An Examination of Codewords with Optimal Merit Factor

Michael W. Cammarano and Anthony G. Kirilusha*

April 6, 1999

Abstract

We examine the codewords with best possible merit factor (minimum sum of squares of periodic autocorrelations) for a variety of lengths. Many different approaches were tried in an attempt to find construction methods for such codewords, or for codewords with good but non-optimal merit factors.

*The authors thank Hew lett-Packard for their generous support during the summer of 1998, and Dr. James Davis, University of Richmond, for his extensive support and assistance.

1 Introduction

1.1 Periodic Autocorrelations and Merit Factors for Binary Sequences

Our general objective this summer was to find construction methods for binary codewords with low autocorrelations. Specifically, we wanted to choose the codewords with the best merit factors for a variety of sequence lengths without resorting to exhaustive search. Initially, we thought that previous work [Bernasconi] would allow for an interesting avenue for investigation, but after acquiring a copy of the paper we decided that further examination of the simulated annealing method would be inappropriate for our circumstances. However, we hoped that an unexpectedly elegant construction might exist, given the encouraging recent discovery relating Golay pairs to Reed-Muller codes [Davis/Jedwab].

At first, our work was equally concerned with minimizing periodic and aperiodic autocorrelations. Both represent measures of the self-similarity of a binary sequence. For a binary sequence d of length N comprised of ± 1, these measures can be simply represented by:

Aperiodic Autocorrelation $X_g(d) = \sum_{m=0}^{N-1-g} d_m d_{m+g}$

Periodic Autocorrelation $C_g(d) = \sum_{m=0}^{N-g} d_m d_{m+g}$, indices taken modulo N.

The parameter g represents the period or offset at which the sequence is compared against itself. The crucial difference between the measures is that the periodic autocorrelation treats the sequence as if it were circular, whereas the aperiodic does not. Note that in the trivial case of $g = 0$, $X_0(d) = C_0(d) = N$. We will only be concerned with $1 \leq g \leq N - 1$. Sequences with low autocorrelations have been sought for some time, and traditionally the problem is stated in terms of maximizing the merit factor F of a sequence, where:

$$F = \frac{N^2}{2\sum_{k=1}^{N-1} C_k(d)^2}$$

Put simply, we wish to find sequences of a particular length with the minimal sum of squares of autocorrelations. For periodic autocorrelations, optimal sequences can be found for a number of lengths thanks to the close relationship between circular sequences and cyclic difference sets [Mertens/Bessenrodt]. For many N, constructions methods are known that yield difference sets of order N. These methods can be directly adapted to form binary sequences of length N with minimal periodic autocorrelations, termed perfect sequences by Mertens and Bessenrodt. However, cyclic difference sets (and hence perfect sequences) exist only for certain values of N, and different approaches are required to handle other values of N. The concept can be extended to so-called almost-perfect seqences, which relate in turn to almost-difference sets [Mertens/Bessenrodt]. However, almost difference sets are quite outside the realm of the established field of difference sets. Without

that broad base of theoretical ! support, almost perfect sequences aren't as readily useful. However, Mertens and Bessenrodt do provide an effective construction for $N = pq$ where p and q are primes.

These difference and almost-difference set techniques still leave many cases unaccounted for, notably $N = 2^m$. Much of our effort was directed at this case, which we hoped might be subject to a radically different approach via a recursive construction. We focused primarily on periodic autocorrelations because of the significant symmetry inherent.

Finally, we should note that the autocorrelation of a sequence is a generalization of the cross-correlation of two sequences. We will write the periodic correlation of two codewords d and e as:

$C_g(d, e) = \sum_{m=1}^{N} d_m e_{m+u}$, indices taken modulo $N + 1$.

2 Linear Codes

Let us consider all the ideal sequences of length $N = 2^m$ for $m = 3$, 4, and 5. It is fairly obvious that the all-0 and the all-1 sequences will not be included among the ideal sequences, since their periodic autocorrelations would be equal to $N - 1$ for every period, thus producing the largest possible sum of squares. This, however, does not necessarily exclude a possibility of linear realtionship between the ideal sequences, which is the question we will examine in this section of the report.

To begin with, let us consider all 32 ideal sequencef for $N = 8$ ($m = 3$). To begin with, we must find an appropriate offset, such that the all-0 and the all-1 sequences will be included in the set of sequences we are dealing with. For the purpose of this explanation we will choose the sequence corresponding to 11 (00001011) to be the offset, which we will subtract from every ideal sequence of length 8. Thus, the set of offsets will include the 00000000 and the 11111111 sequences, which will be obtained by subtracting 11 from itself and from it's inverse. Thus, we can now check and see if there is a linear code underlying the 32 offsets we have obtained. We can see that sequences 11111111 (255), 00010001 (17), 01010101 (85), and 00100111 (39) make up a generator matrix that accounts for half of the offset sequences. However, no linear code underlies the other 16 sequences, and as far as we know there is no sequence which added to the above four will provide a generator matrix that can account for all 32 sequences.

Furthermore, as we went to examine the ideal sequences for $N = 16$ ($m = 4$), it became evident that ideal sequences have virtually no linear relationship to one another. While it was possible to find triplets of offset ideal sequences that would in linear combinations with one another produce 8 offset ideal sequences, no such qudruples (or higher) were found. Although an exhaustive search was not performed, it is clear that no linear code exists that would account for all 256 ideal sequences (see figure blah). Similarly, ideal sequnces do not seem to be cosets of first or second order Reed-Muller code in higher order Reed-Muller. In light of our findings for $m = 3$ and 4, length 32 sequences were not investigated for linearity. Although the number of ideal sequences of length 2^m seems to always be a power of two (which initially suggested a possible linear relationship), this can be explained by the concept of equivalence classes discussed earlier.

3 Recursive Constructions

In working with sequences of length 2^m, it is natural to look for a construction that builds a set of sequences by combining two sequences of length 2^{m-1}. The most simple example is the concatenation of two length-N sequences, a and b, to form a length-$2N$ sequence that we will write $a \mid b$. Thus, concatenating 0011 with 0101 yields 00110101. For any recursive construction to be helpful in approaching autocorrelation problems, we need to know how the correlations of the component sequences interact to yield the autocorrelations of the resulting sequence.

3.1 Simple Concatenation and the Plotkin Construction

Note that the middle autocorrelation of $a \mid b$ is twice the cross-correlation of a and b – that is, $C_{a \mid b}(N) = 2C_{a,b}(0)$. This is a straightforward result: Since subscripts in the periodic autocorrelation are taken modulo $2N$:

$$C_{a \mid b}(N) = \sum_{m=0}^{2N-1} a_{mmodN} b_{mmodN} = 2 \sum_{m=0}^{N-1} a_m b_m$$

Unfortunately, none of the other autocorrelations of $a \mid b$ can be expressed simply in terms of the autocorrelations and cross-correlations of a and b. For that reason, this construction provides little help in generating low autocorrelation sequences.

Closely related is the Plotkin construction, $a \mid a+b$. By logic similar to that above, $C_{a \mid a+b}(N) = N - 2w(b)$. This is slightly more useful than simple concatenation since it depends only on the weight of b rather than on the cross-correlation of a and b. One apparent pattern for lengths 8, 16, and 32 was that the middle autocorrelation for all optimal sequences of a given length is constant - 4 for lengths 8 and 16, and -4 for length 32. By the above formula, therefore, the choices for b are constrained to sequences of a particular weight. This reduces the search space significantly, but not nearly enough to make searching longer sequences feasible. If it were possible to place additional constraints on either a or b, this could potentially be useful. We broke all the optimal codewords for all lengths into their Plotkin components a and b in the hope that additional limiting conditions might emerge, but we weren't able to recognize any further patt! erns in the components.

3.2 Interleave Construction

Interleaving sequences, or alternating the bits of a and b to yield $a_0 b_0 a_1 b_1 \ldots a_{N-1} b_{N-1}$, is a potentially much more useful method in that every autocorrelation of the resulting sequence, aSb, can be expressed as a simple combination of the correlations of the components. Note that $(aSb)_{2j} = a_j$, and $(aSb)_{2j+1} = b_j$. From that observation, two simple formulas follow that suffice to describe all autocorrelations of aSb, one for g-even and one for g-odd.

g-even: $C_{aSb}(2j) = C_a(j) + C_b(j)$

g-odd: $C_{aSb}(2j+1) = C_{a,b}(j) + C_{b,a}(j+1) = C_{a,b}(j) + C_{a,b}(N-j-1)$

One use of this construction that might seem promising would be to interleave sequences a and a' that are Golay pairs, yielding a sequence where all the autocorrelations for even values of g would be zero. Of course, for all odd values of g, the autocorrelations of aSa' would depend on the cross-correlations of a and a'. For example, we might interleave $z = 00010010$ with $z' = 00011101$ (two sequences that comprise a golay pair). The resulting sequence, $(zSz') = 0000001101011001$, is not optimal. It is, however, a Golay sequence of the next higher length. This may or may not seem surprising, but it can be explained by an examination of the effect of the interleaving operation on Reed-Muller codes. Sequences that are Golay pairs with respect to aperiodic autocorrelation seem to also be pairs with respect to periodic autocorrelation, and thus are produced by the same pattern of second order Reed-Muller terms. A series of lemmas exist describing the concatena! tion of two RM codewords in terms of RM of the next higher length. A similar description is possible for the effect of interleaving two RM codewords. For any codeword $a \in RM(2,m)$, $(aS\mathbf{0}) = (ax_m + a) \in RM(2,m+1)$. As an example, consider $a = x_1 \in RM(2,2)$:

$(x_1 \in RM(2,2)S\mathbf{0}) = (0011S0000) = 00001010 = x_1x_3 + x_1 \in RM(2,3)$

Similarly, for any $b \in RM(2,m)$, $(\mathbf{0}Sb) = bx_m \in RM(2,m+1)$. In the above example, 00010010 is the stereotypical Golay coset leader, $x_1x_2 + x_2x_3$, and 00011101 is its pair, $x_1x_2 + x_2x_3 + x_1$. By using the above interleaving properties of RM, we find that:

$(zSz') = x_1x_2x_4 + x_2x_3x_4 + x_1x_2 + x_2x_3 \ + \ x_1x_2x_4 + x_2x_3x_4 + x_1x_4$

$= x_1x_2 + x_2x_3 + x_1x_4$, which is itself a Golay coset leader in $RM(2,4)$.

The Golay cosets for aperiodic seem to be a subset of those for periodic. Comparable Reed-Muller patterns should exist for the periodic case, although we did not pursue this area of investigation.

We did proceed to break all optimal sequences of lengths 8, 16, and 32 into their interleaved components. The results for length 32 were quite striking: all optimal sequences were composed by interleaving two sequences that we shall call an almost-Golay pair. All but one of the autocorrelations for each pair of component sequences were complementary. The same was true of the components of the length 8 optimal sequences, although not for the length 16, suggesting that almost-Golay pairs might be involved in lengths 2^m for odd m. Consider one of the optimal length-32 sequences:

00100101001000110000010001011111

It is composed by interleaving the following two sequences, which are given together with their autocorrelations:

0100010100000011: 0 4 0 0 -4 4 4 4 4 4 -4 0 0 4 0 (144)
0011000100101111: 0 -4 0 0 4 -4 -4 0 -4 -4 4 0 0 -4 0 (128)

As stated above, the autocorrelations are complementary in all but the middle position. This is true for the components of all the optimal length 32 sequences. Note that not every almost-Golay pair can be interleaved to form an optimal sequence. For any given almost-Golay pair, either of the components can be rotated to form N different sequences, all of which have the same autocorrelations and hence still form a Golay pair with the other component. Only a small number of these almost-Golay pairs result in an optimal sequence when interleaved, however. Nevertheless, even though only a small fraction of almost-Golay pairs produce an optimal codeword when interleaved, this would still be an overwhelming improvement over exhaustive search, provided that a technique existed for generating almost-Golay pairs as easily as the technique for generating true Golay pairs. Unfortunately, appealing to Reed-Muller representations of almost-Golay sequences did not reveal anything like the! simple patterns tha t exist for typical Golay sequences. If a construction for almost-Golay pairs existed that was as powerful as the construction for Golay pairs, and if this pattern holds, then optimal sequences could be found for higher lengths 2^m, m odd.

4 Individual Investigations

4.1 Sequences Corresponding to Prime Numbers

Let us consider all the seuences for which $N = 2^m$. By reviewing the formula for periodic autocorrelation presented above, we can convince ourselves that a cyclic shift of a codeword will not alter the values of its periodic autocorrelations. Same holds true if we were to invert the sequence (flip every bit) and then perform cyclic shifts on the inverse. Finally, taking a mirror image of the sequence will also leave the values of periodic autocorelations unchanged. We define taking mirror image as rewriting the sequence from right to left, making the last bit first and first bit last. Thus, a mirror image of 0001011 would be 1101000. Therefore, we can split the codespace into equivalence classes of $4N$ and $2N$ (if mirror image is equivalent to one of the cyclic rotations) sequences which all have the same values for periodic autocorrelations of periods 1 to $N - 1$.

Let us remind ourselves that the purpose of this project was to investigate binary sequences which yield the smallest sum of squares of autocorrelations compared to other sequences of the same length. For $m = 3$, there are 32 sequences which achieve the minimum value of 16 (to which we will refer as ideal sequences), while for $m = 4$ there are 256 sequences which yield the value of 48, and for $m = 5$ (maximum value of m we investigated) there are 2048 sequences which give the value of 80. If we assume that the mirror image of an ideal sequence is never equivalent to a cyclic rotation of the original sequence, we may now split all ideal sequences into equivalence classes of $4N$ sequences. This suggests 1 equivalence class for $m = 3$, 4 equivalence classes for $m = 4$, and 16 equivalence classes for $m = 5$. We have performed exhaustive searches on sequences of lenght 8 and 16 and were able to confirm the existence of exactly 1 and 4 equivalence classes respectively, ! while a partial search of length 32 sequences has yielded 16 "class representatives", from which we were able to generate and confirm all 2048 possible ideal sequences. If we consider the ideal sequences for $N = 8$ ($m = 3$), they can all be generated from sequence 0001011, which is equivalent to 11 (prime number). Furthermore, if we consider the four equivalence classes of ideal sequences for $N = 16$ ($m = 4$), every equivalence class has at least one prime number representative - i.e. at least one sequence in each of the four classes is equivalent to a prime number. Thus, one can generate all the ideal sequences for $N = 16$ by finding all the sequences equivalent to 0000110010111001, 0000111011101101, 0001001111101011, or 0010100111110011 by using the algorithm of cyclic shifting, inverting, and "mirroring" the sequence.

This data led us to believe that there may be a connection between prime numbers and ideal sequences - since periodic autocorrelation tends to measure periodicity of the sequence, one might have concluded that prime numbers above a certain value (depending on the value of N) will be equivalent to binary sequences with lowest possible periodicity. With that in mind we have investigated the ideal

sequences of length 32 (m = 5), and through partial computerized search were able to generate all 2048 ideal sequences of the appropriate length. Unfortuantely, NO ideal sequence of length 32 were prime. Since no conclusive data is available on length 64 sequences ($m = 6$).

4.2 Pursuing an Algebraic Structure for Periodic Sequences

Taking note of the observations regarding sequences whose decimal representations were prime, it seemed interesting to consider the effect of conventional multiplication on binary sequences and their autocorrelations. Customary arithmetic multiplication of binary sequences of length N is easiest to think of as the sum of a series of shifts of one of the codewords, as in the following example:

$$
\begin{array}{l}
\quad 0001001 = 18 = 2^0 + 2^3 \\
\times 0001010 = 10 \\
- - - - - - - \\
\quad 00001010 = 10 \times 2^0 = 10 \\
+00001010000 = 10 \times 2^3 = 80 \\
- - - - - - - \\
= 00001011010 = 90
\end{array}
$$

To see how this is of interest in regard to autocorrelations, consider $C_3(0001011)$. This autocorrelation can be thought of as the weight of the (bitwise) sum of d and d rotated by g bits, as shown here:

$$
\begin{array}{l}
\ 0001010 \\
+1010000 \\
- - - - - - \\
= 1011010
\end{array}
$$

Note the similarity to the result of arithmetic multiplication! The point is, there is a strong similarity between the operation performed by arithmetic multiplication of binary sequences and that performed in computing certain autocorrelations. To make the analogy even stronger, we could depart from using arithmetic multiplication and instead use a variety of multiplication chosen to mimic the autocorrelation operation as closely as possible. Arithmetic multiplication differs from Autocorrelation computing in that it uses shifts of a codeword rather than rotations, and that it performs arithmetic addition rather than bitwise addition. If these two aspects are changed, we are left with what we shall call cyclic multiplication. Recall that cyclic codes can be thought of as polynomials modulo $x^N - 1$. The behavior of the cyclic multiplication we wish to consider is equivalent to the multiplication of polynomials modulo $x^N - 1$. The following example may clarify the process:

$$
\begin{array}{l}
\ 00101001 = 2^0 + 2^3 + 2^5 \\
*01011010 = a
\end{array}
$$

– – – – ––
01011010 = a rotated 0 bits
11010010 = a rotated 3 bits
+01001011 = a rotated 5 bits
– – – – ––
= 11000011 bitwise (not arithmetic) sum of the rotations.

The cyclic multiplication operation can be described by the following formula:

$$(a * b)_n = \sum_{i=0}^{N-1} a_{N-1-i} b_{n+i}$$

Having defined this operation, we can observe that it is indifferent to rotation – any rotation of a cyclicly multiplied by any rotation of b will produce a rotation of $a * b$. Similarly, the converse of a cyclicly multiplied by b yields the converse of $a*b$, and vice versa. Since all rotations and converses of a sequence have equivalent autocorrelations, we will simply lump all rotations of a sequence and its converse into an equivalence class and treat that class as a single entity. We can then find that the set of all equivalence classes for a given length of sequence comprise an algebraic structure with respect to cyclic multiplication. Because of the close relationship between the operation of cyclic multiplication and the computation of autocorrelations, it seemed possible that study of this structure might provide insight into the autocorrelation problem. As noted earlier in section 5.1, prime numbers (with respect to conventional multiplication) seemed to be ! involved in optimal autocorrelation sequences. It seems possible that this is not purely coincidental, and that there might be a concep t of primacy within the algebra of cyclic multiplication that would be helpful. However, we were unable to pursue this area any further.

5 Conclusions

Overall, our attempts to better understand optimal sequences met with little success. Nevertheless, we hope that the observations presented about almost-Golay pairs and cyclic multiplication might hold promise for future advances. With further study, these ideas might pan out, despite our lack of immediate success.

References

[Macwilliams/Sloane] F.J. MacWilliams, N.J.A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland Mathematical Library, NY. 1986.

[Bernasconi] Bernasconi. "Low Autocorrelation Binary Sequences: Statistical Mechanics and Configuration Space Analysis" *Journal Physique* 48:559-567, 1987.

[Mertens/Bessenrodt] S. Mertens, C. Bessenrodt. "On the Ground States of the Bernasconi Model" *paper* July 10, 1997.

[Davis/Jedwab] J.A. Davis, J. Jedwab. "Peak-to-Mean Power Control and Error Correction for OFDM Transmission Using Golay Sequences and Reed-Muller Codes" *Electronic Letters* 13 Feb. 1997, Vol. 33, No. 4, pp 267–268.

[Golay] M.J.E. Golay. "Complementary Series" *IRE Trans. Int. Theory* 1961, 7, pp. 82–7.