Build a Sporadic Group in Your Basement

Paul E. Becker, Martin Derka, Sheridan Houghten, and Jennifer Ulrich

Abstract. All simple finite groups are classified as members of specific families. With one exception, these families are infinite collections of groups sharing similar structures. The exceptional family of sporadic groups contains exactly twenty-six members. The five Mathieu groups are the most accessible of these sporadic cases. In this article, we explore connections between Mathieu groups and error-correcting communication codes. These connections permit simple, visual representations of the three largest Mathieu groups: M_{24}, M_{23}, and M_{22}. Along the way, we provide a brief, non-technical introduction to the field of coding theory.

1. INTRODUCTION In 1873, Emile Mathieu published a description of a 5-transitive group of permutations on 24 symbols. A simple group with 244,823,040 elements, it is now generally known as M_{24} – the largest of the five Mathieu groups.

In 1951, Ralph Stanton described Mathieu's groups as follows:

"... with but five exceptions, all known simple groups fall into infinite families; these five unusual simple groups were discovered by Mathieu and, after occasioning some discussion, were relegated to the position, which they still hold, of freakish groups without known relatives."[**23**, p. 164]

The timing of Stanton's comment was most unfortunate. In the new field of binary communication theory, the "freakish" group M_{24} was about to become a star. Claude Shannon's famous paper, "A Mathematical Theory of Communications," had explored the theoretical limits of communication in real-world systems just three years earlier. Marcel Golay responded with a uniquely useful communication scheme, in which 2^{12} symbols were represented by easily distinguished binary vectors of length 23. A parity check digit was soon added, and the extended binary Golay code became the most important example in the theory of error-correcting codes. The automorphism group of this structure is M_{24} [**14**, p. 251].

We may update Stanton's statement in light of the complete classification of finite groups. Simple groups are groups whose only normal subgroups are trivial. All finite groups may be constructed with simple groups as their component parts. The classification theorem states that almost all finite simple groups may be sorted into four general types. These types are the cyclic groups of prime-power order, the alternating groups, the projective unimodular groups, and the groups of Lie type. There are exactly twenty-six groups which defy this classification, and are collectively called the sporadic groups [**21**, p. 269]. The Mathieu groups are the most accessible of these sporadic cases.

Stanton's interest in Mathieu's group was based on early work in classifying simple groups. He cited a text from the year 1901:

"Dickson has shown that there are infinitely many group orders g with the property that there exists two simple groups of order g, the lowest such value of g is 20,160."[**23**, p. 164]

Stanton then proved that M_{24} is unique, if not "freakish."

Lemma 1. The only simple group of order 244, 823, 040 is the Mathieu group M_{24}.

Mathieu's original construction of M_{24} utilized three apparently arbitrary permutations a, b, c [**9**, p. 209] where:

$$a = (1, 2, 3, \ldots, 23)$$
$$b = (3, 17, 10, 7, 9)(5, 4, 13, 14, 19)(11, 12, 23, 8, 18)(21, 16, 15, 20, 22)$$
$$c = (1, 24)(2, 23)(3, 12)(4, 16)(5, 18)(6, 10)(7, 20)\ldots$$
$$\ldots(8, 14)(9, 21)(11, 17)(13, 22)(19, 15).$$

Searching for alternative constructions of the Mathieu groups has become something of a cottage industry. R. T. Curtis, who presented M_{24} as a group of actions on the faces of an icosatetrahedron, commented:

"Mathieu himself constructed the groups by 'gluing together' copies of linear fractional groups in a very clever but hardly 'natural' manner." [**5**, p. 423]

This article offers a construction for three of the Mathieu groups which is "natural" in a practical sense. We tie Golay's extremely practical communication code to Mathieu's intellectual oddities. Along the way, we provide the following: some background in coding theory; two simple constructions of Golay's error-correcting code; a correlation of those constructions which produces the sporadic group M_{24}; and a clear visual description of M_{24}. Finally, we build two more of Mathieu's groups as stabilizer subgroups within M_{24}. We hope to show that the Mathieu groups have a fascinating structure which goes far beyond their original definitions.

2. ERROR CORRECTING CODES Shannon's 1948 paper [**22**] introduced a mathematical model for the transmission of digital information. His basic model incorporated a finite set of possible signals, a sender, a receiver, and a communication channel connecting them. He assumed the channel's properties are known, including the types (and probabilities) of errors which the channel introduces. We will be concerned with symmetric binary channels, where the primary issue is random substitution errors. In such channels, any digit 1 may be replaced by a 0 with a fixed (small) probability p. Any 0 may likewise be replaced by a 1 with the same probability. For convenience, we denote the field of binary numbers as F.

Formally, a binary *code* with *length n* is a set of vectors chosen from F^n, where each vector represents a corresponding signal. The vectors are called *codewords* or simply *words*. Used in this sense, the term "code" has nothing to do with secrecy. *Error-correcting* codes contain enough redundancy that (limited) random errors can be detected and corrected. A code is *linear*, with *dimension k* if it forms a subspace of dimension k within F^n. Immensely useful examples of linear error-correcting codes were introduced by Hamming and Golay soon after Shannon's publication; these codes are discussed below.

A short linear code As an introductory example, consider the code:

$$S = \{[0, 0, 0, 0], [1, 1, 1, 1], [1, 0, 1, 0], [0, 1, 0, 1]\}.$$

This set of vectors forms a subspace of F^4 with dimension 2, so S is linear with length $n = 4$ and dimension $k = 2$. To satisfy tradition, we stack a basis for a linear code and call it a *generator matrix*. Code S is the rowspace of generator matrix M, below:

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

We could, of course, choose a different basis for S; generator matrices are not unique.

Linear codes are classified by length, dimension, and a third parameter: minimum distance. The *distance* between two codewords is the number of positions in which they differ. For example, $d([1, 0, 1, 0], [1, 1, 1, 1]) = 2$. The *minimum distance* of a code is the smallest distance between two distinct codewords; the minimum distance of example S is $d = 2$. The support of a codeword $\hat{x}$, written $supp(\hat{x})$, is the set of nonzero coordinates in the word. The weight of $\hat{x}$ is $d(\hat{x}, \hat{0}) = |supp(\hat{x})|$, the number of ones it contains. For linear codes, minimum distance and minimum weight are equal. Ordering parameters as (n, k, d), we describe example S as a $(4, 2, 2)$ code.

Self-dual codes Orthogonality is an important concept in coding theory. The ordinary dot-product of vectors (mod 2) is an inner product on F^n. We say vectors $\hat{x}$ and $\hat{y}$ in F^n are *orthogonal* if $\hat{x} \cdot \hat{y} = 0$; orthogonal vectors meet (share ones) in an even number of positions. For a linear code C, the set $C^{\perp} = \{\hat{x} : \hat{x} \cdot \hat{c} = 0 \ \forall \hat{c} \in C\}$ also forms a subspace of F^n, the dual code of C. A code is *self-dual* if $C = C^{\perp}$; our example, S, is self-dual.

Lemma 2. The words of a binary self-dual code C are easily recognized; they are exactly those vectors in F^n which are orthogonal to all rows of any chosen generator matrix for C.

Our communication channel permits substitution errors during transmission; received vectors may not even be words in the code. *Nearest-neighbor decoding* attempts to correct such errors by treating a received vector $\hat{x}$ as a codeword $\hat{c}$ which minimizes $d(\hat{x}, \hat{c})$. Code S is unsuitable for such use; the vector $[0, 0, 0, 1]$ has two nearest neighbors, $[0, 0, 0, 0]$ and $[0, 1, 0, 1]$.

A linear error-correcting code Nearest neighbor decoding allows reliable correction of $\lfloor \frac{d-1}{2} \rfloor$ errors per codeword. With minimum distance $d = 2$, our example S is zero error-correcting. We can increase the minimum distance by lengthening the codewords. Code T extends S to a $(6, 2, 2)$ linear code with minimum distance 3:

$$T = \{[0, 0, 0, 0, \mathbf{0}, \mathbf{0}], [1, 1, 1, 1, \mathbf{1}, \mathbf{1}], [1, 0, 1, 0, \mathbf{1}, \mathbf{0}], [0, 1, 0, 1, \mathbf{0}, \mathbf{1}]\}.$$

If the vector $\hat{x} = [0, 0, 0, 1, 0, 0]$ is received, we now assume $\hat{x}$ represents the unique nearest neighbor $[0, 0, 0, 0, 0, 0]$ in T. Code T is one error-correcting, but is no longer self-dual. It is also quite inefficient, requiring 6 digits to represent only 4 distinct signals. Error-correcting codes exist which are both self-dual and efficient; the Hamming and Golay codes (below) are famous examples.

The automorphism group of a code A (linear binary) code C is equivalent to another code D if C can be obtained from D via a permutation of the coordinates of all codewords. Formally, coordinate permutations form distance-preserving vector-space homomorphisms on F^n; as such, they map (n, k, d) codes to (n, k, d) codes. Recall that code S is generated by

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

Exchanging the second and third coordinates in code S produces an equivalent code, generated by the matrix

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

It is common to select a single, standard code from any set of equivalent codes. In this article we approach equivalence a little differently. Equivalent codes represent the same information and structure, expressed as different subspaces of F^n. We view two equivalent codes as two *models* for that underlying structure. Different models reveal different properties of the structure.

An *automorphism* of a code is a coordinate permutation which preserves the original code (as a set of vectors). In matrix terms, automorphisms transform generator matrices into new generator matrices for the same code. The automorphisms of a code C form a group under composition, denoted by $Aut(C)$.

The group $Aut(S)$ consists of those permutations which either fix or exchange $[1, 0, 1, 0]$ and $[0, 1, 0, 1]$. (The codewords $[0, 0, 0, 0]$ and $[1, 1, 1, 1]$ will be fixed by every permutation.) The identity permutation, fixing every column, is in $Aut(S)$. Automorphisms which fix the last column must also fix the second, and those fixing the third column also fix the first. These rules describe the automorphisms $(1, 3)$ and $(2, 4)$. Some automorphisms fix no columns: $(1, 3)(2, 4)$ fixes both words, while $(1, 2, 3, 4)$, $(1, 4, 3, 2)$, $(1, 4)(2, 3)$, and $(1, 2)(3, 4)$ exchange them. Two generators are sufficient to produce the 8 elements of $Aut(S)$; specifically, $Aut(S) = \langle(1, 2, 3, 4), (1, 3)\rangle$.

The Hamming code and its automorphisms While Shannon was introducing a theory of digital communication, one of his colleagues at Bell Telephone Laboratories was developing an extremely practical error-correcting code [**2**]. Hamming proposed a binary linear code which could correct a single substitution error in any codeword [**13**].

Hamming's code can be built in simple stages from our code S. Define the blocks:

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \overline{I} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \text{ and } \quad J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Substitute these blocks into the generator matrix M

$$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \longrightarrow \begin{bmatrix} I & 0 & \overline{I} & J \\ 0 & I & J & \overline{I} \end{bmatrix},$$

obtaining a commonly used generator matrix:

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

In modern terminology, the rowspace of N forms a self-dual (8,4,4) code, denoted H. The code is doubly-even, as all codewords have weights which are multiples of 4. We call this the *block-substitution model* of the extended Hamming code. (The reason for calling it "extended" will be explained soon.)

The automorphism group, $Aut(H)$, is surprisingly large, with 1344 elements. Some of these automorphisms are apparent from the generator matrix N. Simply cycling the rows of N reflects the automorphism $\epsilon = (1, 2, 3, 4)(5, 6, 7, 8)$. A normal subgroup of eight elements can be "lifted" from automorphisms of the underlying code S. For example, if we apply the permutation $(1, 3)(2, 4)$ to the blocks which built N, we obtain an automorphism of H: $(1, 5)(2, 6)(3, 7)(4, 8)$.

A code automorphism need not preserve any particular generator matrix. The permutation $\eta = (1, 2, 3, 4, 6, 8, 5)$ is an automorphism of code H, but distorts matrix N.

To explain η, and to develop the full group $Aut(H)$, we consider another model of the same code.

The set of quadratic residues (mod 7) is the set of numbers q such that the equation $x^2 \equiv q \pmod 7$ has a solution. This set, $\{1, 2, 4\}$, is used to construct the *quadratic residue model* of Hamming's code. Create a binary vector of length 7 by placing ones in positions 1, 2, and 4. "Extend" the vector by placing an extra one in an eighth position; this parity-check position ensures the code will be even. Produce consecutive rows of a matrix by applying the coordinate permutation $\theta = (1, 2, 3, 4, 5, 6, 7)(8)$. The resulting matrix

$$P = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

generates a code, V, with dimension 4. (Further applications of θ result in vectors which are linear combinations of rows in P. Thus, θ is an automorphism of V.)

To verify V is equivalent to the block-constructed code H, we only need to find a permutation ω which embeds H into V.

Equivalence of the two models If a mapping ω takes the code H into the the code V, it cannot be unique. (For any $\delta \in Aut(V)$, the composition $\delta \cdot \omega$ will achieve the same goal.) We should have some flexibility in constructing ω. In particular, we will suppose that ω fixes the first four coordinates of H. (This is something of a gamble, but it turns out to be an acceptable restriction.)

We rely on a well-known combinatorial structure associated with the Hamming code. In combinatorics, a $t-(v, k, \lambda)$ design is a collection of $k-$subsets called "blocks," chosen from a set of v elements called "varieties." Each $t-$subset of varieties occurs in exactly λ blocks.

Any chosen model, C, of the Hamming code contains a design. The coordinates $\{1, 2, \ldots, 8\}$ are the varieties, while the supports of (nonzero) minimum-weight words form the blocks. Specifically, $\{supp(\hat{x}) : \hat{x} \in C, d(\hat{x}, \hat{0}) = 4\}$ is a $3-(8, 4, 1)$ design, which is unique up to permutation of coordinates [**15**, pp. 103-104]. In other words, every 3-tuple of nonzero coordinates defines a unique minimum-weight vector in the Hamming code.

We construct $\omega : H \to V$ by assuming both codes are models of the Hamming code. We match minimum-weight vectors between the codes, using three nonzero coordinates to determine unique pairings. Vectors in a pair each have a fourth non-zero position. Comparison of these positions describes the action of ω on a single coordinate. If the resulting map takes H into V, then we will know that our assumption was correct.

We work with the matrices N and P, which generate H and V, respectively. Suppose that ω fixes the first 4 coordinates of H. We use these 4 coordinates to pair vectors in $row(N)$ with vectors in $row(P)$. As an example, the sum of the second, third, and fourth rows of N is $[0, 1, 1, 1, \mathbf{1}, 0, 0, 0]$. This is mapped to the exactly one vector, the sum of the second and fourth rows in P: $[0, 1, 1, 1, 0, 0, \mathbf{1}, 0]$. We conclude that ω maps coordinate 5 to coordinate 7. Three similar pairings appear below:

$$\omega : [1, 1, 0, 1, 0, 0, \mathbf{1}, 0] \to [1, 1, 0, 1, 0, 0, 0, \mathbf{1}],$$

$$\omega : [1, 1, 1, 0, 0, 0, 0, \mathbf{1}] \to [1, 1, 1, 0, 0, \mathbf{1}, 0, 0],$$

$$\omega : [1, 0, 1, 1, 0, \mathbf{1}, 0, 0] \to [1, 0, 1, 1, \mathbf{1}, 0, 0, 0].$$

From these pairings, we conclude ω maps coordinate 5 to coordinate 7, coordinate 7 to coordinate 8, coordinate 8 to coordinate 6, and coordinate 6 back to coordinate 5. Our permutation must be $\omega = (5, 7, 8, 6)$. Code V is self-dual, ω takes $row(N)$ into $row(P)$, and our two codes have the same dimension. From Lemma 2, we may conclude $\omega(H) = V$. Codes H and V are equivalent models of Hamming's code.

The permutation $\theta = (1, 2, 3, 4, 5, 6, 7)(8)$, used to create the quadratic residue model, is an automorphism of that model. We claimed, above, that the block-substitution model also admitted an automorphism η of order 7. The composition $\omega^{-1} \cdot \theta \cdot \omega$ maps words in the block-substitution model to words in the quadratic residue model, applies θ to those words, then maps the results back. The automorphism η is simply θ translated into the block-substitution model:

$$\begin{aligned}\eta &= \omega^{-1} \cdot \theta \cdot \omega \\ &= (1, 2, 3, 4, 6, 8, 5).\end{aligned}$$

Our two models of the extended Hamming code reveal distinct features of its internal structure. Think of the models as x-rays taken from two different viewpoints. These two views, expressed as automorphisms η and $\epsilon = (1, 2, 3, 4)(5, 6, 7, 8)$, are sufficient to describe the entire group $Aut(H)$. Formally, $Aut(H) = \langle \epsilon, \eta \rangle$, a group with 1344 elements.

3. THE EXTENDED GOLAY CODE AND $\mathbf{M}_{24}$ Hamming's work was paralleled in an incredibly brief article by Golay [**11**], which introduced a multiple-error-correcting code of length 23. The code was later extended with a twenty-fourth (parity check) column, producing a (24, 12, 8) three-error correcting code.

The extended binary Golay code has been extensively used in deep-space communications. In particular, it was used by the Voyager mission to encode color photographs of Jupiter and Saturn [**16**]. This code has many astonishing properties.

Theorem 1 (Pless [20]). *Let C be a linear binary* $(24, 12, d)$ *code. Then the following statements are equivalent:*

1. *the minimum weight of C is $d = 8$;*
2. *C is equivalent to the Golay code.*

Theorem 2 (Huffman, Pless [14]). *The full automorphism group of the extended binary Golay code is isomorphic to M_{24}.*

Lemma 3. Let M be the set of minimum-weight words in a model of the extended Golay code. The set $\{supp(\hat{x}) : \hat{x} \in M\}$ forms a $5 - (24, 8, 1)$ design. Furthermore, this design is unique up to permutation of coordinates [**20**].

The design described in Lemma 3 is named after Witt, who showed it to be unique long before the Golay code was created [**26**]. The proof of Theorem 1 was based on this lemma.

The primary goal of this article is to construct the automorphism group M_{24} by correlating two simple models of the Golay code. The first is essentially Golay's original model, based on quadratic residues mod 23. We develop the second via further substitutions into the block-substitution Hamming code. These models provide two different glimpses into the automorphism group M_{24} – different enough to construct the group.

Constructing the extended Golay code (I) The *quadratic residue model* of the extended Golay code begins with the set $\{1, 2, 3, 4, 6, 8, 9, 12, 13, 16, 18\}$ – the quadratic residues (mod 23). A binary vector of length 23 is created with ones in these positions; an additional 1 appears in position 24. Consecutive rows of a generator matrix, Q, are produced using the coordinate permutation

$$\sigma = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23)(24),$$

as follows

$$Q = \begin{bmatrix} 11110101100110010100000 & 1 \\ 01111010110011001010000 & 1 \\ 00111101011001100101000 & 1 \\ 00011110101100110010100 & 1 \\ 00001111010110011001010 & 1 \\ 00000111101011001100101 & 1 \\ 10000011110101100110010 & 1 \\ 01000001111010110011001 & 1 \\ 10100000111101011001100 & 1 \\ 01010000011110101100110 & 1 \\ 00101000001111010110011 & 1 \\ 10010100000111101011001 & 1 \end{bmatrix}.$$

The rowspace of Q forms a self-dual code, with length 24, dimension 12, and minimum weight 8. This is Golay's original construction; we will call this model R.

The twelve rows of the generator matrix Q were produced by repeated applications of the permutation σ. Note that further applications of σ produce vectors which are linear combinations of those twelve rows. Therefore, we make the following trivial (but useful) observation about R.

Lemma 4. The permutation

$$\sigma = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23)(24)$$

is an automorphism of the quadratic residue model of the extended Golay code.

Constructing the extended Golay code (II) We now propose an alternative model of Golay's code, based on an interesting fact. As a vector space, the extended Golay code is equivalent to a direct sum of three copies of Hamming's code.

We build Golay's code in stages from the generator matrix, M, of our trivial self-dual code S. Recall that the block-substitution model of Hamming's code came from substitutions into M:

$$M = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix},$$

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

Perform similar substitutions into matrix H, using 3×3 blocks:

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad ; \quad \overline{I} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \quad ; \quad J = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix};$$

to create the matrix G, below.

$$G = \begin{bmatrix} I & 0 & 0 & 0 & \overline{I} & I & I & J \\ 0 & I & 0 & 0 & J & \overline{I} & I & I \\ 0 & 0 & I & 0 & I & J & \overline{I} & I \\ 0 & 0 & 0 & I & I & I & J & \overline{I} \end{bmatrix}.$$

In expanded form, we have:

$$G = \begin{bmatrix} 100 & 000 & 000 & 000 & 011 & 100 & 100 & 111 \\ 010 & 000 & 000 & 000 & 101 & 010 & 010 & 111 \\ 001 & 000 & 000 & 000 & 110 & 001 & 001 & 111 \\ & & & & & & & \\ 000 & 100 & 000 & 000 & 111 & 011 & 100 & 100 \\ 000 & 010 & 000 & 000 & 111 & 101 & 010 & 010 \\ 000 & 001 & 000 & 000 & 111 & 110 & 001 & 001 \\ & & & & & & & \\ 000 & 000 & 100 & 000 & 100 & 111 & 011 & 100 \\ 000 & 000 & 010 & 000 & 010 & 111 & 101 & 010 \\ 000 & 000 & 001 & 000 & 001 & 111 & 110 & 001 \\ & & & & & & & \\ 000 & 000 & 000 & 100 & 100 & 100 & 111 & 011 \\ 000 & 000 & 000 & 010 & 010 & 010 & 111 & 101 \\ 000 & 000 & 000 & 001 & 001 & 001 & 111 & 110 \end{bmatrix}.$$

The matrix G generates a self-orthogonal code, B, with dimension 12 and minimum distance $d = 8$ [**7**]. By Theorem 1, we are justified in calling this the *block-substitution model* of Golay's code. Again, we make a relatively trivial observation.

Lemma 5. The permutation $\rho = (1, 2, 3)(4, 5, 6) \ldots (22, 23, 24)$ is an automorphism of the block-substitution model of the extended Golay code.

Theoretically, our models of Golay's code must be equivalent. Unfortunately, this is not enough; we will need a specific equivalence map.

Equivalence of the Golay models We wish to link the block-constructed model of Golay's code to the quadratic residue model. Symbolically, we are seeking a permutation of coordinates $\chi : B \to R$.

In section 2, we derived an equivalence mapping between two models of the Hamming code. A embedded combinatorial design allowed unique matchings of codewords with their images. Our current situation is very similar. Readers wishing to avoid a detailed derivation of the mapping χ should feel free to skip ahead. Our narrative continues with Lemma 6, on page 10.

Lemma 3 provides a $5 - (24, 8, 1)$ design embedded in the Golay code. The words of weight 8 contain a basis for the Golay code [**19**], and they are uniquely determined by five nonzero coordinates. We use fixed 5-tuples to match words $\hat{w} \in B$ with their images $\chi(\hat{w}) \in R$, then reverse-engineer χ from those pairings.

Mathieu's 1873 publication showed that his new group was 5-transitive. For ordered 5-tuples $x = (x_1, x_2, x_3, x_4, x_5)$ and $y = (y_1, y_2, y_3, y_4, y_5)$ chosen from $\{1, 2, \ldots, 24\}$, there is some element of M_{24} taking x to y.

Since B and R are models of the Golay code, we know $Aut(B) \cong Aut(R) \cong M_{24}$, and at least one equivalence map $\kappa : B \to R$ exists. For any $\alpha \in Aut(B)$ and $\beta \in Aut(R)$, the composition $\chi = \beta \cdot \kappa \cdot \alpha$ is also an equivalence map. Whatever permutation is represented by $\kappa \cdot \alpha$, the 5-transitivity of M_{24} guarantees a choice of β so that χ fixes coordinates 1 through 5.

Although our models were constructed very differently, they are not disjoint. The intersection code, $B \cap R$, has dimension 2, and is generated by $\hat{1}$ and $\hat{z} = [0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 0, 0, 1, 1, 0, 1, 1]$. It seems reasonable to assume that χ fixes $B \cap R$ in general, and $\hat{z}$ in particular.

The 5-transitivity of $Aut(R)$ guarantees the existence of some χ fixing coordinates 1 through 5. We can do better, as $Aut(B)$ is also 5-transitive. A total of ten fixed coordinates should be possible, although we cannot promise they will be consecutive.

Our assumption that χ fixes $\hat{z} \in B \cap R$ effectively partitions the set of coordinates. The nonzero coordinates in $supp(\hat{z})$ cannot be interchanged with those where $\hat{z}$ contains zeros. Assuming that χ has at least five fixed points allows a finer partition. We define three subsets of $\{1, 2, \ldots, 24\}$; our mapping χ may permute coordinates within these sets, but not between them. Let F be the set of fixed points for χ. Let C be the set of coordinates, not fixed by χ, where $\hat{z}$ is nonzero: $C = supp(\hat{z}) \setminus F$. Finally let D be the remaining coordinates: $D = \overline{C} \setminus F$.

We are expecting to identify ten fixed points, but we only know five of them at this time. For the moment, we can only say:

$$F \supseteq \{1, 2, 3, 4, 5\},$$
$$C \subseteq \{6, 9, 10, 15, 16, 17, 20, 21, 23, 24\},$$
$$D \subseteq \{7, 8, 11, 12, 13, 14, 18, 19, 22\}.$$

Codes B and R contain unique words of weight 8 starting with five consecutive ones. Matching these words starts the process of determining χ:

$$\chi : [1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1]$$
$$\to [1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0].$$

We find that $\chi : \{18, 21, 24\} \to \{16, 18, 21\}$. Now $18 \notin C$, so 18 is not mapped to 16 or 21. We now know $18 \in F$, leaving $\chi : \{21, 24\} \to \{16, 21\}$. We are expecting additional fixed points, so we guess that 21 is fixed, and $\chi : 24 \to 16$.

We now have seven fixed coordinates: $F \supseteq \{1, 2, 3, 4, 5, 18, 21\}$. As this collection grows, we are able to identify more matchings. We concentrate on words whose support meets our (rapidly shrinking) set C in only one coordinate. Each code has sixteen words satisfying this requirement and intersecting our known subset of F in exactly four coordinates. As paired 5-tuples give unique matchings, paired 4-tuples yield two choices. In the example below, nonzero coordinates falling in C are highlighted. The pair of vectors

$$\hat{b_1} = [0, 0, 1, 1, 1, 0, 1, 1, \mathbf{1}, 0, 0, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]$$

and

$$\hat{b_2} = [0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, \mathbf{1}, 0, 1, 0, 0, 0, 1, 0, 0]$$

in B map to the pair of images

$$\hat{r_1} = [0, 0, 1, 1, 1, 0, 0, 0, 0, \mathbf{1}, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0]$$

and

$$\hat{r_2} = [0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, \mathbf{1}].$$

We conclude $\chi : \{9, 16\} \to \{10, 24\}$. Eight similar pairings yield four distinct facts:

$$\chi : \{9, 16\} \to \{10, 24\},$$
$$\chi : \{10, 15\} \to \{6, 9\},$$
$$\chi : \{6, 17\} \to \{17, 20\},$$
$$\chi : \{20, 23\} \to \{15, 23\}.$$

Again, we expect considerable flexibility in choosing χ. We already know 24 is mapped to 16; it appears that 16 could be mapped back to 24. A similar situation is possible with 9 and 10. Since we have no information about 17, perhaps it is fixed. We now know χ maps 15 to 6 and 6 to 20. Coordinate 20 is either mapped to 23 or back to 15. Choosing the first case, we have several closed cycles in χ: (6, 20, 23, 15); (9, 10); and (16, 24).

The action of χ on C is determined. Our collection of established fixed points now contains eight elements. It is large enough to uniquely determine numerous vectors whose supports meet D in only one point. Matching these vectors between the codes, we can predict the action of χ on coordinates in D. The actions of χ on F, C, and D are then determined, yielding a potential equivalence map.

Lemma 6. The permutation

$$\chi = (6, 20, 23, 15)(7, 12, 11, 8, 22, 19)(9, 10)(16, 24)$$

is an equivalence map from the block-constructed model (B) to the quadratic-residue model (R) of the extended Golay code. It fixes the intersection of those two codes.

Proof. As we are working with self-dual codes, Lemma 2 makes it easy to check that χ maps B into R. The dimensions of the two codes are equal, so they are equivalent via χ. ∎

Building the sporadic group $\mathbf{M}_{24}$ We now have two automorphisms arising from different models of the extended Golay code. The permutation σ cycles the coordinates $1, 2, \ldots, 23$; it is an automorphism of the quadratic residue model. The permutation ρ is an automorphism of the block-substitution model; it partitions the coordinates $1, 2, \ldots, 24$ into consecutive 3-cycles. We also have the map χ, which links the two models.

Conjugation by χ transforms σ into an automorphism of the block-substitution model; specifically, we define $\tau = \chi^{-1} \cdot \sigma \cdot \chi$. The group generated by ρ and τ is simple, with 244,823,040 elements; these facts were verified with the software package GAP [**10**]. By Lemma 1, the group must be isomorphic to M_{24}. Alternatively, we could argue that $\langle \rho, \tau \rangle$ is a subgroup of the automorphism group of Golay's code. As its order matches the order of M_{24}, Theorem 2 confirms that the groups are isomorphic.

Theorem 3. *The permutations*

$$\tau = \chi^{-1} \cdot \sigma \cdot \chi$$
$$= (1, 2, 3, 4, 5, 15, 19, 11, 10, 9, 12, 7, 13, 14, 23, 24, 17, 18, 22, 6, 21, 8, 20)(16)$$

and

$$\rho = (1, 2, 3)(4, 5, 6) \ldots (22, 23, 24)$$

generate the sporadic group M_{24}.

4. VISUALIZING THE MATHIEU GROUPS We claim this construction of M_{24} is the simplest possible. As M_{24} is not cyclic, it requires at least two generators. Our generators are obvious from the models they are based on, and those models are easily constructed from rather simple ideas. Perhaps more importantly, this construction of M_{24} is easy to represent visually.

We represent M_{24} as a permutation group on a collection of 24 beads. Place 23 of these beads in a ring, with the remaining bead in the center. Label the beaded ring, in order, with the numbers appearing in permutation τ. Reserve the number 16 to label the center bead.

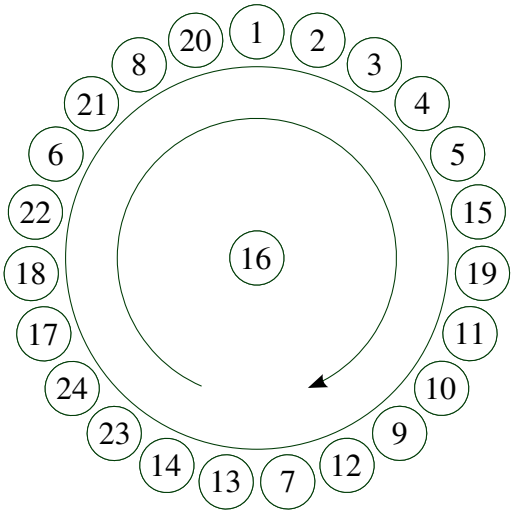

Figure 1. The permutation τ

Figure 1 shows the permutation τ, which fixes bead 16. Figure 2 shows the permutation ρ, which moves beads on triangles. It is easy to see why M_{24} is so large. The eight triangles comprising ρ display eight very different behaviors with respect to our ring of beads. There is little, if any, correspondence between τ and ρ, allowing seemingly endless variety in combinations of the two permutations.

The groups M_{23} and M_{22} We have constructed M_{24}, the largest member of the Mathieu family consisting of M_{24}, M_{23}, M_{22}, M_{12}, and M_{11}. What can we say about its relatives?

The Mathieu groups were introduced in two articles by Emile Mathieu. The first article, in 1861, dealt primarily with M_{12}, a 5-transitive group of permutations on 12

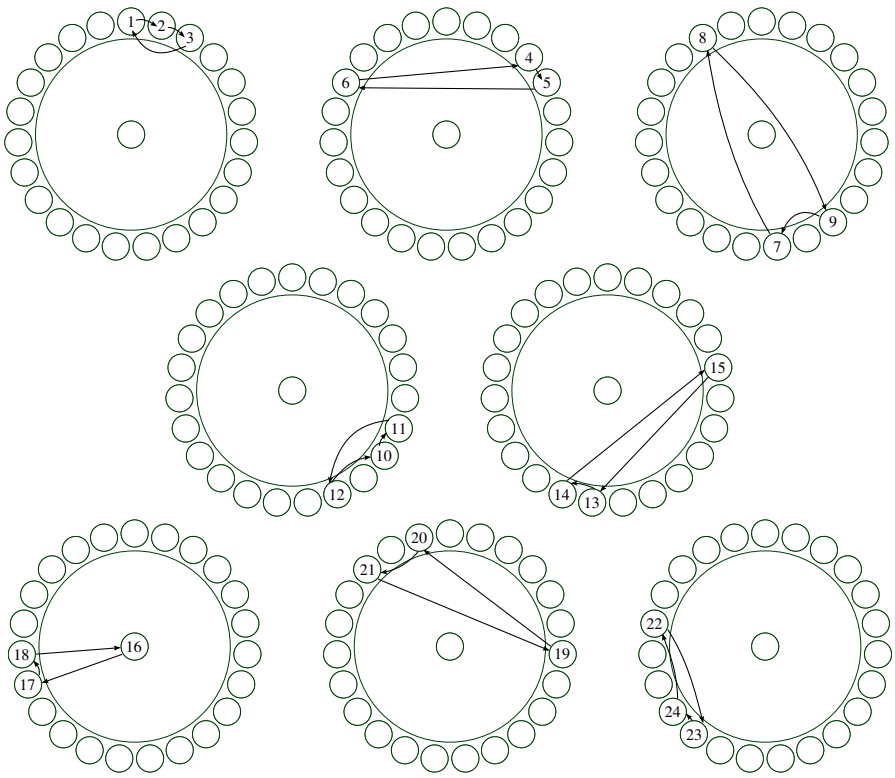

Figure 2. Disjoint cycles comprising the permutation ρ

coordinates [**17**]. The second article, twelve years later, dealt with the much larger group M_{24} [**18**]. Other authors showed, circa 1900, that these were simple groups [**9**, p. 209]. In other words, the Mathieu groups were the first sporadic groups. Aside from the full symmetric and alternating groups, S_n and A_n, these are the only 4- and 5-transitive groups [**9**, p. 177].

Mathieu's 1873 article introduced a 4-transitive permutation group on 23 coordinates, now known as M_{23}. That 4-transitive group required two generators, a and b, appearing on page 2 of this article. He then extended M_{23} by adding a third generator, c, forming the 5-transitive M_{24}. This definition was not entirely convincing in its time; at least one paper was published denying the existence of M_{24}. The group's nature was firmly established in 1938, when Witt constructed M_{24} as the permutation group of the combinatorial design bearing his name [**24**, p. 191].

The two groups M_{23} and M_{22} are usually described as stabilizer subgroups within M_{24}. Their subscripts refer to the number of permuted coordinates. The larger of these, M_{23}, is generated by stabilizing any single coordinate in $\{1, 2, \ldots, 24\}$. It has $\frac{|M_{24}|}{24}$ elements [**9**, p. 177].

Using our beaded-ring description of M_{24}, construction of M_{23} is easy. Recall that $M_{24} \cong \langle \rho, \tau \rangle$. Permutation τ already fixes the bead in the center of our ring, number 16. We replace ρ, which does not fix bead 16, by $f = \rho \cdot \tau \cdot \rho$. This new permutation moves bead 16 onto position 17 of the ring, rotates it to the position previously held by bead 18, and then cycles it back to the center. We summarize this computationally:

$$f(16) = [\rho \cdot \tau \cdot \rho](16)$$

$$= [\rho \cdot \tau](17)$$
$$= [\rho](18)$$
$$= 16.$$

The automorphisms τ and f generate the full stabilizer of bead 16; this is the Mathieu group $M_{23} \cong \langle \tau, \rho \cdot \tau \cdot \rho \rangle = \langle \tau, f \rangle$.

The group M_{22} is usually defined as the stabilizer, within M_{23}, of any additional coordinate from $\{1, 2, \ldots, 24\}$. It is simple, 3-transitive on 22 points, and has $\frac{|M_{24}|}{24 \cdot 23}$ elements [**9**, p. 177].

To construct M_{22} from M_{23}, we start with the generator f. Note that f fixes both beads 16 and 1. We replace τ by an permutation which fixes these same beads. This new permutation uses τ^{-1} to rotate bead 1 a few positions counter-clockwise, then bounces bead 1 back to its original position by applying combinations of f and τ. It is mildly entertaining to verify that $g = \tau \cdot f^{-1} \cdot \tau \cdot f \cdot \tau^{-8}$ performs this strange task. The Mathieu group M_{22} is isomorphic to $\langle f, g \rangle$.

One might ask how far this production of sporadic stabilizer subgroups can continue. The sequence is interrupted by the stabilizer of three coordinates, which is isomorphic to a projective linear group [**24**, p. 190]. With a little extra care, however, we can move on to a sporadic stabilizer of 12 coordinates.

The remaining Mathieu groups Mathieu's first article introduced M_{12} and M_{11} as multiply-transitive permutation groups acting on 12 and 11 coordinates, respectively. Their constructions were very similar to the constructions for M_{24} and M_{23}. The smaller group, M_{11}, was generated by two permutations. The larger group required a third generator [**9**, p. 209]. A more modern description would say that M_{11} is the stabilizer subgroup, within M_{12}, of a single coordinate.

Both groups are isomorphic to subgroups of M_{24}. Specifically, if $\hat{w}$ is any weight-12 word in Golay's code, then the subgroup stabilizing $supp(\hat{w})$ is isomorphic to M_{12} [**9**, p. 206]. Stabilizing any thirteenth coordinate would produce M_{11}.

To construct M_{12}, we could return to the vector $\hat{z} \in B \cap R$. The subgroup of permutations in $Aut(B)$ fixing $supp(\hat{z})$ is isomorphic to M_{12}. After computing M_{12}, we could derive M_{11} by fixing one more coordinate. Modeling this process with the ring-of-beads imagery might be difficult.

All members of the Mathieu family exist as subgroups within M_{24}. This was first proven by Frobenius, and apparently was not known by Mathieu himself [**24**, p. 190]. Oddly, M_{12} cannot be directly constructed the way we constructed M_{24}. Our approach for M_{24} relied on a quadratic residue code of length 23; no such code exists for length 11.

5. CONCLUSION There is an extensive literature on connections between the Mathieu groups and geometry, combinatorics, and coding theory. Numerous constructions of the Mathieu groups have been developed, each justified as "simple" and "natural."

The constructions of M_{24}, M_{23}, and M_{22} presented above are simple and natural in the following senses. The set of generators for each group is as small as possible. The generators are easily derived from two models of an important error-correcting code. Finally, the generators are easy to visualize as permutations on a beaded ring.

Although its derivation relied on very simple ideas, the mapping χ which links the two models may seem less natural. Perhaps we could do better. Recall that the automorphism groups $Aut(B)$ and $Aut(R)$ are very, very large. For any $\alpha \in Aut(B)$

and any $\beta \in Aut(R)$, the map $\beta \cdot \chi \cdot \alpha$ also links the two models. Among the many millions of valid choices for $\beta \cdot \chi \cdot \alpha$, there may be one which is obvious in hindsight. Perhaps this simple construction of M_{24} could be made just a little more "natural."

REFERENCES

1. E. Berlekamp, Coding theory and the Mathieu groups, *Information and Control* **18** (1971) 40–64.
2. ———, *Key Papers in The Development of Coding Theory*, IEEE Press, New York, 1974.
3. J. Conway, Three lectures on exceptional groups, in *Finite Simple Groups*, Edited by M. Powell and G. Higman, Academic Press, New York, 1971. 215–246.
4. ———, The Golay codes and the Mathieu groups, in *Sphere Packings, Lattices, and Groups*, Edited by J. Conway and N. Sloane, Springer, New York, 1988. 299–330.
5. R. Curtis, Natural constructions of the Mathieu groups, *Math. Proc. Cambridge Philos. Soc.*, **106** (1989) 423 – 429.
6. ———, Geometric constructions of the 'natural' generators of the Mathieu groups, *Math. Proc. Cambridge Philos. Soc.* **107** (1990) 19–26.
7. M. Derka, *Generator Matrix Based Search for Extremal Self-Dual Binary Error-Correcting Codes*, M.Sc. thesis, Brock University, St. Catharines, ON, 2012.
8. L. Dickson, *Linear Groups*, Leipzig, 1901.
9. J. Dixon, B. Mortimer, *Permutation Groups*, Springer, New York, 1996.
10. The GAP Group, *GAP — Groups, Algorithms, and Programming, Version 4.4.12* (2008) http://www.gap-system.org.
11. M. Golay, Notes on digital coding, *Proc. of the IRE* **37** (1949) 657. Reprinted in *Key Papers in The Development of Coding Theory*, Edited by E. Berlekamp, IEEE Press, New York, 1974.
12. The GUAVA Group, *GUAVA: A GAP4 Package for computing with error-correcting codes, Version 3.12* (2012) http://www.gap-system.org.
13. R. Hamming, Error detecting and error correcting codes, *Bell System Technical Journal* **29** (1950) 147–160. Reprinted in *Key Papers in The Development of Coding Theory*, Edited by E. Berlekamp, IEEE Press, New York, 1974.
14. W. Huffman, V. Pless, *Fundamentals of Error-Correcting Codes*, Cambridge Univ. Press, Cambridge, 2003.
15. W. Huffman, V. Pless (eds), *Handbook of Coding Theory, Volume 1*, Elsevier, New York, 1998.
16. R. Ludwig, J. Taylor, *Voyager Communications*, JPL Deep Space Communications and Navigation Systems Center of Excellence, Pasadena, CA, 2002.
17. E. Mathieu, Memoire sur etude des fonctions de plusieurs quantités, *J. Math Pures. Appl. (Liouville) (2)* **(6)** (1861) 241–323.
18. ———, Sur la fonction cinq fois transitive de 24 quantités, *J. Math Pures. Appl. (Liouville) (2)* **(18)** (1873) 25–46.
19. L. Paige, A note on the Mathieu groups, *Canad. J. Math.* **9** (1957) 15–18.
20. V. Pless, On the uniqueness of the Golay codes, *J. Combin. Theory Ser. A*, (1968) 215–228.
21. J. Rotman, *Advanced Modern Algebra*, Second edition, AMS, Providence, RI, 2010.
22. C. Shannon, A mathematical theory of communication, *Bell System Technical Journal*, **27** (1948) 379–423.
23. R. Stanton, The Mathieu groups, *Canad. J. Math.* **(3)** (1951) 164–174.
24. R. Wilson, *The Finite Simple Groups*, Springer, New York, 2009.
25. E. Witt, Die 5-fach transitiven gruppen von Mathieu, *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg* **12** (1938) 256–264, http://dx.doi.org/10.1007/BF02948947.
26. ———, Uber Steinersche Systeme, *Abhandlungen aus dem Mathematischen Seminar der Universitat Hamburg* **12** (1938) 265 - 275, http://dx.doi.org/10.1007/BF02948948.

PAUL E. BECKER received his PhD in mathematics from Central Michigan University. He is an associate professor of mathematics at Penn State Behrend. In his spare time, he kayaks and operates a family blueberry farm.
School of Science, Penn State Behrend, Erie PA 16563
peb8@psu.edu

MARTIN DERKA earned an M.Sc. degree in computer science at Brock University, and is currently completing a PhD at the University of Waterloo. He also holds a M.Sc. degree in computer science from Masaryk University. Martin enjoys rock and metal music, a good cup of coffee, tech startup culture, and the outdoors.

David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON, N2L 3G1 Canada
mderka@uwaterloo.ca

SHERIDAN HOUGHTEN received her PhD degree in computer science from Concordia University, Montreal. She is a professor, and former department chair, of computer science at Brock University. Her research interests encompass bioinformatics, computational intelligence, coding theory, and combinatorial optimization.
Department of Computer Science, Brock University, St. Catharines, ON, L2S 3A1 Canada
shoughten@brocku.ca

JENNIFER ULRICH received her M.S. degree in mathematics from Texas A&M University. She holds a B.S. degree from Penn State Behrend, where she is currently a lecturer in mathematics.
School of Science, Penn State Behrend, Erie PA 16563
jkm154@psu.edu