J. E. Griffith

COMPANY CONFIDENTIAL

PROJECT STRETCH

STRETCH MEMO NO. 51

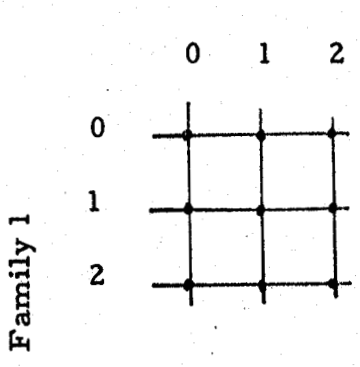
SUBJECT: Error-Correcting Codes Using Compatible Families
BY: F. P. Brooks
DATE: December 3, 1956 (release date)

ERROR-CORRECTING CODES USING COMPATIBLE FAMILIES

Introduction

Erroneous elements of a binary set can be corrected by inversion when they are identified. A large class of error-correcting codes are based upon this principle and are designed to identify erroneous code elements. A second class of codes are those, such as the Reed code, in which the erroneous elements are not specifically identified, but which yield the proper information from an error-disturbed configuration upon application of the decoding algorithm. Codes of the first class only will be treated here.

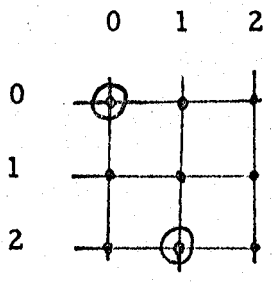
Error-identifying codes in general can be considered as dividing the space of code points into families of mutually exclusive and exhaustive subsets.



Any bit will therefore lie in one, and only one, subset of each family. If an indication of the subsets in which the error lies can be given, and if families can be so constructed that each bit is the sole intersection of the subsets in which it lies, single error identification and thus correction is possible.

In general, however, if each bit is identified by s such families, the code points can be considered as members of an s-dimensional finite space. If each subset contains the same number of points m and this is equal to the number of subsets in each family, the space will be called complete and the families proper.

While f families of m subsets are sufficient to identify m^f points, multiple errors in general will give ambiguous identifications for the erroneous bits. Consider a case for f = 2, m = 3:



If the erroneous subsets are identified as 0, 1 of the vertical family and 0, 2 of the horizontal family, this indicates that 0, 0 and 1, 2 or 0, 1 and 2, 0 or any three or all four of bits 0, 0; 0, 1; 2, 0; 2, 1 are in error. In short, the two errors belong to two subsets of each family, and these subsets have four intersections.

Ambiguities of this sort can be resolved by reducing the number of points by t^w so that m families are used to identify the t^{m-w} points. Let m-w = s. Then s families are sufficient to identify each point, and w redundant families lie in the s-dimensional space for the resolution of ambiguities.

The problem of choosing such families in a satisfactory manner is one of a geometry of coincidence, for the goal is to establish families so that all erroneous bits are the coincidences of the erroneous subsets, and all such coincidences are erroneous bits. This permits one-step (non-sequential) error correction with circuitry that operates on each bit independently.

Ashenhurst* has developed a suitable coincidence geometry in connection with the related problem of improving switching and selection ratios of coincident-current core memories. His results will be briefly reviewed.

The Commutative Ring with Unity.

Let a set of elements be given and let operations of addition and multiplication be defined that every pair of elements a, b has a unique sum $a + b$ and the product ab in the set. The resulting system is called a commutative ring if the ordinary commutative: $\left\{\begin{array}{l}(a + b = b + a)\\ ab = ba\end{array}\right\}$, associative: $\left\{\begin{array}{l}a + (b + c) = (a + b) + c,\\ a(bc) = (ab)\,c\end{array}\right\}$ and distributive: $a(b + c) = ab + ac$, laws hold, and if the equation $a + x = b$ has solutions for all a and b. Consider a ring R with a finite number of elements m.

There is then a unique element 0 which satisfies $a + x = a$ for each element in R. Further, there is a unique element $x = -a$ such that

* Ashenhurst, R. L., the Structure of Multiple Coincidence Selection Systems, Ph. D. Thesis Harvard University, Cambridge, Massachusetts, 1956.

$a \not x = 0$ for any a in R.

$$(a)(b) = -ab, \quad (-a)(-b) = ab$$

$$(a)(0) = 0$$

If $ab = 0$ for some $b \neq 0$, then a is called a divisor of zero. Zero is trivally a divisor of zero, and there may be others.

A commutative ring need not have an element b that satisfies $ab = a$. If it does, however, it is known as a commutative ring with unity. Only such rings will interest us here. There can be only one such unity element, and if a is not a divisor of zero, $ab = 1$ has a unique solution a^{-1}. Further, $ax = b$ has a unique solution $x = ba^{-1}$ if a is not a divisor of zero. $ax = b$ has no solution if a is a divisor of zero and b is not. If both are divisors of zero, there is either no solution or at least two.

An element a for which $ax = 1$ has solutions is called a divisor of unity. Thus, every element in R is either a divisor of zero or a divisor of unity, but not both.

The Algebra of Integers Modulo m (J_m)

The algebra of integers modulo m is a good example of a commutative ring with unity. Let $(P)_m$ = P mod m stand for the remainder $0 \leq j < m$ of the division P/km where k is an integer sufficiently large to satisfy the condition on j. The modulo m algebra J_m has the following rules for addition and multiplication:

$$a_m \not b_m = (a \not b)_m$$

$$a_m b_m = (ab)_m$$

For example, for m = 3 the addition table is

	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

and the multiplication table is

	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Such an algebra may have nontrivial divisors of zero. Consider the tables for m = 4.

ADDITION

	0	1	2	3	
0	0	1	2	3	0
1	1	2	3	0	1
2	2	3	0	1	2
3	3	0	1	2	3

MULTIPLICATION

	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

The Galois Field (G_m)

A Galois field is a commutative ring with unity that contains no non-trivial divisor of zero. Such a field contains $m = p^q$ elements, where p is prime and $q > 0$. There is only one such field for each p and q. For $q = 1$, the field is equivalent to the algebra of integers modulo p. An example is G_4, the field of order 2^2

ADDITION

	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

MULTIPLICATION

	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	2	1	2

There are still other examples of commutative rings with unity, but these will be of primary interest.

The Coincidence Geometry of Two Dimensions, $C_2(R)$, Based on a Commutative Ring with Unity

Let R contain m elements. A point of $C_2(R)$ is defined by an ordered pair of elements (x, y) its coordinates. The space of $C_2(R)$ is the set of m^2 points derived by letting x and y assume all positive values.

A subset consisting of the points for which one coordinate has a fixed value is a coordinate subspace or coordinate line. Each such contains m points. Any pair of coordinate subspaces $S(x, y_o)$ and $S(x_o, y)$ intersect at one point (x_o, y_o).

A line in $C_2(R)$ is a set of points satisfying the equation

$$ax + by = c$$

where a and b are its coefficients. A line is said to be proper if it contains exactly m points. The coordinate lines are proper. A line is oblique if it is proper, and if it has exactly one point in common with each horizontal and each vertical coordinate line.

A line is oblique if and only if its coefficients a and b are both divisors of unity. An oblique line can thus be expressed by $x + \beta y = \gamma$ where $\beta = b/a$ and $\gamma = c/a$. β is called the orientation and describes the family to which the line belongs. γ is the x - intercept of the line.

There is one, and only one, oblique line of a given orientation that contains a given point. Therefore, a family of oblique lines divides the space into mutually exclusive and exhaustive subsets just as the coordinate lines do.

Since β may not be a divisor of zero, there may be μ families of μ lines each, where μ is the number of divisors of unity in R. At any point, there are exactly μ oblique lines through the point.

For any R, $c \leq u$ families of lines may be selected so that no two lines through a point have any other point in common. Such a set of families is called compatible. A set of oblique lines is compatible with respect to a given point if, and only if the parallel set is compatible with respect to the origin. This simplifies examinations for compatibility. A set of $f > 1$ oblique lines (families) through a point with orientations $\beta_1, \beta_2 \ldots\ldots\ldots \beta_f$ is compatible if, and only if, $\beta_i - \beta_j$ is a divisor of unity for all i and j. Over a modulo m algebra, the largest compatible set of oblique lines (families) which may be found contains m' - 1 families, where m' is the least prime factor of m. Clearly the lines passing through (0, 0) and (1, 1); (1, 2);..... (1, m' - 1) are compatible, since any prime divisor of zero must be a prime factor of m. The line through (0, 0) and (1, m') is not, and those through (1, m' + 1), (1, m' + 2) etc. have β differeing from the first set by m'. Including the two co-ordinate families, m' + 1 compatible families can be found, of which m' - 1 are oblique (redundant) and may be used for identifying higher order errors.

These concepts can be extended, and Ashenhurst derives a similar geometry C_s for s-dimensional spaces, where the coordinate subspaces are hyperplanes, and oblique hyperplanes are added to provide redundancy in selection. The conditions on compatibility are somewhat more complicated, and considerably more work must be done before the direct application of the s-dimensional coincidence geometry to error correction can be made.

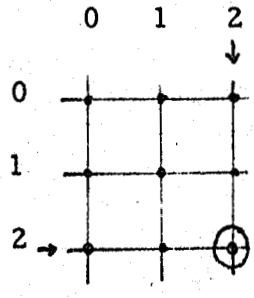
Investigations into Applications of the Coincidence Geometry

The application of the two dimensional geometries C_2 (R) to error correcting codes has been investigated in some detail.

It appears that k / s - 1 compatible families (or k - 1 redundant families) are necessary and sufficient for the bitwise correction of the k^{th} order errors in an s-dimensional space, although this has not yet been proved. Several examples may help make this palnsible:

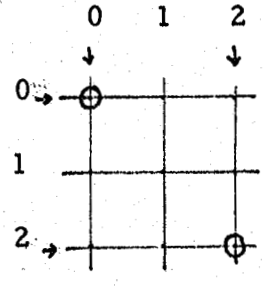
s = 2

k = 1

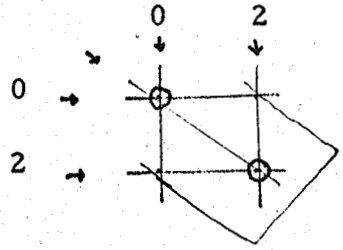


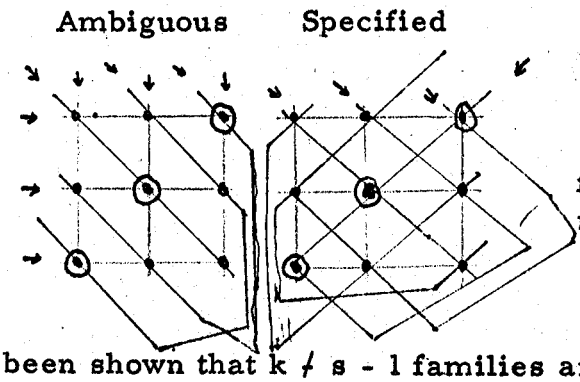
f = 2 families are necessary and sufficient

K = 2



f = 3 can be considered as f=2 and f = 3 families are necessary and sufficient

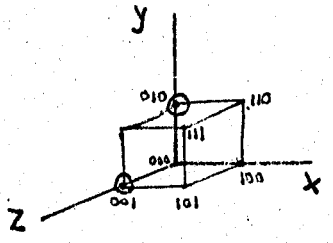




f = 4 families are necessary and sufficient

It has been shown that $k \neq s - 1$ families are necessary and sufficient for $s = 2$ $k = 1, 2, 3, 4, 5$.

For $s = 3$, $k = 2$, consider



$$\begin{bmatrix} x & y & z \\ (0 & 0 & 0), (0 & 0 & 1) \\ (0 & 1 & 0), (0 & 1 & 1) \end{bmatrix} \begin{bmatrix} (1 & 0 & 0), (1 & 0 & 1) \\ (1 & 1 & 0), (1 & 1 & 1) \end{bmatrix}$$

then planes $z = 0$, $x = 0, 1$, $y = 0, 1$ give error indications and a family of oblique planes is necessary. It can also be shown to be sufficient. Thus, $f = 4 = k \neq s - 1$ families are needed.

For $s = 4$, $k = 2$ the same holds, as can be seen by conceiving of the previous example as constituting one of the two 2 x 2 x 2 cubes in a 2 x 2 x 2 x 2 hypercube (tesseract).

A second unproved but intuitively obvious theorem is that any error of order lower than k can be bitwise corrected in a system that contains enough compatible families to provide k^{th} order correction. However, an error of order $r < k$ cannot always be corrected in a k x k space that will not contain $r \neq s - 1$ compatible families even if $r \neq s - 1$ compatible families could be placed in an r x r space. For example, for $s = 2$, $k = 4$, only 3 compatible

families can be drawn. If a fourth in $C_2(J_4)$ the other diagonal, is drawn it will not be compatible, and the four families will not give unambiguous correction of all triple errors, even though in a 3 x 3 space four compatible families could be drawn and triple errors corrected.

A consideration of some two dimensional systems will show the implications of Ashenhurst's requirements for a line (family) to be oblique and for a set of such to be compatible.

Consider a 3 x 3 system with families according to $C_2(J_3)$.

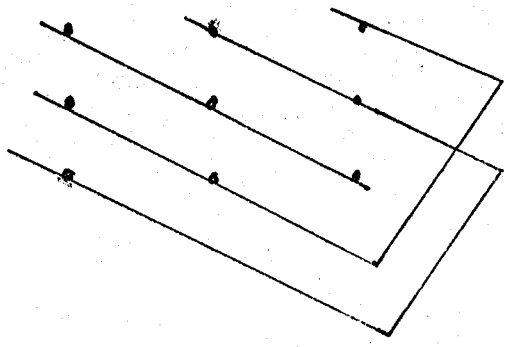
00 00	01 11	02 22
10 21	11 02	12 10
20 12	21 20	22 01

Each element can be identified by two coordinates, and two other coordinates can be attached indicating the subset of the redundant families to which the element belongs. The redundant families have the equations

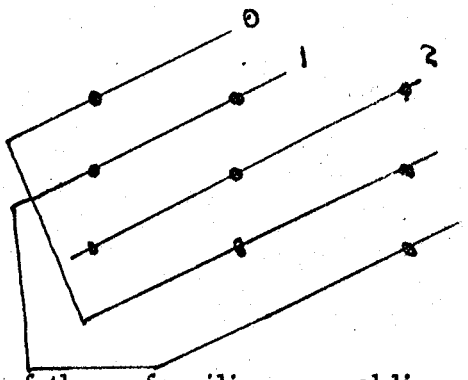
$x - y = \gamma$ $\beta = -1 = 2$ the principal diagonal and the lines parallel to it

and $x + y = \gamma$ $\beta = 1$ The other diagonal and the lines parallel to it.

$x \neq 2y = \gamma$ $\gamma = 0$ $\gamma = 1$ $\gamma = 2$

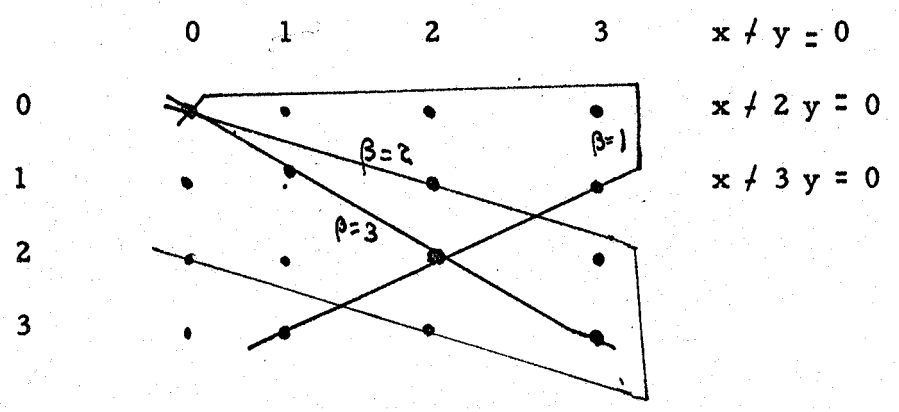


$x \neq y =$ $\gamma = 0$ $\gamma = 1$ $\gamma = 2$



Both of these families are oblique and they are compatible with each other. By the definition of obliqueness, the whole set of redundant and coordinate lines will always be compatible if the redundant lines are compatible.

In a 4 x 4 system of two dimensions, the potential oblique lines through the origin ($\gamma = 0$, $\beta = 1, 2, 3$) can be used to represent the families created by C_2 (J_4).

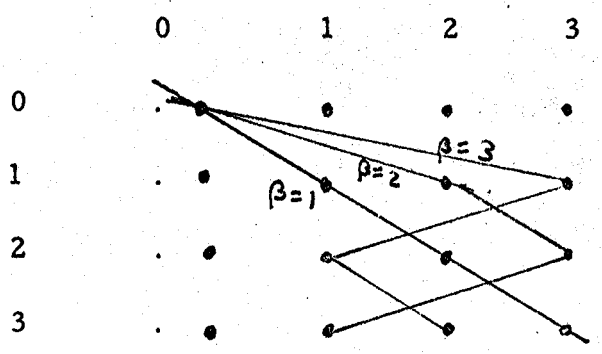


It will be seen that the potential oblique lines $\beta = 1, 2, \ldots\ldots$ m - 1 pass through 0, 0 and each of the points (1, y)other than (1,0) . The line for $\beta = 2$ does not satisfy the definition for obliquity, since it includes the point 2, 0 that lies on the coordinate line through the origin. This is in accordance with the theorem, since 2 is a divisor of zero in J_4.

The lines other than $\beta = 2$ are oblique, and they constitute the families of the two diagonals. They do not satisfy the definition for compatibility, however, since $\beta = 1$ and $\beta = 3$ cross at the origin and again at 2, 2. This agrees with the theorem for $\beta_i - \beta_j = 1 - 3 = -2 = 2$, a divisor of zero in J_4.

If one proceeds to use the three redundant families for error correction without regard to obliquity and compatibility, it will be found that all error bits are inverted, and for some errors, some correct bits are made wrong by the correcting process. It is not known whether the number made wrong is always smaller than the number corrected. If it is, correction could be done in steps.

This is not a profitable investigation to continue, however, since compatible families would yield one-step bitwise correction, and for any $m = p^q$ where p is prime and $q > 0$, a geometry on the Galois field can be constructed. Because the Galois field contains no nontrivial divisors of zero, all of the oblique families (m - 1) will be compatible, so m-order errors can be corrected in a $s = 2$ space. For $m = 4 = 2^2$, the lines through the origin are



They are indeed compatible with themselves and the coordinate lines, and do provide for one-step bitwise correcting of all quadruple errors.

This whole investigation has assumed that the families are proper: each contains m subsets of exactly m points. Further investigation may properly examine families for which this is not the case.

A more basic set of assumptions is that in a k^{th} order correction system all the subsets containing $1 \leq r \leq k$ errors are always indicated as erroneous, and that this indication itself is never in error. For $k > 1$, simple one bit parity checks fail to suffice for indicating erroneous subsets, since they do not satisfy either assumption, much less both.

The Problem of Embedding Error Indications Within Checked Arrays

If the error indications are transmitted over the same noisy channel as the information bits, the second assumption is unrealistic for practical use. To overcome this difficulty, the error indicator bits must be embedded in the space which is checked. We shall next consider this problem.

A series of sums of any s-dimensional array of numbers along the coordinate lines will yield the same final result regardless of the order of the summations. That is, $\sum_{i=1}^{I} \cdots \sum_{r=1}^{R} \left(\sum_{j=1}^{J} n_{ij\cdots r} \right) = \sum_{j=1}^{J} \cdots \sum_{r=1}^{R} \left(\sum_{i=1}^{I} n_{ij\cdots r} \right)$

for any number of indices. This property guarantees that a parity check can be imbedded in an array, for $\left(\sum n_{ij} \right) \bmod 2 = \sum \left[(n_{ij}) \bmod 2 \right]$

By the same token, s-dimensional arrays may be consistently Hamming checked along all coordinate dimensions, since such a Hamming check is a set of multiple summations mod 2 over various subsets of the space. Now can the error indicating bits for redundant families be consistently embedded in an array which is to be checked? This is not in general possible. Consider a 2 x 2 information array, parity checked in both coordinate dimensions so that it is a 3 x 3 array. The erroneous subset will only be positively identified for a single error by a simple parity check, so only single error correction is possible. Only two families are needed, and the coordinate families suffice. However, a consistently parity checked diagonal family cannot be constructed.

Positive identification of an erroneous subset that may contain two errors can be done with no code more cheaply than the Hamming code used for double error detection. Thus, 4 x 4 information bits must be embedded in a 7 x 7 array to identify double errors in a coordinate line. However, none of the six possible compatible oblique families in the 7 x 7 array permit consistent Hamming checking, so a third family cannot be used for double error correction.

It is worth noting that even had such a consistent family been found, the scheme would have guaranteed correction of only double errors, while the same number of bits used according to Elias' scheme would have guaranteed triple error correction.

It is possible that additional consistent families can be constructed by using more unconstrained bits for error indication along the coordinate lines. Thus, in the 7 x 7 array, the number of information bits was reduced to 3 x 3, leaving 7 bits unconstrained for the satisfaction of Hamming checks on a third problem. This has not been pursued to the end, but no general way to construct such a third family is apparent.

Conclusions

The compatible family approach does not appear apt to prove profitable for direct application where the error-indication bits are directly embedded in the checked array and where proper families are used.

Further investigation may prove the desirability of using weaker families, and while direct application of the theory in the obvious manner does not appear profitable, the notion of compatible families may prove a useful concept for the general theory of error correcting and detecting codes.

F. P. Brooks
September 13, 1956

FPB:bl