NOTES

An Application of Set Theory to Coding Theory

NOGA ALON
Tel Aviv University
Ramat-Aviv, Tel Aviv, Israel

ANDY LIU
University of Alberta
Edmonton, T6G 2G1, Canada

"As high-speed electronic communication becomes commonplace, there is a tremendous need for better transmission schemes—ones that minimize the effect of inevitable transmission errors, ones that protect confidential or secret messages, ones that route messages most efficiently. Many of the best schemes are based on patterns or properties of classical algebraic and geometric objects, originally studied for their intrinsic interest. Mathematically, these are the subjects of information theory, coding and encryption" [**5**].

In this note, we will confine our attention to one aspect of information theory: error-correcting codes. We first state our assumptions about the setting in which these codes come into play. Messages are in the form of sequences of 0's and 1's. In transmission, the only errors that may occur are in the form of digit-reversals; that is, an error may change a 0 to a 1, or vice versa. The transmission device handles blocks of l digits at a time, and we know the maximum number of errors per transmission.

If the original message is broken up into blocks of length l, and transmitted as is, potential transmission errors will compromise the reliability of the received message. To trade off economy for accuracy, the original message is broken up into words of length $m < l$, and each word is augmented with $l - m$ digits in such a way that the correct word can be deciphered despite possible errors.

An error-correcting code may be defined as a pair of companion procedures. The first, that of determining how the $l - m$ additional digits are to be chosen, is called encoding. The second, that of recovering the correct word from the received block, is called decoding. The ratio m/l is a measurement of the efficiency of the code.

The subject of error-correcting codes is of immense scope and depth, as detailed in the monumental treatise by MacWilliams and Sloane [**6**]. An excellent exposition by Thompson [**10**] shows the interrelationship between codes and many other mathematical structures. Our primary purpose is to give another example along this line. We will show how a recent set-theoretic result of Frankl and Pach [**2**] provides an alternative justification for a family of known codes.

Error-correcting codes may be based on very simple ideas (see for example [**1**] and [**8**]), but these tend to suffer in efficiency. The extended Hamming codes (see [**3**] and [**4**]), discovered early in the history of information theory, enjoy the best of both worlds. We will begin by describing such a code in set-theoretic language.

Suppose we have a transmission device which handles 15 digits at a time, with at most 1 error per transmission. We break up the message into words of length 11.

TABLE I

a	a	a	a		a	a	a				a			
b	b	b		b	b			b	b			b		
c	c		c	c		c		c		c			c	
d		d	d	d			d		d	d				d
1	0	1	1	0	0	0	1	1	0	0	0	1	1	0

TABLE I shows how the word 10110001100 is to be encoded. All 15 nonempty subsets of $\{a, b, c, d\}$ are listed, each occupying a column over the horizontal line. A vertical line separates the 4 one-element subsets from the other 11. Beneath the horizontal line under these 11 subsets, the word is copied with one digit in each column.

The extra digit to be placed under $\{a\}$ is obtained as follows. Consider all other subsets which *include* $\{a\}$, namely, $\{a, b, c, d\}$, $\{a, b, c\}$, $\{a, b, d\}$, $\{a, c, d\}$, $\{a, b\}$, $\{a, c\}$ and $\{a, d\}$. (We could have referred to these as subsets which contain the element "a." However, anticipating our main result, we prefer to describe them using the relation of set inclusion.) Under the corresponding columns, there are four 1's. We append a 0 under $\{a\}$ so that the total number of 1's under all subsets which include $\{a\}$ is even. A similar process is used to determine the remaining three appended digits.

TABLE II

a	a	a	a		a	a	a				a			
b	b	b		b	b			b	b			b		
c	c		c	c		c		c		c			c	
d		d	d	d			d		d	d				d
1	0	1	1	0	0	1	1	1	0	0	0	1	1	0

Suppose the received block is 101100111000110 as shown on the bottom line of TABLE II. Because the total number of 1's under the columns corresponding to subsets which include $\{a\}$, $\{b\}$, $\{c\}$ and $\{d\}$ are 5, 4, 5 and 4 respectively, it is clear that an error has occurred. The parities for $\{a\}$ and $\{c\}$ are disturbed. Since there is at most one digit-reversal, it must have occurred under $\{a, c\}$. Thus the message can be decoded correctly.

More generally, suppose the blocks are of length $l \geqslant 3$. Let n be the greatest integer such that $2^n - 1 \leqslant l$. The word length is chosen to be the greatest integer m such that $m \leqslant 2^n - \binom{n}{0} - \binom{n}{1}$. We use an n-element set instead of $\{a, b, c, d\}$. The word is copied under the columns corresponding to the subsets of size at least two. If $m < 2^n - \binom{n}{0} - \binom{n}{1}$, the appropriate number of these columns are omitted. The digits under the columns corresponding to the subsets of size one are determined by the parity condition as in TABLE I.

The extended Hamming codes cannot correct multiple transmission errors. For instance, a double-error on the digits under $\{a, b\}$ and $\{a\}$ has the same net effect as a single error on the digit under $\{b\}$. However, our set-theoretic approach suggests a generalization immediately. Again, we use an example.

Suppose we have a transmission device which handles 15 digits at a time, with up to 3 errors per transmission. We break up the messages into words of length 5.

TABLE III is set up just like TABLE I, but with an additional vertical line separating the two-element subsets from the rest. The word 10110 is copied beneath the

TABLE III

a	a	a	a		a	a	a				a				
b	b	b		b	b			b	b			b			
c	c		c	c		c		c		c			c		
d		d	d	d			d		d	d					d
1	0	1	1	0	0	0	1	1	0	0	0	1	1	0	

horizontal line under the columns corresponding to subsets of size at least three.

The digit under $\{a, b\}$ is obtained by ensuring that the total number of 1's under the subsets which include $\{a, b\}$ is even. It will be 0 since the number of 1's under the relevant subsets other than $\{a, b\}$ itself is even (2, under $\{a, b, c, d\}$ and $\{a, b, d\}$). The other five digits between the vertical lines are similarly determined. The table is then completed as in the extended Hamming code.

TABLE IV

a	a	a	a		a	a	a				a				
b	b	b		b	b			b	b			b			
c	c		c	c		c		c		c			c		
d		d	d	d			d		d	d					d
1	0	1	0	0	1	0	1	1	0	0	0	1	1	1	

Suppose the block 101001011000111 is received as shown on the bottom line of TABLE IV. Parity checks reveal that violations occur with respect to $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{c, d\}$, $\{b\}$ and $\{c\}$. We use P, Q and R to denote the (at most) three subsets of $\{a, b, c, d\}$ such that the digits beneath them are reversed.

Note that each of $\{b\}$, $\{c\}$ and $\{a, d\}$ is a subset of an odd number (1 or 3) or P, Q and R, while each of $\{a\}$, $\{d\}$ and $\{b, c\}$ is included in an even number (0 or 2) of them. It follows that $\{a, d\}$ is a subset of exactly 1 of P, Q and R, say P. Moreover, one of Q and R, say Q, includes $\{a\}$ while the other, R, includes $\{d\}$.

Similarly, $\{b, c\}$ is not a subset of any of P, Q and R, and that each of $\{b\}$ and $\{c\}$ is a subset of exactly 1 of them. Violations with respect to $\{a, c\}$ and $\{c, d\}$ imply that the element "c" is in P, while the absence of a violation with respect to $\{b, d\}$ means that "b" cannot be in R. Hence "b" is in Q, which is consistent with the violation with respect to $\{a, b\}$.

Thus the errors occur under $P = \{a, c, d\}$, $Q = \{a, b\}$ and $R = \{d\}$, and decoding is completed. This type of code is known as the punctured Reed-Muller codes (see [7] and [9]).

More generally, suppose the blocks are of length $l \geqslant 3$. Let n be the greatest integer such that $2^n - 1 \leqslant l$. The maximum number of errors per transmission that can be accommodated is $2^{n-1} - 1$. Let t be the least integer such that $2^t - 1$ is greater than or equal to the maximum number of transmission errors. The word length is chosen to be the greatest integer such that $m \leqslant 2^n - \binom{n}{0} - \binom{n}{1} - \cdots - \binom{n}{t}$. We use an n-element set instead of $\{a, b, c, d\}$. Additional vertical lines separate subsets of size up to t. The word is copied under the columns corresponding to the subsets of sizes at least $t + 1$. The digits under the columns corresponding to the subsets of sizes $t, t - 1, \ldots, 1$ are determined recursively by the parity condition.

If $m < 2^n - \binom{n}{0} - \binom{n}{1} - \cdots - \binom{n}{t}$, the appropriate number of subsets of sizes at least $t + 1$ are omitted. It should be pointed out that in such cases, the codes can no longer be called the punctured Reed-Muller codes. It seems natural to refer to them as the extended Reed-Muller codes. In particular, the extended Hamming codes are special cases of the extended Reed-Muller codes.

Although the decoding schemes for both are clearly related, that for the extended

Reed-Muller codes, as illustrated by decoding TABLE IV, does seem rather ad hoc. We now come to the set-theoretic result which guarantees that, in fact, the errors can always be uniquely identified. For practical applications, decoding tables can be constructed.

To facilitate the description of parity-disturbances, we introduce a notation. For any set S and any positive integer t, let S^t denote the collection of nonempty subsets of S of sizes at most t. If $S = \{a, b, c, d\}$, then $S^2 = \{\{a\},\{b\},\{c\},\{d\},\{a,b\},\{a,c\},\{a,d\},\{b,c\},\{b,d\},\{c,d\}\}$.

We also remind the reader of the associative operation of symmetric difference between two sets. $A \triangle B$ yields the set of elements belonging to exactly one of A and B. In general, $A_1 \triangle A_2 \triangle \cdots \triangle A_k$ yields the set of elements which belong to an odd number of the A_i.

To see how this notation helps describe parity-disturbances, we return to the example in TABLE IV. We have

$$\begin{aligned}
&\{a,c,d\}^2 \triangle \{a,b\}^2 \triangle \{d\}^2 \\
&\quad = \{\{a\},\{c\},\{d\},\{a,c\},\{a,d\},\{c,d\}\} \\
&\qquad \triangle \{\{a\},\{b\},\{a,b\}\} \triangle \{\{d\}\} \\
&\quad = \{\{b\},\{c\},\{a,b\},\{a,c\},\{a,d\},\{c,d\}\}.
\end{aligned}$$

In the general case of the extended Reed-Muller codes, if the transmission errors occur under the subsets $A_1, A_2, \ldots, A_r$, this will result in some parity-disturbances. Specifically, if D is a subset of the n-element set such that the size of D is at most t, then the parity with respect to D will be disturbed if and only if D is a subset of an odd number of the A_i, that is, if and only if D belongs to $A_1^t \triangle A_2^t \triangle \cdots \triangle A_r^t$.

The decoding problem is to determine the A_i, given the subsets of the same type as D. Thus it is crucial to know that the subsets A_i may be determined uniquely.

Suppose to the contrary that there are two distinct collections of subsets, $\{A_1, A_2, \ldots, A_r\}$ and $\{B_1, B_2, \ldots, B_s\}$, which give rise to the same pattern of parity-disturbances. Then we have $A_1^t \triangle A_2^t \triangle \cdots \triangle A_r^t = B_1^t \triangle B_2^t \triangle \cdots \triangle B_s^t$ or equivalently $A_1^t \triangle \cdots \triangle A_r^t \triangle B_1^t \triangle \cdots \triangle B_s^t = \varnothing$. If $A_i = B_j$ for some i and j, we may remove them. Since each of r and s is at most $2^t - 1$ (the maximum number of errors per block), $r + s \leqslant 2^{t+1} - 2$. This contradicts the following result.

THEOREM. *Let $S_1, S_2, \ldots, S_r$ be distinct nonempty finite sets. If $S_1^t \triangle S_2^t \triangle \cdots \triangle S_r^t = \varnothing$, then $r \geqslant 2^{t+1} - 1$ and this is best possible.*

Proof. If $S_1, S_2, \ldots, S_r$ are the $2^{t+1} - 1$ nonempty subsets of a $(t+1)$-element set, then we have $S_1^t \triangle S_2^t \triangle \cdots \triangle S_r^t = \varnothing$. This shows that $r \geqslant 2^{t+1} - 1$ is best possible. We must now show that this inequality always holds.

We begin by noting that because the S_i are distinct, the largest one among them is a subset of itself but of no other S_i. That is, there is certainly *some* set which is a subset of an odd number of the S_i. Thus it will be possible to choose a nonempty set B of minimal cardinality k which is a subset of an odd number of the S_i. If $k \leqslant t$, then B belongs to $S_1^t \triangle S_2^t \triangle \cdots \triangle S_r^t$. Since this collection is empty, we have $k \geqslant t + 1$.

Next we claim that if C is any one of the $2^k - 1$ nonempty subsets of B, then there will be an odd number (and hence at least one) of the S_i having the property that $S_i \cap B = C$. Of course, if $C_1 \neq C_2$, then the sets S_i with $S_i \cap B = C_1$ must be distinct from the sets S_j with $S_j \cap B = C_2$. That is, if this claim is true, there must be at least $2^k - 1$ different sets S_i. In other words, $r \geqslant 2^k - 1 \geqslant 2^{t+1} - 1$, and the theorem will be proved.

We shall prove this claim by a "descending" inductive argument on $|C|$. When $|C| = k$, we have $C = B$, and $S_i \cap B = B$ for an odd number of S_i by the definition of B. Suppose then that the claim is true for all subsets of B having cardinality *greater* than j, $1 \leqslant j < k$. We will show that this implies that it must also be true for any subset of B of cardinality j.

We consider such a subset C, restricting our attention to those S_i which include C. We expect that some of these S_i will contain elements of $B - C$. Indeed, if D is any one of the 2^{k-j} subsets of $B - C$, we can form the class of S_i such that $S_i \cap B = C \cup D$.

It is our goal to show that the class corresponding to $D = \varnothing$ has an odd number of members. We establish this via three observations.

(1) *The total number of classes is even.*

More explicitly, there are exactly 2^{k-j} classes.

(2) *The total number of S_i under consideration is even.*

This follows from the definition of B (in particular, the minimality condition) and the fact that $|C| < |B|$.

(3) *All but perhaps one class contain an odd number of the S_i.*

This is the induction hypothesis, which applies to all classes except that corresponding to $D = \varnothing$.

It follows from these observations that the final class, corresponding to $D = \varnothing$, must necessarily contain an odd number of the S_i. The claim is thus proved, and the theorem follows.

Our theorem is essentially a paraphrase, in set-theoretic language, of a result in [2]. Conversely, this result could have been derived from the knowledge that the extended Reed-Muller codes do work.

The authors thank the referees for helpful suggestions which improved the presentation of this paper.

REFERENCES

1. F. L. Alt, A Bell Telephone Laboratories' computing machine (I), *Math. Tables and Aids to Comput. (Math. Comput.)* 3 (1948/49), 1–13.
2. P. Frankl and J. Pach, On the number of sets in a null t-design, *Europ. J. Comb.* 4 (1983), 21–23.
3. M. J. E. Golay, Notes on digital computing, *Proc. IRE (IEEE)* 37 (1949), 657.
4. R. W. Hamming, Error-detecting and error-correcting codes, *Bell System. Tech. J.* 29 (1950), 147–160.
5. A. Jaffe, Ordering the universe, *Renewing U.S. Mathematics*, National Academy Press, Washington, DC, 1984, p. 148.
6. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977, pp. 372–377.
7. D. E. Muller, Application of Boolean algebra to switching circuit design and to error detection, *IRE Trans.* EC-3 (1954), 6–12.
8. M. Rabenstein, An example of an error-correcting code, this MAGAZINE 58 (1985), 225–226.
9. I. S. Reed, A class of multiple-error-correcting codes and the decoding scheme, *IRE Trans.* IT-4 (1954), 38–49.
10. T. M. Thompson, *From Error-Correcting Codes through Sphere Packings to Simple Groups*, MAA, Washington, DC, 1983.