10-13-2005

Big Cwatsets and Hamming Code

Matthew Davis
Rose-Hulman Institute of Technology

Thomas M. Langley
Rose-Hulman Institute of Technology, langley@rose-hulman.edu

Norah Mazel
Rose-Hulman Institute of Technology

BIG CWATSETS AND HAMMING CODES

Matthew Davis, Thomas Langley and Norah Mazel

MS TR 05-03

October 13, 2005

Department of Mathematics
Rose-Hulman Institute of Technology
http://www.rose-hulman.edu/Class/ma/

FAX: (812) 877-8883 **PHONE: (812) 877-8391**

BIG CWATSETS AND HAMMING CODES

MATTHEW DAVIS, THOMAS LANGLEY, AND NORAH MAZEL

ABSTRACT. In contrast to Lagrange's Theorem in finite group theory, we show that the ratio of the largest proper cwatset of degree d to the size of binary d-space approaches 1 as d approaches infinity. We show how to explicitly construct large cwatsets as cosets of Hamming codes, and discuss many open questions that arise.

1. INTRODUCTION

Cwatsets are sets of binary words with algebraic structure that are not quite additive groups. An example is the set $F = \{000, 110, 101\}$. Note that F is not closed under binary addition since, for example, $110 + 101 = 011$. So F is not a subgroup of binary 3-space. But

$$F + 110 = \{110, 000, 011\} = F^{(1,2)}$$
$$F + 101 = \{101, 011, 000\} = F^{(1,3)}$$

where, for example, $F^{(1,2)}$ is the set F with the first and second bits of each word interchanged. So F is **c**losed **w**ith **a** **t**wist, leading to the following definition.

Definition 1.1. *Let C be a subset of binary d-space and let S_d denote the symmetric group on d symbols. Then C is a cwatset of degree d if for each $\mathbf{c} \in C$ there exists $\sigma \in S_d$ such that $C + \mathbf{c} = C^\sigma$.*

For a given word $\mathbf{c}$ the permutation σ need not be unique. Indeed, $F^{(1,2)} = F^{(1,2,3)}$. Also, any subgroup G of binary d-space is a cwatset since $G + \mathbf{g} = G = G^{id}$ for any $\mathbf{g}$ in G where id is the identity permutation in S_d.

Hardigan [4] introduced the set F in the late 1960s as a tool for bootstrapping confidence intervals for the mean of a random variable. This set was rediscovered in the late 1980s for the same purpose by Erich Friedman, a student at Rose-Hulman Institute of Technology working with professor Gary Sherman. The above defining condition followed shortly [7], spawning a rich algebraic and combinatorial theory of cwatsets, with interesting applications to group theory and the theory of graphs and hypergraphs [1, 2, 3, 5]. Many of the main results are due to the work of undergraduates, both at Rose-Hulman's NSF sponsored REUs and elsewhere.

A natural question to ask when considering any algebraic structure is how big can proper substructures be? For example, Lagrange's Theorem prevents a proper subgroup of a finite group from containing more than half the elements of the group. The goal of this

Date: 10.13.2005.

paper is to show that something decidedly different happens with cwatsets. Define a proper cwatset to be a cwatset that is not all of binary d-space (which is itself a cwatset). Then we will prove the following.

Theorem 1.2. *Suppose C is a proper cwatset of degree d of maximal size. Set $\rho_d = |C|/2^d$. Then ρ_d approaches 1 as d approaches infinity.*

In particular the following have been established (see [6]:)

d	2	3	4	5	6	7
ρ_d	1/2	3/4	3/4	3/4	3/4	7/8

The jumps in ρ_d at degrees 3 and 7 suggest that something interesting is happening at degrees $2^n - 1$. We will prove Theorem 1.2 by showing that the essence of the above pattern holds in general, specifically that $\rho_{2^n-1} \geq \frac{2^n-1}{2^n}$. This coupled with the fact that ρ_d is a nondecreasing function of d will give the result. We will demonstrate explicitly how to construct the large cwatsets that realize these ratios, leading to an unexpected connection with algebraic coding theory.

On the way to building these large cwatsets, we start in Section 2 with a crucial definition and by showing that the ratio ρ_d is nondecreasing. We continue in Section 3 by showing how to double the size of a cwatset by adding a binary string to every element of the cwatset. We then generalize this construction to allow us to add a group instead of just a single string in Section 4. With this construction in hand, the game of building large cwatsets comes down to finding the right large group, and it is here that the connection to coding theory arises. The right groups turn out to be Hamming codes, and we discuss their use in building cwatsets in Section 5. Finally, in Section 6 we discuss some related questions that arise from our construction of large cwatsets. We will pose a number of open questions along the way.

2. The ratio ρ_d is nondecreasing

To show that the ratio ρ_d is a nondecreasing function of d, we need a definition and some notation. Since the structure of a cwatset lies not only in the binary words but also in the associated permutations, the set of all word-permutation pairs associated with a cwatset is well-studied.

Definition 2.1. *For a cwatset C of degree d, the* maximal covering group *of C is the set*

$$M_C = \{(\sigma, \mathbf{c}) \in S_d \times Z_2^d \mid C + \mathbf{c} = C^\sigma\}$$

where Z_2^d is the set of all binary words of length d.

For example, M_F is the set

$$\{(id, 000), ((2,3), 000), ((1,2), 110), ((1,2,3), 110), ((1,3), 101), ((1,3,2), 101)\}.$$

Calling M_C the maximal covering group is no accident as there is indeed a group structure on the set $S_d \times Z_2^d$ that is suggested by the defining condition $C + \mathbf{c} = C^\sigma$. We can rewrite this as

$$C^\sigma + \mathbf{c} = C,$$

suggesting a possible action of pairs $(\sigma, \mathbf{c})$ on sets of binary strings, defined by

$$C^{(\sigma,\mathbf{c})} = C^{\sigma} + \mathbf{c}.$$

If we apply this operation twice, we obtain

$$\begin{aligned}(C^{(\sigma,\mathbf{c})})^{(\tau,\mathbf{d})} &= (C^{\sigma} + \mathbf{c})^{(\tau,\mathbf{d})} \\ &= (C^{\sigma} + \mathbf{c})^{\tau} + \mathbf{d} \\ &= (C^{\sigma})^{\tau} + \mathbf{c}^{\tau} + \mathbf{d} \\ &= C^{\sigma\tau} + \mathbf{c}^{\tau} + \mathbf{d} \\ &= C^{(\sigma\tau,\mathbf{c}^{\tau}+\mathbf{d})}\end{aligned}$$

where we read products of permutations from left to right so that $\sigma\tau$ is the permutation obtained by first applying σ and then τ. This now suggests the group structure on $S_d \times Z_2^d$, since for our operation to be a group action we would like to have

$$(C^{(\sigma,\mathbf{c})})^{(\tau,\mathbf{d})} = C^{(\sigma,\mathbf{c})(\tau,\mathbf{d})}.$$

So we set $(\sigma, \mathbf{c})(\tau, \mathbf{d}) = (\sigma\tau, \mathbf{c}^{\tau} + \mathbf{d})$. It is routine to check that the set $S_d \times Z_2^d$ with this binary operation is indeed a group, called the *wreath product* of S_d by Z_2 and denoted $S_d \wr Z_2$ (this group is also called the *hyperoctahedral group* or the group of *signed permutations*). The set M_C is then the stabilizer of C under the action defined above, and is therefore a subgroup of $S_d \wr Z_2$. The cwatset C is covered by M_C in the sense that every element of C is contained as a component of an element of M_C and M_C is called the maximal covering group since it contains all possible pairs $(\sigma, \mathbf{c})$ that respect the defining cwatset condition.

We are now ready to show that ρ_d is nondecreasing.

Proposition 2.2. *If ρ_d is as defined in Theorem 1.2, then $\rho_{d+1} \geq \rho_d$ for all d.*

Proof. For a cwatset C, define $C \times Z_2$ to the set obtained by appending both a 0 and a 1 to each word in F. So for example,

$$F \times Z_2 = \left\{ \begin{array}{cc} 0000 & 0001 \\ 1100 & 1101 \\ 1010 & 1011 \end{array} \right\}$$

It is routine to check that if a permutation σ in S_3 is paired with a word $\mathbf{c}$ in M_F, then the natural embedding of σ in S_4 (where 4 is a fixed point) will pair with the two words obtained by appending 0 and 1 to $\mathbf{c}$ in $M_{F\times Z_2}$. So $F \times Z_2$ is a cwatset. Similarly, so is $C \times Z_2$ for any cwatset C. So we can always double the size of a cwatset by increasing the degree by 1. Therefore if C is a cwatset that realizes ρ_d, then

$$\rho_{d+1} \geq \frac{|C \times Z_2|}{2^{d+1}} = \frac{2|C|}{2^{d+1}} = \frac{|C|}{2^d} = \rho_d.$$

3. Doubling the size of a cwatset

We now begin our quest to construct large cwatsets of a given degree. We start with the natural question: can we build large cwatsets from smaller cwatsets? Fortunately the answer is yes, and in a very natural way.

Example 3.1. Let's again start with $F = \{000, 110, 101\}$. To build a new cwatset (of degree 3), add the word 111 to each word in F and take the union of F and these new words:

$$W = \{000, 111, 110, 001, 101, 010\}.$$

It is routine to check that W is a cwatset, and that the word-permutation pairs in its maximal covering group are

words	permutations
$000, 111$	$id, (2,3)$
$110, 001$	$(1,2), (1,2,3)$
$101, 010$	$(1,3), (1,3,2)$

(so 000 and 111 are both paired with id and $(2,3)$ in M_W). So we can double the size of F and obtain a new cwatset by adding 111. However, adding an arbitrary word does not result in a cwatset in general (try adding 001 for instance). Is 111 the only word that we can use? Well for F, the answer is yes. But the answer in general is no. So let's examine the situation a little closer to see precisely what is special about 111. The key lies in the structure suggested by the above table. After adding 111 to F, we view W as a union of cosets of the group $G = \{000, 111\}$ by elements of F. In particular,

$$W = \left\{ \begin{array}{ccc} 000 & 110 & 101 \\ 111 & 001 & 010 \end{array} \right\} = G \cup (G + 110) \cup (G + 101).$$

To confirm that W is a cwatset, we need to pair each word $\mathbf{c}$ in W with a permutation σ such that $W + \mathbf{c} = W^\sigma$. A few observations:

- The cosets of G by elements of F are all distinct because the word that we chose to add, 111, is not in F and is not equal to the sum of any two elements of F. If, for example, we had tried adding $011 = 110 + 101$ instead so that $G = \{000, 011\}$, then 110 would be in both $G + 110$ and $G + 101$.

- The set W is closed under addition by 111, that is, $W + 111 = W$. So 111 is paired with the same permutations in M_W as 000.

- Given a word $\mathbf{c}$ in F, any permutation σ that is paired with $\mathbf{c}$ in M_W will also be paired with $\mathbf{c} + 111$. This follows since

  $$W + (111 + \mathbf{c}) = (W + 111) + \mathbf{c} = W + \mathbf{c}.$$

  So there is a natural pairing of words and their complements by 111 in the maximal covering group of the new cwatset. Note that this pairing does not depend specifically on our choice of 111. Had we used another word to form the new cwatset there would still be a pairing of the original words and their complements by this new word.

Now, exactly what is it about 111 that enables us to find a permutation that pairs with each element of F in the new maximal covering group M_W? Well, from the table above, we see that each permutation that paired with F in M_F also pairs with F in

M_W. To see precisely why this is, let's follow the action of an element of M_F on W, say $((1,2,3),110)$. For emphasis, we'll start by writing W as it was constructed using 111:

$$W = \left\{ \begin{array}{lll} 000 & 110 & 101 \\ 000+111 & 110+111 & 101+111 \end{array} \right\}.$$

Now compute $W+110$ and $W^{(1,2,3)}$ and compare:

$$W+110 = \left\{ \begin{array}{lll} 110 & 000 & 011 \\ 110+111 & 000+111 & 011+111 \end{array} \right\}$$

while

$$\begin{aligned} W^{(1,2,3)} &= \left\{ \begin{array}{lll} 000^{(1,2,3)} & 110^{(1,2,3)} & 101^{(1,2,3)} \\ 000^{(1,2,3)}+111^{(1,2,3)} & 110^{(1,2,3)}+111^{(1,2,3)} & 101^{(1,2,3)}+111^{(1,2,3)} \end{array} \right\} \\ &= \left\{ \begin{array}{lll} 000 & 011 & 110 \\ 000+111 & 011+111 & 110+111 \end{array} \right\}. \end{aligned}$$

Now, the top rows of $W+110$ and $W^{(1,2,3)}$ are equal because $((1,2,3),110)$ is in M_F and the top row of W is just F. The heart of our construction lies in what what makes the bottom rows equal. Drumroll...the bottom rows are equal precisely (and simply) because $111^{(1,2,3)} = 111$. Because 111 is fixed by $(1,2,3)$, the element $((1,2,3),110)$ that first lived in M_F now also lives in M_W. This will be true for any element of M_F since any permutation fixes 111, so every every element can be paired with a permutation in M_W, and therefore W is a cwatset.

One more observation: although 111 is fixed by every permutation that appears in M_F, to guarantee that W is a cwatset it is only necessary that that 111 be fixed by *at least one permutation* paired with each word of F in M_F. With that in mind, our example illustrates the following theorem.

Theorem 3.2. *Suppose C is a cwatset of degree d and $\mathbf{x}$ is a binary word such that*

- *$\mathbf{x}$ is not the sum of any two elements of F,*
- *for each $\mathbf{c}$ in C there is an element $(\sigma, \mathbf{c})$ in M_C such that $\mathbf{x}^\sigma = \mathbf{x}$.*

Set $G = \{\mathbf{0}, \mathbf{x}\}$. Then $\cup_{\mathbf{c}\in C}(G+\mathbf{c})$ is a cwatset whose size is twice that of C.

Note that since the composition of two permutations that fix $\mathbf{x}$ also fixes $\mathbf{x}$, another way of stating Thoerem 3.2 is to say that $\cup_{\mathbf{c}\in C}(G+\mathbf{c})$ is a cwatset if $\mathbf{x}$ is fixed by some covering group of C (but not necessarily all of M_C). In fact, the heart of Theorem 3.2 lies in the fact that if $(\sigma, \mathbf{c})$ is in the maximal covering group of the original cwatset and σ fixes $\mathbf{x}$, then $(\sigma, \mathbf{c})$ is in the maximal covering group of the new cwatset. The converse of this statement is also true, that is, if $(\sigma, \mathbf{c})$ is in the maximal covering group of the original cwatset but σ does not fix $\mathbf{x}$, then $(\sigma, \mathbf{c})$ is not in the maximal covering group of the new cwatset (this is a special case of Theorem 6.2 below). So we emphasize again that all of the original maximal covering group may not go through to the new cwatset. Here's an example.

Example 3.3. Let $F_2 = \{000000, 111100, 110011\}$. This cwatset has the same structure as F if we consider each 2-bit block as a "super bit". To understand the maximal

covering group, consider adding 111100:

$$F_2 + 11\,11\,00 = \{11\,11\,00, 00\,00\,00, 00\,11\,11\}.$$

So for example, just as $(1,2,3)$ is paired with 110 in M_F, we can pair any permutation that cycles the three "super bits" with 111100 in M_{F_2}. The most natural of these is $(1,3,5)(2,4,6)$, but any of the following will work:

$$\begin{array}{cccc} (1,3,5)(2,4,6) & (1,4,5)(2,3,6) & (1,3,5,2,4,6) & (1,4,5,2,3,6) \\ (1,3,6)(2,4,5) & (1,4,6)(2,3,5) & (1,3,6,2,4,5) & (1,4,6,2,3,5). \end{array}$$

Now consider adding 101010 to F_2, obtaining the new set

$$W_2 = \left\{ \begin{array}{ccc} 000000 & 111100 & 110011 \\ 101010 & 010110 & 011001 \end{array} \right\}.$$

This 101010 meets the criteria of Theorem 3.2 since it it is not the sum of two elements of F_2 and is fixed by the permutations in the following covering group of F_2:

$$\{(id, 000000), ((1,3,5)(2,4,6), 111100), ((1,5,3)(2,6,4), 110011)\}.$$

So W_2 is a cwatset, and the elements of this covering group are in M_{W_2}. However, the element $((1,3,6)(2,4,5), 111100)$ of M_{F_2} is not in M_{W_2} since $101010^{(1,3,6)(2,4,5)} = 011001$.

It is also worth noting that the permutation $(1,3,5)(2,4)$ fixes 101010 but does not appear in M_{F_2}, in contrast to Example 3.1 where every permutation in S_3 fixes 111 and every permutation appears in M_W.

The converse of Theorem 3.2 is an open question.

Question 3.4. *If C is a cwatset and $\mathbf{x}$ is a binary word such that $D = C \cup (C + \mathbf{x})$ is a cwatset, then must $\mathbf{x}$ be fixed by a covering group of C?*

There are two components to this question. First, are there elements $(\sigma, \mathbf{a})$ of M_D that are not elements of M_C? If not, then the answer to Question 3.4 is yes. Second, if there are such elements, then can they be the only elements of M_D? That is, can all of the elements of M_C be lost in the transition to M_D?

4. Growing cwatsets by taking cosets of larger groups.

So now we know when we can use a single word $\mathbf{x}$ to double the size of a cwatset. The new cwatset is just the union of cosets of the group $\{\mathbf{0}, \mathbf{x}\}$ by elements of the cwatset. Fortunately, it is easy to generalize our construction to obtain cwatsets that are unions of cosets of larger groups. This will allow us to obtain large enough cwatsets of degrees $d = 2^n - 1$ to show that the ratio of the maximum size of a proper cwatset of degree d to the size of all of binary d-space space goes to 1 as d gets large.

Example 4.1. Start with the following degree 4 version of F: $F_4 = \{0000, 1100, 1010\}$ and let G be the group $\{0000, 1111, 0001, 1110\}$. Then W_4 is the disjoint union of the

cosets of G by elements of F_4:

$$W_4 = \left\{ \begin{array}{ccc} 0000 & 1100 & 1010 \\ 1111 & 0011 & 0101 \\ 0001 & 1101 & 1011 \\ 1110 & 0010 & 0100 \end{array} \right\} = G \cup (G + 1100) \cup (G + 1010).$$

So now we have built a new cwatset by starting with a small cwatset and then adding a group and its cosets. Here are the pairings in M_{W_4}:

words	permutations
$0000, 1111, 0001, 1110$	$id, (2,3)$
$1100, 0011, 1101, 0010$	$(1,2), (1,2,3)$
$1010, 0101, 1011, 0100$	$(1,3), (1,3,2)$

Note the following similarities to Example 3.1:

- The cosets of G by elements of F_4 are all distinct since G does not contain the sum of any two distinct elements of F_4 (and in particular does not contain any elements of F_4 except 0000).

- The set W_4 is closed under addition by elements of G.

- Elements in the same coset of G are paired with the same permutations in M_{W_4}. This follows since if c is in F_4 and $\mathbf{g}$ is in G, then

  $$W_4 + (\mathbf{g} + \mathbf{c}) = (W_4 + \mathbf{g}) + \mathbf{c} = W_4 + \mathbf{c}.$$

  So any permutation σ that is paired with $\mathbf{c}$ in M_{W_4} is also paired with each element of the coset $G + \mathbf{c}$.

- Every element of G is fixed by the permutations that appear as components of elements of M_{F_4}. This is in fact is a stronger condition than we need in general. All that is required to force W_4 to be a cwatset is that for each element $\mathbf{c}$ of F_4, there is an element $(\sigma, \mathbf{c})$ in M_{F_4} such that $G^\sigma = G$. This equality is a set equality so that σ does not necessarily need to fix each element of G, but it must take elements of G to other elements of G. This condition is the necessary generalization of the condition in Theorem 3.2 that $\mathbf{x}^\sigma = \mathbf{x}$.

With these observations, an argument similar to the one preceding Theorem 3.2 leads to the following theorem.

Theorem 4.2. *Let C be a cwatset of degree d and let G be a group in Z_2^d that does not contain the sum of any two distinct elements of C. Set $D = \bigcup_{\mathbf{c} \in C}(G + \mathbf{c})$. Then D is a cwatset if for each $\mathbf{c} \in C$, there exists $(\sigma, \mathbf{c}) \in M_C$ such that $G^\sigma = G$.*

Again, this begs the question of the converse.

Question 4.3. *Can we expand a cwatset C in the same manner as in Theorem 4.2 to obtain a cwatset D, but also have an element $\mathbf{c}$ in C such that in M_D, $\mathbf{c}$ is not paired*

with any of the permutations that it is paired with in M_C? In other words, does M_D have to contain a subgroup of M_C that covers C?

5. Big cwatsets and Hamming codes

So now we've reduced the game of finding big cwatsets to finding the right cwatset to start with and taking the union of cosets of the right big group by elements of the cwatset. For degree 3, the correct cwatset is F, which we can double to produce W, the maximal size cwatset of degree 3. So the correct group is $G = \{000, 111\}$. The next jump in ρ_d appears at degree 7, and it turns out that the correct cwatset to use is a degree 7 generalization of F, and a degree 7 generalization of G. We'll call the cwatset K_7:

$$K_7 = \left\{ \begin{array}{ccccccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right\}.$$

The big question is what is the degree 7 generalization of G? To satisfy the conditions of Theorem 4.2, we need a large group that is fixed under permutations that appear as components of elements of M_{K_7}. Fortunately we have a lot of permutations to choose from. Consider adding 1100000 to K_7:

$$K_7 + 1100000 = \left\{ \begin{array}{ccccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right\} + 1100000 = \left\{ \begin{array}{ccccccc} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{array} \right\}.$$

So adding 1100000 to K_7 moves the the column with 6 ones from the first column to the second column. Therefore if we want $K_7 + 1100000 = K_7^\sigma$, the only restriction on σ is that it must send 1 to 2. So any permutation that sends 1 to 2 will be paired with 1100000 in M_{K_7}. Similarly, any permutation that sends 1 to 3 will be paired with 1010000 and so on. So in fact every permutation in S_7 appears as a component of an element of M_{K_7} (just as every permutation in S_3 appeared in a pairing in M_F). Therefore we need to find a group G that is fixed by at least one permutation that sends 1 to i for any i. In group theory lingo, this is equivalent to requiring the group of permutations that fixes G to be transitive.

So how do we find G? Well, let's review what properties it needs to have. It can't contain the elements of K_7 (except the 0 word), or the sum of any two distinct elements of K_7. So G can't contain any words with two 1s. Such words are often said to have *weight* two. So we need a big group with no weight two words, of degree $2^3 - 1$, which is fixed by a transitive group of permutations. Well, if you happen to mention those qualifications to someone familiar with a little algebraic coding theory, you're likely to

hear the words "Hamming code", which turns out to be precisely the group that we need.

So what is a Hamming code? First of all, conveniently enough for our purposes, Hamming codes are defined for degrees $2^n - 1$. Coding theorist like to talk about the distance between words in a code, which is defined to be the number of bits in which the two words differ. For example, the distance between 101 and 110 is 2, since they differ in the second and third bits. The distance between any two words in a Hamming code is at least 3. Because of this, Hamming codes are said to have minimum distance 3. Since Hamming codes are groups, they contain $\mathbf{0}$, and so Hamming codes do not contain any words of weight 1 or 2, which meets our criteria.

Hamming codes are useful in the theory of error correcting codes in the following way. If elements of a Hamming code are to be sent over a communications line (wireless or otherwise) and a single bit flips due to noise, the resulting word will be closest to a unique word in the Hamming code (since the minimum distance between words is 3) and therefore the original message word can be recovered. So Hamming codes are called 1-error correcting codes.

Here's how you get them. For degree $d = 2^n - 1$, form an $n \times 2^{n-1}$ matrix whose columns are all nonzero binary strings of length n. For example, for degree 7 we might choose

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

The associated Hamming code H is just the null space of this matrix if we write binary words as column vectors. In this way H can be viewed as a vector space over the field Z_2, and is therefore a binary group.

The properties of H that we need are well-known and can easily be deduced from the matrix A. First, for a binary word $\mathbf{c}$ written as a column vector, the product $A\mathbf{c}$ is just a sum of the columns of A that correspond to the positions of the 1s in $\mathbf{c}$. Since the minimum number of columns of A that sum to $\mathbf{0}$ is 3, H contains no words of weight 1 or 2. So in particular, H does not contain the sum of any two distinct elements of K_7.

How big is H? Well, since A has three linearly independent rows, the dimension of its row space, and therefore its column space, is 3. Since A has 7 columns, the dimension of its null space is $7 - 3 = 4$. So H has dimension 4 as a vector space over Z_2, and therefore has cardinality $2^4 = 16$. So taking the union of the 7 cosets of H by elements of K_7 yields a set of size $7 \cdot 16 = 112$, which is 7/8 of the 128 strings of length 7. So if this union is a cwatset, we will have found a cwatset that realizes ρ_7.

So now we just need to verify that for each i, $1 \leq i \leq 7$, we can find a permutation σ that takes 1 to i and such that $H^\sigma = H$. This again follows from A.

Lemma 5.1. *Let H be a Hamming code of length $2^n - 1$ and let G be the group of all permutations σ that satisfy $H^\sigma = H$. Then G is transitive.*

Proof. The proof lies in the structure of the defining matrix A. Permuting columns of A and then taking the null space of the resulting matrix gives a code obtained by permuting the corresponding bits of H. Consider the matrix A above. Let's start with $i = 2$. We would like to modify A so that column 1 ends up in column 2, and do the modification in such a way the the nullspace is unchanged. So all we do is add the first row to the third row, obtaining

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}.$$

Note that this results in a permutation σ of the columns of A that sends the first column to the second column. Since the null space, and therefore H, hasn't changed, we have $H^\sigma = H$. We can do similar row operations to send the first column to columns 3 and 4. To send the first column to columns 5, 6, or 7, we first need to interchange rows one and three, to get a one to the top of the column. For example, here's how to send column 1 to column 7:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \longrightarrow \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Again, the main point is that performing row operations just results in a permutation of the columns, and therefore a permutation of H, but leaves the null space of H unchanged.

The above example generalizes easily to an arbitrary degree $d = 2^n - 1$. The corresponding Hamming code H will have dimension $2^n - 1 - n = d - n$, and therefore size 2^{d-n}. Constructing K_d in the same manner as K_7, and taking cosets of H by elements of K_d will produce a cwatset of size $d(2^{d-n})$. Dividing by the total number of binary strings of length d, we obtain a ratio of

$$\frac{d(2^{d-n})}{2^d} = \frac{d}{2^n} = \frac{2^n - 1}{2^n}.$$

So we have illustrated the following.

Theorem 5.2. *If ρ_d is as in Theorem 1.2, then for $d = 2^n - 1$ we have $\rho_d \geq \frac{2^n-1}{2^n}$.*

This theorem, along with our observation that ρ_d is nondecreasing (Proposition 2.2), finally proves Theorem 1.2: the ratio of the size of the largest cwatset of degree d to the size of binary d-space goes to 1 as d goes to infinity. Again we emphasize that this is in contrast to most algebraic structures, where there is typically a quantum leap between the size of the largest proper substructure and the size of the host structure. There is one caveat that we should mention. We have been careful to avoid using the term *subcwatset*. The reason is that there is more to being a subcwatset than just being a subset that is also a cwatset. The subcwatset structure must necessarily also be reflected in the corresponding maximal covering group, and in fact there is more than one possible definition of subcwatset (see [3]). However, in each definition any cwatset is always a subcwatset of the entire space, so that our result does parallel the idea of

other substructures.

We are still left with some obvious open questions, which we state as conjectures.

Conjecture 5.3. *The cwatsets built from Hamming codes are maximal. That is, we know that at degrees $d = 2^n - 1$ that ρ_d is at least $\frac{2^n-1}{2^n}$. We believe that in fact this is an equality.*

Conjecture 5.4. *The ratio ρ_d changes only at $d = 2^n - 1$.*

6. The identity group and cwatset structure

We will conclude with some observations and questions that arrive from the work of the previous sections. Now that we have a method for constructing larger cwatsets from smaller ones, the next question is, of course, are all big cwatsets constructed in this way? In our examples above, the cwatset elements paired with the identity permutation in the maximal covering group played a central role. In any cwatset, these words form a group. With that in mind, a definition.

Definition 6.1. For a cwatset C, define the *identity group* of C, denoted C_{id}, to be the set of elements of C that are paired with id in M_C (so C is closed under addition by these elements).

With this definition, any cwatset in a sense behaves like our examples above, with C_{id} playing the part of the group G. Specifically, we have the following theorem.

Theorem 6.2. *If C is a cwatset, then*

1. *C is a union of cosets of C_{id}.*
2. *If $(\sigma, \mathbf{a})$ is in M_C, then $(\sigma, \mathbf{b})$ is in M_C if and only if $\mathbf{a}$ and $\mathbf{b}$ are in the same coset of C_{id}.*
3. *If $(\sigma, \mathbf{c})$ is an element of M_C, then $C_{id}^{\sigma} = C_{id}$.*

Proof. For the first statement, let M_{id} be the set of elements of M_C that have id as their first component. Then M_{id} is a subgroup of M_C so M_C is a disjoint union of left cosets of M_{id}. But the projection of a left coset of M_{id} onto binary space will be a coset of C_{id} since

$$(\sigma, \mathbf{a})M_{id} = \{(\sigma, \mathbf{a})(id, \mathbf{c}) \mid (id, \mathbf{c}) \in M_{id}\} = \{(\sigma, \mathbf{a} + \mathbf{c}) \mid \mathbf{c} \in C_{id}\}.$$

For the second statement, the proof follows from our observation preceding Thoerem 4.2. For the third statement, note that if $C_{id} = \{\mathbf{0}\}$, then every permutation fixes C_{id} so the result is trivial. Now suppose $\mathbf{x}$ is in C_{id} and $(\sigma, \mathbf{a})$ is in M_C. Then $(id, \mathbf{x})$ is in M_C, so $(id, \mathbf{x})(\sigma, \mathbf{a}) = (\sigma, \mathbf{x}^{\sigma} + \mathbf{a})$ is also in M_C. But then $\mathbf{a}$ and $\mathbf{x}^{\sigma} + \mathbf{a}$ have to be in the same coset of C_{id}, namely $C_{id} + \mathbf{a}$. So $\mathbf{x}^{\sigma} + \mathbf{a} = \mathbf{y} + \mathbf{a}$ for some $\mathbf{y}$ in C_{id}. Therefore $\mathbf{x}^{\sigma} = \mathbf{y}$ for some $\mathbf{y}$ in C_{id}. So $C_{id}^{\sigma} = C_{id}$ as required.

One note: in each of Examples 3.1, 3.3, and 4.1, the group G is identical to the identity group of the larger cwatset because $\mathbf{0}$ is the only word paired with the identity permutation in the original cwatset. If the original cwatset has a nontrivial identity group, then the group G will be a subgroup of the identity group. Here's an example.

Example 6.3. Consider the cwatset F' that was obtained by appending both a 0 and 1 to each word in F:

$$F' = \{0000, 0001, 1100, 1101, 1010, 1011\}.$$

Then $F'_{id} = \{0000, 0001\}$. If we double F' by adding 1111, we get

$$D = \left\{ \begin{array}{cccccc} 0000 & 0001 & 1100 & 1101 & 1010 & 1011 \\ 1111 & 1110 & 0011 & 0010 & 0101 & 0100 \end{array} \right\}$$

So $D = \cup_{c \in F'}(G + c)$ where G is the group $\{0000, 1111\}$. But because 0001 is also paired with the identity permutation in $M_{F'}$, the identity group of D is $\{0000, 1111, 0001, 1110\}$. So D can also be considered a union of cosets of this identity group:

$$D = \left\{ \begin{array}{ccc} 0000 & 1100 & 1010 \\ 1111 & 0011 & 0101 \\ 0001 & 1101 & 1011 \\ 1110 & 0010 & 0100 \end{array} \right\}$$

Now that we know there is a natural representation of a cwatset as a union of cosets of its identity group, a natural question to ask is do all cwatsets with nontrivial identity groups have the same structure as the cwatsets we built in the preceding sections. Specifically, if C_{id} is nontrivial, can you always find a set of representatives of the cosets of C_{id} (or the cosets of some subgroup of C_{id}) that is itself a cwatset? The answer, interestingly enough, is no.

Example 6.4. Consider the set

$$C = \left\{ \begin{array}{cccc} 0000 & 1101 & 0111 & 0110 \\ 1111 & 0010 & 1000 & 1001 \end{array} \right\}$$

It can be shown that C is a cwatset with $C_{id} = \{0000, 1111\}$. However, there is no way to take a representative from each of the four cosets of C_{id} and form a cwatset. For example, suppose we choose the representatives $R = \{0000, 1101, 0111, 0110\}$. Then, for example, $R + 1101 = \{1101, 0000, 0100, 1011\}$ which has no element of weight 2. So $R + 1101$ can't be a permutation of R. A similar argument works for any set of four coset representatives.

We should note that the union of C_{id} and any one of its cosets form a group, and therefore a cwatset. So C contains a cwatset of size 4, it just does not contain one that consists of representatives of cosets of C_{id}.

With our counterexample in hand, we conclude with two questions.

Question 6.5. Given a cwatset nontrivial identity group, when can you find a system of representatives of cosets of the identity group that form a cwatset?

We note, for example, that the situation in Example 6.4 can't happen with an odd degree cwatset. It is easy to show that all the even weight words in any cwatset form a cwatset. So if the degree of C is odd and $C_{id} = \{\mathbf{0}, \mathbf{1}\}$, then the even weight words in C will form a system of representatives of the cosets of C_{id} that is also a cwatset.

But given the discussion following Example 6.4, a refinement of Question 6.5 might be:

Question 6.6. Given a cwatset C with nontrivial C_{id}, if there exists a system of coset representatives R such that the set of weights of the elements of C is preserved under addition by elements of C, is R a cwatset?

Acknowledgment. The authors thank Gary Sherman for suggesting the original problem.

References

[1] Daniel Biss. On the symmetry groups of hypergraphs of perfect cwatsets. *Ars. Comb.*, 56:271–288, 2000.
[2] Nancy Elizabeth Bush and Paul Isihara. The cwatset of a graph. *Math. Mag.*, 74(1):41–47, 2001.
[3] Cary Girod, Matthew Lepinski, Joseph Mileti, and Jennifer Paulhus. Cwatset isomorphism and its consequences, Rose-Hulman MS TR 00-01, 2000.
[4] J. A. Hartigan. Using subsample values as typical values. *J. Amer. Stat. Assoc.*, 64(328):1303–1317, 1969.
[5] Julie Kerr. Hypergraph representations and orders of cwatsets, Rose-Hulman MS TR 96-03, 1996.
[6] Jody Radowicz and Pamela Richardson. Properties of cwatsets of order at most ten, Rose-Hulman MS TR 00-07, 2000.
[7] Gary Sherman and Martin Wattenberg. Introducing...cwatsets! *Math. Mag.*, 67(2):109–117, 1994.