

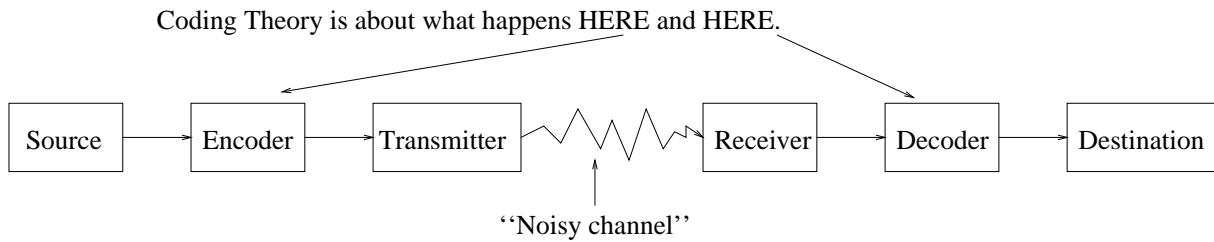
Error-correcting codes from permutation groups  
Ann Cook Prize entry

Robert Bailey

October 2004

# Introduction

First a word of warning: Coding Theory is NOT Cryptography. Cryptography is for *secrecy*, Coding Theory for *accuracy*.



- We want to transmit a message *accurately* along a “noisy channel”, where there may be interference.
- If the received message contains errors, we want to *decode* it to recover the original transmitted message.
- If the possible messages are sufficiently “different”, then this will be possible.
- What do we mean by “different”?  
If messages are strings of the same length, then the *Hamming distance* between two strings is the number of places in which they differ. For example,  $d_H(001, 010) = 2$ .
- Formally, a *code*,  $C$ , is a set of strings of symbols (“codewords”) chosen from some alphabet.
- The *minimum distance* of  $C$  is

$$\min_{\substack{v, w \in C \\ v \neq w}} \{d_H(v, w)\}$$

## Proposition:

If  $C$  has minimum distance  $d$ , then  $C$  can correct  $r = \left\lfloor \frac{d-1}{2} \right\rfloor$  errors.

- “Good” codes have:
  1. a reasonably large number of codewords;
  2. a reasonably large minimum distance;
  3. a usable decoding algorithm.

Note that properties 1 and 2 are mutually incompatible; for a fixed length and alphabet size, as the number of codewords increases they will become closer together, thus reducing the minimum distance. So a compromise will have to be found. Also, note that property 3 is independent of either of the first two!

# Our approach

- The usual approach to coding theory is to use *linear codes*, where the code is a  $k$ -dimensional subspace of an  $n$ -dimensional vector space over a finite field,  $\mathbb{F}_q$ .
- Our approach is to use permutation groups, in the manner described below.
- Let  $G$  be a permutation group acting on  $\Omega$ , where  $|\Omega| = n$ .
- We can write elements of  $G$  as ordered  $n$ -tuples of distinct symbols from  $\Omega$ ,

e.g.  $231794685 \in S_9$

- Can define Hamming distance as before.
- However,

$$\begin{aligned} d_H(g, h) &= \#x \text{ where } x^g \neq x^h \\ &= n - |\text{Fix}(gh^{-1})| \end{aligned}$$

- Thus the minimum distance is

$$\min_{\substack{g \in G \\ g \neq 1}} \{n - |\text{Fix}(g)|\},$$

the *minimum degree* of  $G$ .

- Also useful to us is the following notion:

## Definition:

A *base* for  $G$  is a sequence of points  $(x_1, \dots, x_b)$  from  $\Omega$  such that its pointwise stabiliser is the identity.

- A consequence of this is that the action of  $g \in G$  on a base *uniquely determines that element*. This will be of use in our decoding algorithm – more on which later.
- The minimum degree and base structure are not always easy to determine, but for the following families it is straightforward.

*Sharply  $k$ -transitive groups* (of degree  $n$ )  
 Minimum distance:  $n - k + 1$   
 (no two elements can agree on  $k$  or more points)  
 Base structure: any  $k$  points

$GL(n, q)$  acting on  $\mathbb{F}_q^n \setminus \{0\}$   
 Minimum distance:  $q^n - q^{n-1}$   
 (fixed points sets: vector subspaces of  $\mathbb{F}_q^n$ )  
 Base structure: a basis for  $\mathbb{F}_q^n$

$C_m \wr S_n$  acting on  $\{1, \dots, m\}^n$   
 Minimum distance:  $m$   
 (fixed points occur in multiples of  $m$ )  
 Base structure:  $n$  points, one from each copy of  $\{1, \dots, m\}$

$AGL(n, q)$  acting on  $\mathbb{F}_q^n$   
 Minimum distance:  $q^n - q^{n-1}$   
 (fixed points sets: affine subspaces of  $\mathbb{F}_q^n$ )  
 Base structure: an affine basis for  $\mathbb{F}_q^n$

# A decoding algorithm

- If a received word contains  $r$  errors, then clearly it must contain  $n - r$  correct symbols.
- So, if these correct symbols lie in positions labelled by a base, we can decode successfully.

**PROBLEM:** We can't necessarily tell in which positions the errors are.

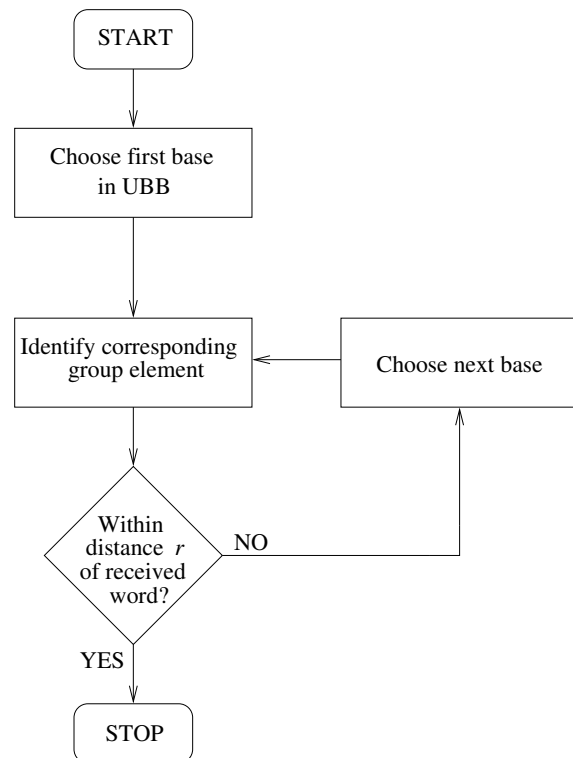
**SOLUTION:** We need a set of bases such that any combination of  $r$  error positions is disjoint from at least one base. We call this an *uncovering-by-bases* (UBB).

- Finding a UBB, in general, is not easy! At least, finding a “good” one (i.e. one that is relatively small) is not easy.
- For sharply  $k$ -transitive group (say of degree  $n$ ), any  $k$ -tuple of points forms a base. So what we require is a set of  $k$ -subsets of  $\{1, \dots, n\}$  such that any  $r$ -set of errors is disjoint from at least one  $k$ -set. In fact, what we have here is that the  $k$ -sets are the complements of the blocks of an  $(n, n - k, r)$ -covering design. (This is where the name “uncovering” comes from.) Clearly, the set of *all*  $k$ -subsets of  $\{1, \dots, n\}$  forms an uncovering, just not a very good one. However, finding an minimal one is more difficult and there is no general method.
- For other groups it is more complicated. To start with, you need to prove that a UBB actually exists, by showing that given an arbitrary set of  $r$  error positions, there is a base disjoint from it. This is non-trivial, but is straightforward. In the worst case, you would need a different base for each set of error positions. Actually *finding* a reasonably small UBB is much more difficult.

The decoding algorithm works as follows:

- Look in the positions of the received word,  $w$ , that are labelled by the first base in the UBB.
- If the symbols appearing are all distinct, identify the unique group element  $g$  (if it exists: existence is only guaranteed if the group is sharply  $k$ -transitive) that corresponds. (There are algorithms in computational group theory that do this.)
- If  $d_H(g, w) \leq r$ , then we have decoded. If not, move to the second base and repeat the process.

The diagram describes this procedure.



# A nice example

- The Mathieu group  $M_{12}$  is a sharply 5-transitive group of degree 12. (That is, it acts on 12 points, and given any two 5-tuples of distinct points, there is a unique group element mapping the first to the second.) It has minimum distance 8, so can correct  $\lfloor \frac{8-1}{2} \rfloor = 3$  errors.
- Since any 5-tuple of points forms a base, we need a set of 5-subsets of  $\{1, \dots, 12\}$  such that any 3-subset is disjoint from at least one 5-set. An example is shown on the right.

1	2	3	4	5
1	2	6	11	12
1	3	7	8	9
1	4	6	7	10
1	5	8	9	11
2	4	8	9	12
2	5	7	10	11
3	4	7	11	12
3	5	6	10	12
3	6	8	9	11
6	7	8	9	10

Suppose we transmit

$$g = 1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12$$

and receive

$$w = 6\ 2\ 1\ 4\ 6\ 6\ 7\ 8\ 9\ 10\ 11\ 12$$

$w$  has errors in positions 1, 3 and 5.

As the algorithm works through the uncovering, it outputs:

Error (repeated symbol);

Error (repeated symbol);

6 3 1 4 12 2 7 8 9 5 10 11, which is distance 6 from  $w$  and is rejected;

Error (repeated symbol);

Error (repeated symbol);

1 2 3 4 5 6 7 8 9 10 11 12, which is distance 3 from  $w$  and is accepted.

## References

- [1] I.F. Blake, Permutation codes for discrete channels, *IEEE Trans. Inform. Theory* **20** (1974), 138–140.

This was the first paper to suggest the use of sharply  $k$ -transitive permutation groups as error-correcting codes.

- [2] D.M. Gordon, *La Jolla Covering Repository*, <http://www.ccrwest.org/cover.html>.

This website contains a very useful database of covering designs, from which uncoverings can be constructed.

- [3] T. Maund, *Bases for Permutation Groups*, D.Phil. thesis, University of Oxford, 1988.

This thesis classifies those permutation groups which act transitively on their bases. This means that all bases will look “the same” in some sense, and so the structure is uncomplicated, making constructing a UBB easier.