:355–359, 1988.
19. János Körner. On the extremal combinatorics of the Hamming space. *J. Combin. Theory Ser. A*, 71(1):112–126, 1995.
20. F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes.* North-Holland, Amsterdam, 1977.
21. Yu. L. Sagalovich. Separating systems. *Problems of Information Transmission*, 30(2):105–123, 1994.
22. Hans Georg Schaathun. Fighting two pirates. In *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, volume 2643 of *Springer Lecture Notes in Computer Science*, pages 71–78. Springer-Verlag, May 2003.
23. Jessica N. Staddon, Douglas R. Stinson, and Ruizhong Wei. Combinatorial properties of frameproof and traceability codes. *IEEE Trans. Inform. Theory*, 47(3):1042–1049, 2001.
24. D. R. Stinson and R. Wei. Combinatorial properties and constructions of traceability schemes and frameproof codes. *SIAM J. Discrete Math.*, 11(1):41–53 (electronic), 1998.
25. D.R. Stinson, Tran Van Trung, and R. Wei. Secure frameproof codes, key distribution patterns, group testing algorithms and related structures. *J. Stat. Planning and Inference*, 86(2):595–617, 2000.
26. Michael A. Tsfasman. Algebraic-geometric codes and asymptotic problems. *Discrete Appl. Math.*, 33(1-3):241–256, 1991. Applied algebra, algebraic algorithms, and error-correcting codes (Toulouse, 1989).
27. S. H. Unger. *Asynchronous Sequential Switching Circuits.* Wiley, 1969.