Properties of interleaved sequences created from m-sequences

by

Kirsten Nelson

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Mathematics

Carleton University

Ottawa, Ontario

Abstract

Maximal-length sequences or m-sequences over a finite field $\mathbb{F}_q$ are a well-known and studied class of sequences with desirable properties such as balance of both individual elements and tuples. Interleaved sequences are created by combining a base sequence $\underline{a}$ of period s and a shift sequence $\underline{e}$ of length T, consisting of elements from $\mathbb{Z}_q \cup \{\infty\}$. This thesis examines interleaved sequences to determine which properties of m-sequences are preserved when the m-sequence is used as a base sequence. First an equivalence relation on shift sequences is defined, with two operations that can be applied to these sequences. Palindromic sequences are defined, and the exact conditions for the interleaved sequence to also be palindromic are given. The period of the interleaved sequence was previously known to divide sT, but this thesis proves that the length of the interleaved sequence can be l then, letting $n = sT/l$ we must have $n \mid T$ and $\gcd(n, s) = 1$. The results of experimentation on using the interleaved sequences to construct covering arrays are given.

Acknowledgements

First and foremost, I thank my supervisors, Daniel Panario and Brett Stevens. Their patience and good humour have made the past two years fly by. Meetings discussing math with other mathematicians is just as amazing as I had hoped it would be, and they have been the biggest part of that. My intention to spend another four or five years with them is the biggest indicator of their quality as supervisors, coaches, editors, and so much more. I hope we share many more pitchers and plates of nachos.

I thank my committee, particularly Lucia Moura and Steven Wang. Their close reading and thoughtful questions improved this thesis on every page.

The hazard of returning to school at my advanced age is that there are a lot of people I need to thank. At Carleton and the University of Ottawa over the past two years, Lucia Moura, Jason Gao, and Steven Wang have widened my knowledge. At the University of Waterloo, I have to thank U.S.R. Murty, Brian Forrest, Naomi Nishimura, and Prabhakar Ragde all for teaching, but also encouraging me and having conversations well beyond the scope of their courses. Blair Madore was a fellow student who I now proudly call a colleague.

Russ and Diane Garrett will always have a special place in my heart for turning me toward mathematics at such a young age, and being exemplary teachers.

The friends I've made in Ottawa (especially Kaitlyn, Juhi, and Thaís) have helped me make the most of these two years. I forget the gap in our ages when we sit down for coffee together. My faraway friends (especially erin-blythe, Mandy, Sayaka, and Sheryl) have nourished me from afar with our virtual chats.

I thank my family for not raising their eyebrows *too* high when I went back to school for the fourth time. Finally, my biggest thanks go to my partner Mike and my son Will. You two inspire me every day. I couldn't have done this without you both cheering me on, commiserating on assignments, and generally being the best men I can imagine in my life.

Contents

1 Introduction

The arrangement of objects in a grid has been a source of puzzles, practical applications, and research for thousands of years [10]. When a grid of numbers has the same sum for all the columns, rows, and diagonals, it is known as a "magic" square. This name gives a flavour for how tantalizing these objects are to mathematicians and puzzlers. Even though there is nothing magic about the square itself, the sense of wonder the first time a child is introduced to one is indeed magical, as is the satisfaction the first time one creates their own, or derives a general pattern to construct these squares.

The first known magic square of order 3 (meaning, the square itself is 3 rows tall by 3 columns wide) is of Chinese origin, and legend says it dates from around 2800 B.C.E., being invented by Fuh-Hi, the mythical founder of Chinese civilization. Leaving legend aside, it is certain that a text from the first century C. E. contains an explicit magic square of order 3 [10]. Magic squares were also known and developed in India. In a work written in 587 CE, a magic square of order 4 is constructed for the purpose of making perfumes using 4 substances selected from 16 different substances [12].

In 1782, Leonhard Euler proposed his *Problème de 36 Officiers* (36 Officers Problem) [2], in which he asked: is it possible to arrange 6 regiments of six officers, each of different ranks, in a 6 by 6 grid so that no rank or regiment will be repeated in any row or column? Euler could not find a solution and proposed that none was possible. Despite the apparent simplicity of the problem, a proof was not found until Gaston Tarry, a French amateur mathematician, provided one in 1901 [11].

Problems such as Euler's found a practical application in the 18th century, when agricultural scientists wanted to design experiments to determine the best growing conditions for wheat. If a field were to be subdivided into a grid, and there were a list of conditions to test, which is the best way to lay out the field? There might be five

levels of watering, high and low land placement, several types of fertilizer, etc. The naive approach is to simply multiply the number of options for each and exhaustively test them all, but this approach creates an unwieldy number of tests [3].

In the 20th century, the mathematical objects used in the testing of wheat became relevant to the testing of software, by generalizing the orthogonal arrays to covering arrays. The number of variables in testing grains pales in comparison to the possible number of interactions in a software program [9]. Covering arrays are lists of tests to be run, ensuring a certain coverage of interactions. In this thesis we aim to advance the knowledge about how to create covering arrays by considering a specific construction method.

Our construction method for covering arrays is based on certain types of sequences over finite fields. Sequences have many applications including pseudorandom number generators, cryptography, communications systems, and testing hardware design [5, Chapter 4]. In our case, we study interleaved sequences, created from existing sequences.

In Section 2, we give background on the finite fields and sequences that are used in this construction, and the current state of covering array research. In Section 3, we define interleaved sequences and their equivalence classes, and present some properties of interleaved sequences. Some of these properties were previously published, and some are new in this thesis. In Section 4, we look at how interleaved sequences can be used to create new covering arrays. Section 5 concludes with some final thoughts and open questions.

2 Background

2.1 Finite Fields

A *finite field* is an algebraic field that contains a finite number of elements. For example, for a prime number p, the set of integers modulo p, denoted by $\mathbb{Z}_p$ or $\mathbb{F}_p$, is a field. The *order* of a field is the number of elements it contains. We can also construct finite fields with an order that is a prime power p^n, usually denoted $\mathbb{F}_{p^n}$. We will commonly use q as the subscript, with the understanding that in the field $\mathbb{F}_q$, q is either prime or a prime power. All results we use from finite fields can be found in Lidl and Neiderreiter [7].

2.1.1 Primitive Polynomials

Polynomials over a finite field are defined similarly to polynomials over the real numbers. However, the structure of finite fields leads to some interesting results that are not obvious by comparing to the real numbers.

A polynomial over a finite field is *irreducible* if it cannot be expressed as the product gh where neither g nor h is a constant. Otherwise, it is *reducible*. If the polynomial has a non-zero constant term, then we also have that the *period* of the polynomial f is the smallest integer t such that $f(x) \mid (x^t - 1)$.

Let α be an element in $\mathbb{F}_{p^n}$. The *minimal polynomial* of α over $\mathbb{F}_p$ is defined as the lowest degree monic polynomial $m \in \mathbb{F}_p[x]$ such that $m(\alpha) = 0$. The minimal polynomial of any element in $\mathbb{F}_{p^n}$ is unique.

A generator of the cyclic group $\mathbb{F}_{p^n}^{\times}$ is called a *primitive element* of $\mathbb{F}_{p^n}$. A *primitive polynomial* is the minimal polynomial of a primitive element of $\mathbb{F}_{p^n}$. All minimal polynomials are irreducible, but not all irreducible polynomials are primitive.

We create an *extension* of a base field by augmenting the base field with a root of an irreducible polynomial of degree n over the base field. Let α be a root of

an irreducible polynomial f of degree n, that is, $f(\alpha) = 0$. We construct $\mathbb{F}_{p^n}$ by $\{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_i \in \mathbb{F}_q\}$. We observe that all finite fields of order q are isomorphic [5, Chapter 3].

2.1.2 The Trace Function

The trace is a useful function that takes as input the elements of an extension field, and returns elements of the base field. It is instrumental in constructing various types of sequences, such as m-sequences.

Definition 1. Let q be a prime. For $\alpha \in F = \mathbb{F}_{q^n}$ and $K = \mathbb{F}_q$, the *trace function* $Tr_{F/K}(x), x \in F$, is defined by

$$Tr_{F/K}(x) = x + x^q + x^{q^2} + \cdots + x^{q^{n-1}}.$$

We can also use this definition to calculate the trace from any extension field to a subfield, not necessarily the base field. We need only consider $\mathbb{F}_q$ where q is the appropriate prime power for the subfield, and n to correspond. For example, we can calculate the trace from $\mathbb{F}_{3^{16}}$ to $\mathbb{F}_{3^8}$ by taking $q = 3^8$ and $n = 2$.

2.2 Sequences over Finite Fields

2.2.1 Linear Recurrence

Consider a recurrence relation of the form $a_n = c_1a_{n-1} + c_2a_{n-2} + \cdots + c_na_0$, with an appropriate number of initial values. The familiar example is that of the Fibonacci sequence, where $a_n = a_{n-1} + a_{n-2}$ for $n \geq 2$, and usually we take $a_0 = a_1 = 1$. It is easy to use the recurrence relation to calculate as many elements of the sequence in turn as we desire: $a_2 = a_1 + a_0 = 1 + 1 = 2$, $a_3 = a_2 + a_1 = 2 + 1 = 3$, etc.

We can modify the recurrence relation to have the form $a_n - c_1a_{n-1} - c_2a_{n-2} - \cdots - c_na_0 = 0$, and recast it as a polynomial by substituting x^i for the a_i values as

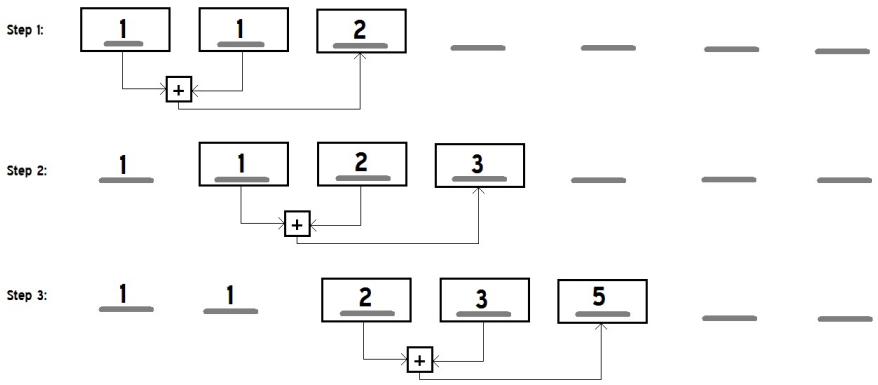
follows:

$$x^n - c_1x^{n-1} - c_2x^{n-2} - c_3x^{n-3} - \cdots - c_{n-1}x^1 - c_nx^0 = 0$$

Now the problem is one of finding the values that make this equation true; that is, finding the *roots* of the equation. In our Fibonacci example, we have the equation $x^2 - x - 1 = 0$.

A *Linear Feedback Shift Register* (LFSR) can be viewed as a simple machine that accepts a set of values as input and gives us the next element in the sequence as output. Each time it gives a new element, it moves a step over and begins the process again with the modified input. The recurrence relation and polynomial both encode the instructions for the machine, as seen in Figure 1.

Figure 1: *A machine taking two integers and outputting their sum, thereby creating the Fibonacci sequence.*



When the instructions in the machine are carried out over a finite field (see Figure 2), the sequence will inevitably start to repeat. This happens because there are only a finite number of possibilities for the numbers inside the machine at any given step. When the same set of input values occurs a second time, the machine will calculate the same output.

Our Fibonacci example over the integers never repeats, because there are an infinite number of possibilities for the two numbers it adds. If we calculate the same

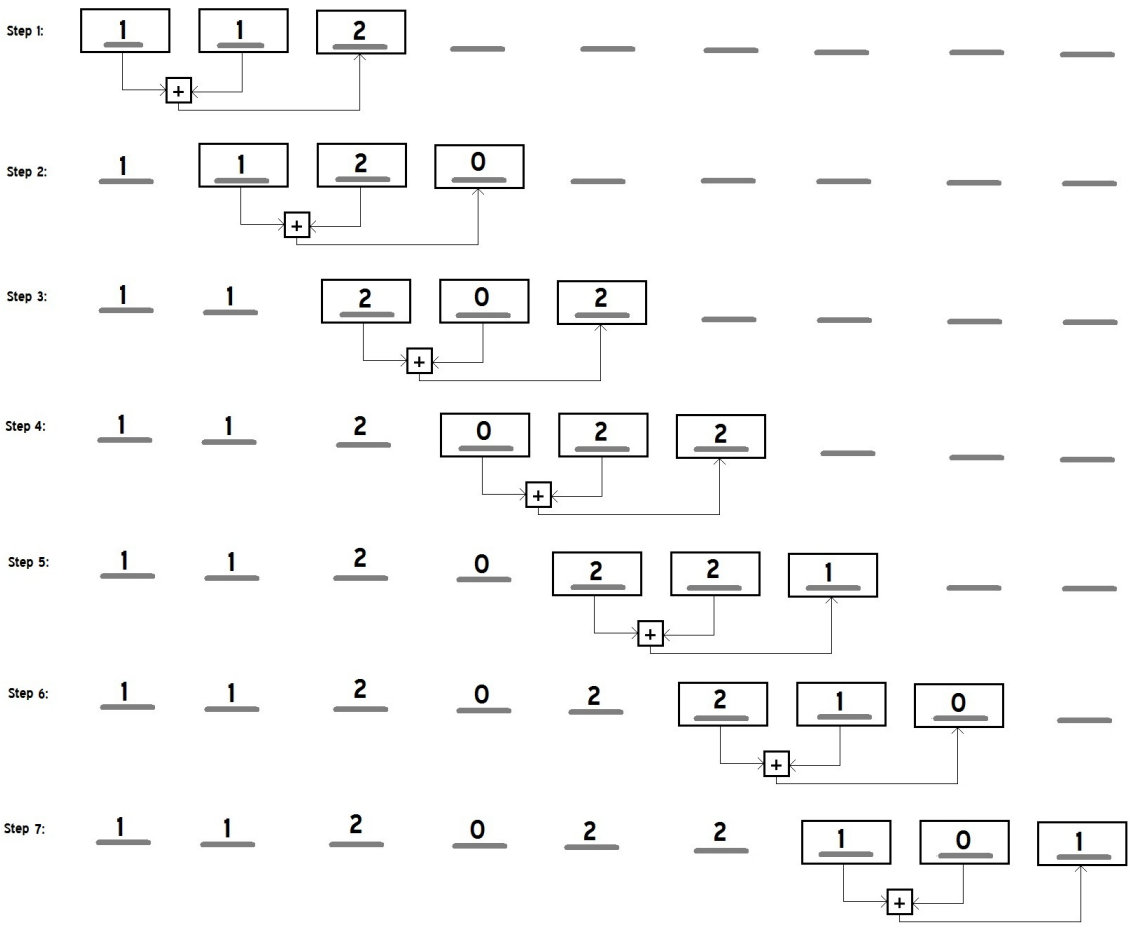
relation over $\mathbb{Z}_3$, though, this pattern emerges:

$$a_0 = 1, a_1 = 1, a_2 = 2, a_3 = 0, a_4 = 2, a_5 = 2, a_6 = 1, a_7 = 0, a_8 = 1, a_9 = 1, \ldots$$

Since $a_8 = a_0$ and $a_9 = a_1$, when the machine adds a_8 and a_9 it will obtain the same result as when it added a_0 and a_1. This means the sequence starts to repeat at this point. We will occasionally use this obvious compact notation for sequences:

$$1120221011202210\ldots.$$

Figure 2: *A modification of the above diagram, where addition is performed modulo 3.*



Since giving the machine input of all zeroes always results in an output of zero,

that starting condition will create a sequence consisting of only zeroes.

A *left shift* of j positions performed on a sequence is the removal of the first j elements and the movement of the others to the left to fill the gap. Formally, for a sequence $\underline{a} = (a_0, a_1, a_2, \ldots)$, the left shift operator L performs this operation:

$$L^{(j)}(a) = (a_j, a_{j+1}, a_{j+2}, \ldots)$$

2.2.2 Properties of sequences

For a finite sequence, we will refer to the *length* of the sequence as simply the number of elements.

We will denote a sequence of infinite length $a_0, a_1, \ldots$ as $\underline{a}$. If there exist integers $r > 0$ and $u \geq 0$ such that $a_{i+r} = a_i$ for all $i \geq u$, then the sequence is said to be *ultimately periodic*. The smallest such number r is called the *period* of the sequence.

We use a family of well-studied sequences as building blocks in our interleaved sequences.

Definition 2. A q-ary sequence generated by an n-stage LFSR is called a *maximal length sequence* (or m-sequence) if it has period $q^n - 1$.

Although any string of elements might be considered a sequence, there are properties that make some sequences more useful than others. The commonly considered properties are the *balance*, *run*, *tuple*, *auto-correlation*, and *coverage* properties. For our purposes, we consider the *balance*, *tuple*, and *coverage* properties.

The *balance* property says that the number of occurrences of each element in a period of the sequence should be as close to the same as possible. For example, the sequence $0010111\ldots$ has three zeroes and four ones, with a difference of 1, which is minimal. The sequence $00000011111222222\ldots$ also has the balance property, since in each period there are six zeroes, six twos, and five ones. In practice, we assume that

the elements with the lower occurrence are also labeled with lower numbers; since the properties do not depend on the actual elements, they can be re-labeled as necessary.

The *tuple* property stipulates that when we scan a sequence starting at each position of one period, for a fixed length t, we see every possible tuple of length t.

For example, under $\mathbb{Z}_2$, the sequence $00010111\cdots$ has the tuple property for length 3, because if we scan the sequence we find the tuples 000, 001, 010, 101, 011, 111, 110, and 100. We occasionally relax the restriction by not requiring the tuple consisting of t zeroes to occur.

The *coverage* property of a sequence generalizes the tuple property by stipulating that we must be able to take any t elements, rather than consecutive elements. The largest possible value of t for a sequence is called the *strength* of the coverage.

As an example, let us take the sequence $0010111\ldots$ of period 7. If we take three consecutive elements, we get the 3-tuples 001, 010, 101, 011, 111, 110, and 100. Since we have seven of the eight possible tuples, this sequence has the 3-tuple property. The zero tuple is a special case. Because our LFSR machinery outputs zero whenever the input values are all zeroes, the zero tuple cannot appear in the final sequence. Therefore we do not require it for the tuple property to hold.

To test the coverage property, let us try another set of three elements, for example, elements in positions 0, 1 and 3. We get the tuples 000, 011, 101, 011, 110, 110, 101. Since we have only the 3-tuples with an even number of ones, this sequence does not have the 3-coverage property.

2.3 Combinatorial Arrays

2.3.1 Orthogonal Arrays and Their Cousins

An *orthogonal array* of size N with k factors, s symbols, strength t and index $\lambda = \frac{N}{s^t}$, denoted by $\mathrm{OA}_\lambda(N, k, s, t)$, is an $N \times k$ array with $s \geq 2$ symbols having the property that in every $N \times t$ subarray, every t-tuple of symbols appears the same number λ

0	0	0
0	1	1
1	0	1
1	1	0

Table 1: *An orthogonal array OA(4,3,2,2).*

of times as a row. Unless otherwise specified, we will assume that $\lambda = 1$ and refer to the orthogonal array as simply an OA, without subscript.

Table 1 shows a very small orthogonal array. It is an $\mathrm{OA}(4, 3, 2, 2)$ with index $\lambda = 1$. Comparing each pair of columns, each 2-tuple in {00, 01, 10, 11} appears once. If we consider the problem of adding more columns, we soon run across the difficulty that we need two zeroes and two ones in each column, and there are $\binom{4}{2} = 6$ ways to do that. We have three columns already, and the other possibilities are all the reverse of an existing column. Since none of those can be added without destroying the 2-coverage of the orthogonal array, we know that it has the maximum number of columns.

As we can see from the definition, some of the parameters of an orthogonal array are dependent on others. Choosing the number of symbols, strength, and index, the size N of the array is fixed at $N = \lambda s^t$. This implies that only certain sizes of OA can be generated. There are also constraints on how many columns k can possibly be in the orthogonal array.

2.3.2 Covering Arrays

A *covering array* is a generalization of an orthogonal array. Instead of requiring that each tuple occurs *exactly* λ times, we require that it occur *at least* λ times and denote it CA_λ(N, k, s, t). Given values for the number of factors k, the number of symbols s, the index λ, and the strength t, the challenge is to find the smallest number of

0	0	0	0
0	1	1	1
1	0	1	1
1	1	0	1
1	1	1	0

Table 2: *The covering array CA(5,4,2,2).*

rows in the array that satisfies the condition. For example, if we take the orthogonal array above, we might want to add a fourth column to represent another factor in the testing. We know from the above that we cannot do this with four rows, but we need to determine how many rows we need to add. The solution in Table 2 shows that we can achieve the property we want with one more row, making a total of 5. This is the minimum number of rows for this set of parameters.

The determination of the required number of rows for a given set of parameters is an active area of research. Tables for the best known values for $2 \leq t \leq 6$, $2 \leq s \leq 25$, and $k < 10000$ are maintained at:

`http://www.public.asu.edu/~ccolbou/src/tabby/catable.html`.

3 Interleaved Sequences

3.1 Background

3.1.1 Definitions

An m-sequence of composite degree n has an associated *shift sequence* for each divisor of n. Let f be a primitive polynomial of degree n, m be an integer dividing n, and α a root of f. Hence, α is a primitive element of $\mathbb{F}_{q^n}$. Take $\gamma = \alpha^T$ a primitive element of $\mathbb{F}^*_{q^m}$, where $T = (q^n - 1)/(q^m - 1)$. For $k \in \mathbb{Z}_{q^n-1}$ define:

$$e_k = \begin{cases} \infty, & \text{if } \mathrm{Tr}^n_m(\alpha^k) = 0, \\ e, & \text{if } \mathrm{Tr}^n_m(\alpha^k) \neq 0 \text{ and } \mathrm{Tr}^n_m(\alpha^k) = \gamma^e. \end{cases}$$

Games [4] defines the sequence $\underline{e} = (e_0, e_1, \ldots, e_{q^n-2}) \in \mathbb{Z}^{q^n-1}_{q^m-1}$ to be the *shift sequence* associated with the polynomial f. As Games notes, the shift sequence is determined by its first T terms, where $T = (q^n - 1)/(q^m - 1)$. We define the shift sequence as the sequence $\underline{e} = (e_0, e_1, \ldots, e_{T-1})$, where $e_{i+T} = e_i + 1 \pmod{q^m - 1}$.

We use symbol T for the length of the shift sequence to acknowledge the standard in interleaved sequence literature, but not symbol t because of the association of t with the strength of an array in design theory [1], a concept we will use.

For example, let $f(x) = x^4 + x + 1$ over $\mathbb{F}_2$ be our primitive polynomial, which has degree $n = 4$. We choose $m = 2$ as our integer dividing n. We calculate $T = (2^4 - 1)/(2^2 - 1) = 5$.

Taking the root of f to be α, we calculate the trace function of the powers of α in this way, after we see that Tr(0)=0;

$$Tr(\alpha) = \alpha + \alpha^2 + \alpha^{2^2} + \alpha^{2^3}$$

$$= \alpha + \alpha^2 + \alpha^4 + \alpha^8$$

$$= \alpha + \alpha^2 + (\alpha + 1) + (\alpha^2 + 1)$$

$$= 0$$

$$Tr(\alpha^2) = \alpha^2 + (\alpha^2)^2 + (\alpha^2)^{2^2} + (\alpha^2)^{2^3}$$

$$= \alpha^2 + \alpha^4 + \alpha^8 + \alpha^{16}$$

$$= \alpha^2 + (\alpha + 1) + (\alpha^2 + 1) + (\alpha)$$

$$= 0$$

$$Tr(\alpha^3) = \alpha^3 + (\alpha^3)^2 + (\alpha^3)^{2^2} + (\alpha^3)^{2^3}$$

$$= \alpha^3 + \alpha^6 + \alpha^{12} + \alpha^{24}$$

$$= \alpha^3 + (\alpha^3 + \alpha^2) + (\alpha^3 + \alpha^2 + \alpha + 1) + (\alpha^3 + \alpha)$$

$$= 1$$

By continuing in this way, we find the sequence is 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, ..., which is periodic of length 15, as we expected.

We can then calculate the shift sequence using the definition. Our trace function is from $\mathbb{F}_{2^4}$ to $\mathbb{F}_{2^2}$, and $\gamma = \alpha^T = \alpha^5$. The values of k are from $\mathbb{Z}_{15}$.

$$Tr_2^4(\alpha^1) = \alpha + \alpha^4 = 1 = \gamma^0, \text{ so } e_0 = 0.$$

$$Tr_2^4(\alpha^2) = \alpha^2 + \alpha^8 = 1 = \gamma^0, \text{ so } e_1 = 0.$$

$$Tr_2^4(\alpha^3) = \alpha^3 + \alpha^{12} = \alpha^{10} = \gamma^2, \text{ so } e_2 = 2.$$

$$Tr_2^4(\alpha^4) = \alpha^4 + \alpha^{16} = 1 = \gamma^0, \text{ so } e_3 = 0.$$

The first $T = 5$ entries uniquely define the shift sequence, so we will end here and define the shift sequence to be $\underline{e} = [\infty, 0, 0, 2, 0]$. We use square brackets as a visual reminder that the sequence is finite and can contain elements of ∞, unlike the m-sequences.

There is an interesting visual representation of m-sequences that are decomposed in this way. If we write a single period of the sequence as a matrix with T columns and s rows, then the original sequence 0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, ... becomes:

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

The first column, where the values of the shift sequence are ∞, is filled with zeroes. The other columns are all versions of the sequence $0, 1, 1$, which is an m-sequence of period 3 associated with the polynomial $x^2 + x + 1$. Furthermore, the values of the shift sequence can be read as the number of left shifts to make to the sequence. In columns 1, 2, and 4, we see the sequence 0, 1, 1, with no shift. In column 3, however, we see the sequence 1, 0, 1, which is $L^2(0, 1, 1, \dots)$.

This decomposition of m-sequences suggests a method of constructing new sequences, which may or not be m-sequences.

A sequence $\underline{u}$ is an *interleaved sequence* if it is constructed from a base sequence $\underline{a}$ of period s and a shift sequence $\underline{e}$ of length T consisting of elements of $\mathbb{Z}_s \cup \{\infty\}$. We construct the interleaved sequence by creating the matrix U in this manner:

$$U = \begin{bmatrix} a_{e_0} & a_{e_1} & a_{e_2} & \cdots & a_{e_{T-1}} \\ a_{e_0+1} & a_{e_1+1} & a_{e_2+1} & \cdots & a_{e_{T-1}+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{e_0+s-1} & a_{e_1+s-1} & a_{e_2+s-1} & \cdots & a_{e_{T-1}+s-1} \end{bmatrix}$$

$$= \begin{bmatrix} u_0 & u_1 & u_2 & \dots & u_{T-1} \\ u_T & u_{T+1} & u_{T+2} & \dots & u_{2T-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_{(s-1)T} & u_{(s-1)T+1} & u_{(s-1)T+2} & \dots & u_{sT-1} \end{bmatrix}$$

To indicate that $\underline{u}$ is the interleaved sequence formed from the base sequence $\underline{a}$ and shift sequence $\underline{e}$, we use the notation $\underline{u} = IL(\underline{a}, \underline{e})$.

The matrix representation shows that we can express u_i in terms of elements of $\underline{a}$ and $\underline{e}$. For any $0 \leq i \leq sT - 1$, with $i = wT + r$, we can express :

$$u_i = \begin{cases} a_{e_r+w} & e_r \neq \infty, \\ 0, & e_r = \infty. \end{cases}$$

A useful consequence of the T columns of the matrix is that moving T positions along the matrix is equivalent to moving down one row. Therefore we will often use the fact that :

$$a_{e_{r+T}+w} = a_{e_r+w+1}.$$

We always assume that $\underline{a}$ has period $s > 1$.

If the degree of polynomial of the m-sequence has multiple divisors, then the m-sequence $\underline{u}$ can be decomposed in different ways. For example, the m-sequence $\underline{u}$ = (0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, ...) has period 63, because the associated polynomial $x^6 + x + 1$ has

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

Table 3: *Interleaved sequence $\underline{u}$ as a 7 by 9 matrix.*

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

Figure 3: *Interleaved sequence $\underline{u}$ as a 3 by 21 matrix.*

degree 6. We can take either 2 or 3 as our divisor of 6. If we take $m_1 = 3$, we have $T = (2^6 - 1)/(2^3 - 1) = 9$ and we write the sequence in a 9 by 7 matrix. In this form we can see that the base sequence is $\underline{a} = (0, 1, 0, 0, 1, 1, 1, \ldots)$ and the shift sequence $\underline{e}_1 = [\infty, 0, 3, 2, 2, 6, 0, 2, 0]$. In the case where $m_2 = 2$, we write a 21 by 3 matrix. Now $\underline{a}_2 = (0, 1, 1, \ldots)$ and $\underline{e}_2 = [0, 0, 0, \infty, 0, 2, \infty, \infty, 0, 0, 1, 2, \infty, 0, \infty, 2, 0, 2, 0, 0, 2]$. The 3×21 matrix is shown in Figure 3.

It is well known that any m-sequence has such a representation, unless the period of the m-sequence is a prime. See Golomb and Gong [5, Theorem 5.2].

Theorem 1. *Let n be a composite number, m be a proper factor of n and $d = (q^n - 1)/(q^m - 1)$. Then any m-sequence over $\mathbb{F}_q$ of degree n can be arranged into a $(q^m - 1) \times d$ array where each column sequence is either an m-sequence over $\mathbb{F}_q$ of degree m or a zero sequence for which all the m-sequences are shift equivalent.*

Games [4, Theorem 2] proves a structural property about $\underline{e}$.

Theorem 2. *Let $\underline{e} = (e_0, e_1, \ldots, e_{q^n-2})$ be the shift sequence associated with the primitive polynomial f. For fixed $k \in \mathbb{Z}_{q^n-1}, k \not\equiv 0 \pmod{T}$, the list of differences $(e_{j+k} - e_j \pmod{q^m - 1} : j \in \mathbb{Z}_T)$ contains each element of $\mathbb{Z}_{q^m-1}$ exactly q^{n-qm} times.*

value of k	list of differences
1	3, 6, 0, 4, 1, 2, 5
2	2, 6, 4, 5, 3, 0, 1
3	2, 3, 5, 0, 1, 6, 4
4	6, 4, 0, 5, 1, 2, 3
5	0, 6, 5, 2, 4, 1, 3
6	2, 4, 6, 5, 4, 1, 3
7	0, 6, 2, 4, 3, 5, 1
8	5, 2, 1, 4, 0, 6, 3

Table 4: *Differences for values of k.*

In our example with $m = 3$ and $\underline{e} = [\infty, 0, 3, 2, 2, 6, 0, 2, 0]$, the sequences of differences for different values of k is given in Table 4. We have that as expected, each of the values from 0 to 6 occurs exactly once in each set of differences. Although $\underline{e}$ has 9 positions, we ignore the differences where a shift of ∞ is involved, so the number of differences each time is 7.

This balance property of the differences suggests a method we can use to create shift sequences, which we explore further in Section 4.

3.2 Equivalence Classes of IL Sequences

Since the shift sequence contains elements of $\mathbb{Z}_s \cup \infty$, and the shift sequence is of length T, for a given length T there are $(s+1)^T$ possible shift sequences. This creates a large pool of potential shift sequences.

We define two shift sequences to be equivalent if the resulting interleaved sequence can be obtained from one another by cyclic shifts and/or reverses. Shifted sequences have the same t-tuple coverage and runs. Reversed sequences have the same level of coverage, although the tuples themselves are reversed. If we fully understand these equivalent forms, our computation experiments can be more efficient.

3.2.1 Rotations

Let $\underline{a}$ be any sequence over $\mathbb{F}_q$. Define $L(a)$ to be the sequence $\underline{a}'$ where $a'_i = a_{i+1}$, for all i. If $\underline{u} = IL(\underline{a}, \underline{e})$, we are interested in how $L(\underline{u})$ decomposes as an interleaved sequence. To this end, we define an operation on shift sequences, which we call *rotation*.

Definition 3. For a shift sequence $(e_0, e_1, \ldots, e_{s-1})$ over $\mathbb{Z}_s$, we define a *rotation* to the left as follows:

$$\mathbb{L}(e_0, e_1, \ldots, e_{T-1}) = (e_1, e_2, \ldots, e_{T-1}, e_0 + 1).$$

The addition is performed in $\mathbb{Z}_s$. When element ∞ is included, we use the convention that $\infty + z = \infty$ for all z in $\mathbb{Z}_s$.

Theorem 3. *Let $\underline{a} = (a_0, a_1, \ldots, a_{s-1})$ be a sequence of period s, $\underline{e} = (e_0, e_1, \ldots, e_{T-1})$ be a shift sequence of length T, and let $\underline{u} = (u_0, u_1, \ldots, u_{sT-1}) = IL(\underline{a}, \underline{e})$. If $\underline{e}' = \mathbb{L}(\underline{e})$ and $\underline{u}' = IL(\underline{a}, \underline{e}')$, then*

$$\underline{u}' = L^{(1)}(\underline{u}).$$

Proof. For any i, $0 \leq i \leq sT - 1$, let $i = wT + r$ where $0 \leq r < T$. Then,

$$u_i = a_{w+e_r}.$$

Similarly, in the shifted sequence, $u'_i = a_{w+e'_r}$. By the definition of rotation, $e'_j = e_{j+1}$ for $0 \leq j < T - 1$, and $e'_{T-1} = e_0 + 1$ otherwise. In the first case,

$$\begin{aligned} u'_i &= a_{w+e'_r} \\ &= a_{w+e_{r+1}} \\ &= u_{i+1}. \end{aligned}$$

In the second case $r = T-1$, so the entry is at the right side of the interleaved matrix. Then,

$$\begin{aligned} u'_i &= a_{w+e'_r} \\ &= a_{w+e_0+1} \\ &= a_{w+1+e_0} \\ &= u_{i+1}. \end{aligned}$$

We have that $u'_i = u_{i+1}$ in both cases, so $\underline{u}' = L(\underline{u})$ as desired. □

In this way we have that rotations of a shift sequence (which are rotations to the left, by definition) correspond to left-shifts on the associated interleaved sequence. We denote this relationship by $L^{(i)}(\underline{u}) \Leftrightarrow \mathbb{L}^{(i)}(\underline{e})$.

3.2.2 Reverses

The other transformation to consider is the reversal of an interleaved sequence.

Definition 4. For any sequence $\underline{a} = (a_0, a_1, \ldots, a_{s-2}, a_{s-1})$, we define a *reverse* operator as follows:

$$\mathrm{Rev}(a_0, a_1, \ldots, a_{s-2}, a_{s-1}) = (a_{s-1}, a_{s-2}, \ldots, a_1, a_0).$$

Two sequences $\underline{a}$ and $\underline{b}$, both of length s, are *reverses* of each other if $a_i = b_{s-1-i}$ for all $0 \leq i \leq s-1$.

Theorem 4. *Let $\underline{a} = (a_0, a_1, \ldots, a_{s-1})$ be a base sequence of period s. Let $\underline{e} = (e_0, e_1, \ldots, e_{T-1})$ be a shift sequence of length T, and $\underline{u} = IL(\underline{a}, \underline{e})$. If $\underline{a}' = Rev(\underline{a})$,*

$\underline{e}' = -(Rev(\underline{e}))$, *and* $\underline{u}' = IL(\underline{a}', \underline{e}')$, *then*

$$\underline{u}' = Rev(\underline{u}).$$

Proof. We want to show that for any u_i in $\underline{u}$, $u_i = u'_{sT-1-i}$.

By the definition of reverse,

$$a'_i = a_{s-1-i}, \quad e'_j = -e_{T-1-j}.$$

Separate i into a multiple of T and a remainder, by letting $i = wT + r$, $0 \leq r < T$, for some w. Then $sT - 1 - i = sT - 1 - wT - r = (s - w - 1)T + (T - r - 1)$.

$$\begin{aligned}
u'_{sT-1-i} &= u'_{(s-w-1)T+(T-r-1)} \\
&= a'_{e'_{(T-r-1)}+(s-w-1)} \\
&= a'_{-e_{T-1-(T-r-1)}+(s-w-1)} \\
&= a'_{-e_r+(s-w-1)} \\
&= a_{s-1-(-e_r+(s-w-1))} \\
&= a_{e_r+w} \\
&= u_i.
\end{aligned}$$

□

For example, take the base sequence $\underline{a} = (0, 1, 2)$ in $\mathbb{Z}_3$ and the shift sequence $\underline{e} = [\infty, 0, 1, 2, 1]$. We create $\underline{a}' = (2, 1, 0)$ and $\underline{e}' = -[1, 2, 1, 0, \infty] = [2, 1, 2, 0, \infty]$. Then,

$$\underline{u} = \begin{bmatrix} 0 & 0 & 1 & 2 & 1 \\ 0 & 1 & 2 & 0 & 2 \\ 0 & 2 & 0 & 1 & 0 \end{bmatrix}, \quad \text{and} \quad \underline{u}' = \begin{bmatrix} 0 & 1 & 0 & 2 & 0 \\ 2 & 0 & 2 & 1 & 0 \\ 1 & 2 & 1 & 0 & 0 \end{bmatrix}.$$

They are reverses of each other.

Now that we have defined a set of equivalence classes of our shift sequences, we would like to choose a canonical representative from each class to represent the entire class. Which shift sequence we choose from each equivalence class is largely a matter of utility and taste. We choose as the representative shift sequence the one that is lexicographically least, using the convention that ∞ is an arbitrary small number. It is desirable to have a property that can be used to determine whether a particular shift sequence is canonical by examining it in isolation, without having to calculate all the equivalent shift sequences given by the possible rotations.

For example, for the shift sequence $[\infty, 0, 3, 2, 2, 6, 0, 2, 0]$ we considered earlier, there are 63 cyclic shifts. We list the first few:

$$[\infty, 0, 3, 2, 2, 6, 0, 2, 0], [0, 3, 2, 2, 6, 0, 2, 0, \infty], [3, 2, 2, 6, 0, 2, 0, \infty, 1],$$
$$[2, 2, 6, 0, 2, 0, \infty, 1, 4], [2, 6, 0, 2, 0, \infty, 1, 4, 3], [6, 0, 2, 0, \infty, 1, 4, 3, 3],$$
$$[0, 2, 0, \infty, 1, 4, 3, 3, 0], [2, 0, \infty, 1, 4, 3, 3, 0, 1], [0, \infty, 1, 4, 3, 3, 0, 1, 3],$$
$$[\infty, 1, 4, 3, 3, 0, 1, 3, 1], [1, 4, 3, 3, 0, 1, 3, 1, \infty], [4, 3, 3, 0, 1, 3, 1, \infty, 2], \ldots$$

Instead of checking each one, we examine the differences between consecutive entries, using the convention that $x - \infty = -\infty$ for all x. If the sequence of differences is lexicographically least in its cyclic shift equivalence class, then $\underline{e}$ is canonical. Our $\underline{e} = [\infty, 0, 3, 2, 2, 6, 0, 2, 0]$ has differences $-\infty, 3, 6, 0, 4, 1, 2, 5, \infty$. They are in lexicographically least order, so this shift sequence is indeed the representative shift sequence for its class.

We conclude this section with a note about the number and placement of ∞ shifts. For two shift sequences to be equivalent, obviously the number of ∞'s must be the same, as well as their relative placement. There are normally $(s+1)^T$ possible shift

sequences. Expanding this, we obtain:

$$s^T + \binom{T}{1} s^{T-1} + \binom{T}{2} s^{T-2} + \cdots + \binom{T}{T-1} s + 1.$$

Each term represents a number of ∞'s in the shift sequences, with the binomial coefficients giving the number of possible placements of them. The first term, s^T, gives the number of shift sequences with no ∞'s. The second gives the number of shift sequences with one ∞. There are $\binom{T}{1} = T$ choices of positions where the ∞ can be, and s^{T-1} ways to arrange the s other symbols in the remaining $T-1$ spots.

3.3 Properties of Interleaved Sequences

3.3.1 Balance

One of Golomb's three randomness postulates (1955) requires that the number of zeroes should be nearly equal to the number of ones in binary sequences. In general, for non-binary sequences, the *balance* property is the following.

Definition 5. Let N_v be the number of $v \in \mathbb{F}_q$ that occur in one period of $\underline{a}$; that is, $N_v = |\{j : a_j = v, 0 \leq j < N\}|$, where N is the period of $\underline{a}$. For each $v \neq \beta$, for a balanced sequence, $|N_v - N_\beta| \leq 1$. In particular, for $N = q^n - 1$, in every period, every nonzero element in $\mathbb{F}_q$ occurs q^{n-1} times, and the zero element occurs $q^{n-1} - 1$ times.

If the element that occurs least is not zero in the presentation of the sequence, the symbols can be re-labeled. Therefore when a sequence satisfies the balance property, we assume the symbols that occur fewer times are the smallest symbols.

Golomb and Gong [5, p. 138] give the number of zero columns (that is, the number of ∞ in $\underline{e}$) when an m-sequence is decomposed. If n is a composite number, let m be a proper factor of n and $d = \frac{q^n-1}{q^m-1}$. Then, any m-sequence over $\mathbb{F}_q$ of degree n can be

arranged into a $(q^m - 1) \times d$ array where each column is either an m-sequence over $\mathbb{F}_q$ of degree m or a zero sequence. There are

$$\frac{q^{n-m} - 1}{q^m - 1}$$

zero columns in the array.

Here we generalize the calculation to any shift sequence. This allows us to create interleaved sequences of various lengths, not constrained by the factors of n.

Theorem 5. *Given an m-sequence $\underline{a}$ of period s over $\mathbb{F}_q$, an interleaved sequence $\underline{u}$ created from $\underline{a}$ satisfies the balance property if and only if the number of ∞ entries in the shift sequence $\underline{e}$ of period T is:*

$$\frac{T-1}{s+1}.$$

Proof. Since m-sequences are balanced, we can assume that $\underline{a}$ of period s over $\mathbb{F}_q$ contains $(s-(q-1))/q$ zeroes and $(s+1)/q$ instances of each other symbol.

We consider the matrix U that defines $\underline{u}$. Each column formed from a shift of $\underline{a}$ contributes $(s-(q-1))/q$ zeroes to the final sequence, and each column corresponding to an ∞ in $\underline{e}$ contributes a full column of zeroes, that is, s zeroes. If the number of columns of zeroes is i, then the total number of zeroes in $\underline{u}$ is:

$$N_0(\underline{u}) = (T-i)\frac{s-(q-1)}{q} + is.$$

Non-zero symbols come only from the columns that are created from $\underline{a}$. For every non-zero element v in $\mathbb{F}_q$,

$$N_v(\underline{u}) = (T-i)\frac{s+1}{q}.$$

We have assumed that $N_v \geq N_0$, and in an m-sequence, we know that $N_v - N_0 = 1$.

Since our base sequence is an m-sequence, we expect to find the same value.

$$(T-i)\frac{s+1}{q} - ((T-i)\frac{s-(q-1)}{q} + is) = 1$$

$$(T-i)(s+1) - (T-i)(s-q+1) - qis = q$$

$$Ts + T - is - i - Ts + Tq - T + is - iq + i - qis = q$$

$$Tq - iq - qis = q$$

$$qis + iq = Tq - q$$

$$i(s+1) = T - 1$$

$$i = \frac{T-1}{s+1}.$$

Therefore we have that the number of ∞'s required to maintain the balance property in the final interleaved sequence must be $(T-1)/(s+1)$, and furthermore, this necessary condition is also sufficient. □

If we begin the calculation by assuming that we want the number of zeroes and other symbols to be exactly equal, then we can show that i needs to be precisely $T/(s+1)$. This puts us in the unusual position of having better balance in the interleaved sequence than in the base sequence.

Throughout this thesis, unless otherwise specified, the number of ∞'s in a shift sequence is assumed to be $i = \frac{T-1}{s+1}$, rounded to the nearest integer.

Because we are dividing by $s+1$, there are clearly s remainders other than 0. Let i' be the value of i rounded to the nearest integer. Then,

$$|i' - i| \leq \frac{1}{2}.$$

The balance is off by the largest amount when $|i' - i| = 1/2$.

3.3.2 Tuple

We begin by showing a small requirement that preserves the tuple property of the base sequenced. It relies on a shift sequence with consecutive entries that are also consecutive numbers, for example, a shift sequence like $[\infty, 0, 3, 4, 5]$.

Theorem 6. *Let $\underline{a}$ be an m-sequence over $\mathbb{F}_q$ of period $q^n - 1$, $\underline{e}$ be a shift sequence of length T, and t be a positive integer. If $\underline{e}$ contains t consecutive entries $(e_j, e_{j+1}, \ldots, e_{j+t-1})$ with the form $(e_j, e_j + 1, \ldots, e_j + t - 1)$, then the interleaved sequence created from $\underline{a}$ and $\underline{e}$ has the t-tuple property.*

Proof. Since $\underline{a}$ has the t-tuple property, in every period of $\underline{a}$, each non-zero t-tuple $(v_1, v_2, \ldots, v_t)$ occurs exactly once. Consider any non-zero t-tuple $(v_1, v_2, \ldots, v_t)$, occurring in $\underline{a}$ in positions $(a_i, a_{i+1}, \ldots, a_{i+t-1})$. Let the t consecutive entries of $\underline{e}$ be $(e_j, e_{j+1}, \ldots, e_{j+t-1})$, which means we can re-write them as $(e_j, e_j + 1, \ldots, e_j + t - 1)$. Now we consider the specific values of $\underline{a}$ from our t-tuple that occur in column j. The values are $(a_{e_{j+i}}, a_{e_{j+i+1}}, \ldots, a_{e_{j+i+t-1}})$. We substitute our values of $\underline{e}$ to obtain $(a_{e_j+i}, a_{e_j+i+1}, \ldots, a_{e_j+i+t})$. Expressing these in terms of the entries of $\underline{u}$, we have $(u_{iT+j}, u_{(i+1)T+j}, \ldots, u_{(i+t-1)T+j})$. Since these are consecutive values in $\underline{u}$, the desired t-tuple appears in $\underline{u}$. We can complete this process for every t-tuple, so the interleaved sequence $\underline{u}$ has the t-tuple property. □

Example 1. The base sequence $\underline{a} = (0, 1, 1)$ has the 2-tuple property (without the 00 tuple). The shift sequence $\underline{e}_1 = [0, 1]$ has two consecutive entries, and the interleaved sequence $\underline{u}_1 = (0, 1, 1, 1, 1, 0)$ also has the 2-tuple property. Although the tuple 11 appears three times, we know that it is guaranteed by the values u_2 and u_3, which correspond to a_{e_0+1} and a_{e_1+1}.

A question is whether an m-sequence $\underline{a}$ with the tuple property for some t can be used to create an interleaved sequence with the tuple property for a larger t. Since we

know that every interleaved sequence can be decomposed into smaller m-sequences, we know it is true for those cases.

Shift sequences with t consecutive entries are not the only shift sequences that preserve the t-tuple property. For example, consider $\underline{a} = (0, 1, 0, 0, 1, 1, 1, \dots)$. With the shift sequence $\underline{e}_1 = [0, 2]$, we have $\underline{u}_1 = (0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 1, 1, \dots)$, which has all seven non-zero 3-tuples. But not all shift sequences of this length do. With the shift sequence $\underline{e}_2 = [0, 1]$, we have $\underline{u}_2 = (0, 1, 1, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 0, \dots)$, which is missing the tuples (0, 1, 0) and (1, 0, 1). More investigation is needed to completely characterize which shift sequences preserve the t-tuple property, and which do not.

3.3.3 Orbit Size

Given our operations $\mathbb{L}$ and Rev, how large is the orbit of $\underline{u}$ under these transformations? $\mathbb{L}$ and Rev generate the dihedral group of order $2|\underline{u}| = 2sT$. The maximum size of the orbit, $2sT$, is the size of the acting group.

If $Rev(\underline{u}) = \mathbb{L}^i(u)$ for some i, then the orbit does not have size greater than sT.

There are cases where the orbit has size less than sT. For example, consider the small example of a palindromic base sequence of period $s = 3$ and a shift sequence of length $T = 5$.

Some orbit sizes are not possible. For example, consider an interleaved sequence of period $d \mid T$, say $\underline{u} = (u_0, u_1, \dots, u_{d-1})$. Because d divides T, we have a matrix form that looks like this:

$$\begin{bmatrix} u_0 & u_1 & \dots u_{d-1} & u_0 & u_1 & \dots u_{d-1} & \dots & u_0 & u_1 & \dots u_{d-1} \\ u_0 & u_1 & \dots u_{d-1} & u_0 & u_1 & \dots u_{d-1} & \dots & u_0 & u_1 & \dots u_{d-1} \\ u_0 & u_1 & \dots u_{d-1} & u_0 & u_1 & \dots u_{d-1} & \dots & u_0 & u_1 & \dots u_{d-1} \end{bmatrix},$$

and thus, either $\underline{a}$ has period 1 or $e_i = \infty$ for all i.

A similar situation occurs when $d = kT$ for $k \mid s$. In this case the matrix form of

$\underline{u}$ looks like this:

$$\begin{bmatrix} u_0 & u_1 & \dots & u_{T-1} \\ u_T & u_{T+1} & \dots & u_{2T-1} \\ \vdots & & & \\ u_{(k-1)T} & u_{(k-1)T+1} & \dots & u_{kT-1} \\ \vdots & & & \\ u_0 & u_1 & \dots & u_{T-1} \\ u_T & u_{T+1} & \dots & u_{2T-1} \\ \vdots & & & \\ u_{(k-1)T} & u_{(k-1)T+1} & \dots & u_{kT-1} \end{bmatrix}.$$

Each column of the matrix has period $k < s$, which contradicts the assumption that the base sequence $\underline{a}$ has period s.

Next we show that the period of an interleaved sequence must be a multiple of the period of the base sequence.

Lemma 1. *Let $\underline{u}$ be an interleaved sequence created from a base sequence $\underline{a}$ of period s and a shift sequence $\underline{e}$ of length T. If $u_{i+l} = u_i$ for all positive integers i, then $l = ks$ for some integer k.*

Proof. Let $d = \gcd(l, sT)$. Then there exists $x, y \in \mathbb{Z}$ such that $xl + ysT = d$ and $u_{i+d} = u_{i+xl+ysT} = u_{i+ysT} = u_i$. Let $d = wT + r$. Taking every dth element of $\underline{u}$ we must have:

$$\begin{aligned} a_{e_0} &= a_{w+e_r} \\ a_{e_1} &= a_{w+e_{r+1}} \\ &\vdots \\ a_{e_{T-r-1}} &= a_{w+e_{T-1}} \end{aligned}$$

$$a_{e_{T-r}} = a_{w+1+e_0}$$
$$\vdots$$
$$a_{e_{T-1}} = a_{w+1+e_{r-1}}.$$

This leads to a system of equations:

$$\begin{aligned}
e_0 &\equiv w + e_r \pmod{s} \\
e_1 &\equiv w + e_{r+1} \pmod{s} \\
e_2 &\equiv w + e_{r+2} \pmod{s} \\
&\vdots \\
e_{T-r-1} &\equiv w + e_{T-1} \pmod{s} \\
e_{T-r} &\equiv w + 1 + e_0 \pmod{s} \\
&\vdots \\
e_{T-1} &\equiv w + 1 + e_{r-1} \pmod{s}.
\end{aligned}$$

Summing these equations, we observe that the terms e_0 through e_{T-1} occur on both sides. Therefore, we have the simplified equation

$$0 \equiv (T-r)w + (w+1)r \pmod{s}$$

$$Tw + r \equiv 0 \pmod{s}$$

$$d \equiv 0 \pmod{s}.$$

We conclude that d and also l are multiples of s. □

Theorem 7. *Let $\underline{a}$ be a base sequence of period s, $\underline{e}$ be a shift sequence of length T, and create the interleaved sequence $\underline{u} = IL(\underline{a}, \underline{e})$. Let l be an integer that is the desired length of the period of the interleaved sequence, and calculate $n = sT/l$. If $u_i = u_{l+i}$ for all integers i, then $n|T$, gcd$(n, s) = 1$, and $e_i = w + \epsilon_i + e_{r+i}$, $0 \leq i < T$, where the addition in the subscript is performed modulo T and:*

$$\epsilon_i = \begin{cases} 0, & 0 \leq i < T - r, \\ 1, & T - r \leq i < T. \end{cases}$$

Proof. First, we have $\frac{sT}{n} = l$, which is equivalent to $\frac{T}{n} = \frac{l}{s}$. From Lemma 1, l must be a multiple of s. Therefore, n divides T.

Consider the least common multiple of l and T. Because l divides sT, $\operatorname{lcm}(l, T) \leq sT$. By way of contradiction, assume $\operatorname{lcm}(l, T) < sT$, and suppose $e_i \neq \infty$. We have $u_i = u_{i+\operatorname{lcm}(l,T)}$ and both these entries are in the same column of the matrix form of $\underline{u}$. Then $a_j = a_{j+\operatorname{lcm}(l,T)/T}$. Thus $\operatorname{lcm}(l, T)/T = s$ and $\operatorname{lcm}(l, T) = sT$ by contradiction.

Let the $\gcd(n, s) = d$. Next, we show d=1. Indeed, let $n = n'd$ and $s = s'd$, so that $\gcd(n', s') = 1$. We have

$$\begin{aligned} lT &= \gcd(l, T) \operatorname{lcm}(l, T), \quad \text{and so} \\ \gcd(l, T) &= \frac{lT}{sT} = \frac{l}{s} = \frac{T}{n}. \\ \text{Hence } n \gcd(l, T) &= T. \end{aligned}$$

We substitute $n = n'd$. Because T/n' is an integer dividing both l and T, and the gcd of two numbers must be at least as large as any divisor, we obtain

$$\gcd(l, T) \geq \frac{T}{n'}.$$

Now

$$\begin{aligned} T &= n \gcd(l, T) \geq \frac{nT}{n'} \\ &= \frac{n' dT}{n'} = dT. \end{aligned}$$

We have shown that $T \geq dT$. Since d is a positive integer, $d = 1$ and the $\gcd(n, s) = 1$ as claimed.

Lemma 1 establishes the existence of the equations. □

Theorem 7 gives us conditions to create interleaved sequences of maximum period. The equations from Lemma 1 give us a method to create the associated shift sequences. These equations sometimes create relationships between the entries in the shift sequence, reducing the number of shift sequences. Depending on the application, we may want a large or small number of shift sequences to generate and test. For this reason it is useful to note that the number of independent values of e_i is $\gcd(l, T)$.

Example 2. If we have a base sequence of period $s = 3$ and a shift sequence of length $T = 4$, we have $sT = 12$. Since n must divide T, we have the possibilities 1, 2, and 4. All of these have a greatest common divisor with $s = 3$ of 1. The corresponding values of l are 12, 6, and 3. We should find that these values partition the possible $4^4 = 256$ shift sequences into orbits of these lengths.

Consider $l = 3$ first. We have $n = \frac{sT}{l} = 4$, and $4|T$ and $\gcd(4, 3) = 1$ as expected. We express l in the form $l = mT + r$, that is, $3 = 0(4) + 3$, which gives our system of equations:

$$e_0 \equiv e_3 + 0 \pmod 3$$

$$e_1 \equiv e_0 + 1 \pmod 3$$

$$e_2 \equiv e_1 + 1 \pmod 3$$

$$e_3 \equiv e_2 + 1 \pmod 3.$$

Since the $\gcd(l, T) = \gcd(3, 4) = 1$, we have only one free variable, e_0. Using the equalities to set up the form of the shift sequence, we have $[e_0, e_0 + 1, e_0 + 2, e_0]$. Setting e_0 to the possible four values, we obtain these shift sequences:

$$[\infty, \infty, \infty, \infty], [0, 1, 2, 0], [1, 2, 0, 1], [2, 0, 1, 2].$$

The first we have seen already. With the base sequence (a_0, a_1, a_2), the second forms the interleaved sequence $\underline{u} = (a_0, a_1, a_2, a_0, a_1, a_2, \ldots)$ of period 3. The third and fourth shift sequences give $\mathbb{L}(\underline{u})$ and $\mathbb{L}^{(2)}(\underline{u})$.

In the second case, where $l = 6$, $n = \frac{sT}{l} = 2$, and $2|T$ and $\gcd(2, 3) = 1$ as expected. We set $l = 6$ where $l = 1(T) + 2$, and set up the system of equations:

$$e_0 \equiv e_2 + 1 \pmod 3$$

$$e_1 \equiv e_3 + 1 \pmod 3$$

$$e_2 \equiv e_0 + 2 \pmod 3$$

$$e_3 \equiv e_1 + 2 \pmod 3.$$

Since the $\gcd(l, T) = 2$, the first two equalities give us all the information we need to determine the shift sequence, which takes the form $[e_0, e_1, e_0 + 2, e_1 + 2]$. The values for e_0 and e_1 can be filled with any two values from $\infty \cup \mathbb{Z}_s$. If we use the pair (∞, ∞), we obtain the trivial case, and if we use any of the pairs (0,1), (1,2), or (2,0), we obtain the three sequences found above in the orbit of length 3. If we take

each remaining pair and generate its orbit, we find these:

$$[\infty, 0, \infty, 2], [0, \infty, 2, \infty], [\infty, 2, \infty, 1], [2, \infty, 1, \infty], [\infty, 1, \infty, 0], [1, \infty, 0, \infty];$$
$$[0, 0, 2, 2], [0, 2, 2, 1], [2, 2, 1, 1], [2, 1, 1, 0], [1, 1, 0, 0], [1, 0, 0, 2].$$

Considering $e_1 = [\infty, 0, \infty, 2]$ and $e_2 = [0, 0, 2, 2]$, we obtain these interleaved sequences of period 6:

$$\underline{u}_1 = (0, a_0, 0, a_2, 0, a_1, 0, a_0, 0, a_2, 0, a_1, \ldots),$$
$$\underline{u}_2 = (a_0, a_0, a_2, a_2, a_1, a_1, a_0, a_0, a_2, a_2, a_1, a_1, \ldots).$$

In the final case, $l = 12$, and since $12 = 3(4) + 0$, our system of equations is, for all i,

$$e_i \equiv e_i + 3 \pmod 3.$$

Every position of the shift sequence is a free variable. Removing the sets of values we have already seen in the previous lengths, we have 240 remaining shift sequences. They are partioned into orbits of length 12, and so we expect there are 20 of them. The representatives for each class are

$$[\infty, \infty, \infty, 0], [\infty, \infty, 0, 0], [\infty, \infty, 0, 1], [\infty, \infty, 0, 2], [\infty, 0, \infty, 0],$$
$$[\infty, 0, 0, 0], [\infty, 0, 0, 1], [\infty, 0, 0, 2], [\infty, 0, 1, 0], [\infty, 0, 1, 1],$$
$$[\infty, 0, 1, 2], [\infty, 0, 2, 0], [\infty, 0, 2, 1], [\infty, 0, 2, 2], [0, 0, 0, 0],$$
$$[0, 0, 0, 2], [0, 0, 1, 0], [0, 0, 1, 2], [0, 0, 2, 0], [0, 1, 0, 2].$$

It is occasionally convenient to ignore the shifts of ∞. For example, if we have the shift sequence $[\infty, 0, 0, \infty, 0]$ over a base sequence of length 3, all of these sequences

produce shift-equivalent interleaved sequences:

$$\infty00\infty0, 00\infty0\infty, 0\infty0\infty1, \infty0\infty11, 0\infty11\infty, \infty11\infty1, 11\infty1\infty, 1\infty1\infty2,$$
$$\infty1\infty22, 1\infty22\infty, \infty22\infty2, 22\infty2\infty, 2\infty2\infty0, \infty2\infty00, 2\infty00\infty.$$

It is easier to consider the shifts of length 3, without the ∞ symbol, of which there are 27. They are in orbits of size 9 with representatives 000, 002, and 010. We can use this to generate the three representatives of the original sequence, which are $\infty00\infty0, \infty00\infty2$, and $\infty01\infty0$. If instead we want the two infinity shifts to be adjacent, we adjust the representatives to be $\infty\infty000, \infty\infty002$, and $\infty\infty010$. From this we observe that if the pattern and number of ∞ shifts is given, we can simply generate the representatives of the orbits by considering only the non-∞ entries.

Corollary 1. *For a base sequence $\underline{a}$ of period s and a shift sequence $\underline{e}$ of length T with no ∞ entries, the period of the interleaved sequence $\underline{u}$ is equal to sT if $sT|s^T$.*

Proof. Suppose $u_{i+l} = u_i$ and $l = \frac{sT}{n}$. In the proof of Theorem 7 we showed that $n \mid T$, and by the assumptions, that $T \mid s^{T-1}$. Let p be any prime divisor of n. Then, $p \mid s^{T-1}$, but since p is prime, $p \mid s$. We have that $\gcd(n, s) = 1$, so a divisor of n cannot also be a divisor of s. This contradiction implies that $l = sT$. □

Define a sequence $\underline{a}$ of period s as *palindromic* if there exists some j such that $a_i = a_{s-1-i+j}$ for all $0 \leq i \leq s-1$. We have three cases in total:

- s is odd. For example, $\underline{a} = (1, 0, 1)$. In these cases j is 0.
- s is even.
  - The sequence is palindromic without any adjustment factor. For example, $\underline{a} = (0, 1, 1, 0)$. In these cases j is 0.
  - We need an adjustment factor, for example, $\underline{a} = (0, 1, 2, 1, \ldots)$. In these cases $j = 1$.

If necessary, we perform left shifts on the base sequence until the adjustment factor is 0 or 1. For example, if $\underline{a} = (0, 1, 1, \ldots)$, we use $\underline{a}' = L^{(2)}(\underline{a}) = (1, 0, 1, \ldots)$; and if $\underline{b} = (1, 2, 1, 0, \ldots)$, we use $\underline{b}' = L^{(3)}(\underline{b}) = (0, 1, 2, 1, \ldots)$. More formally, if $a_i = a_{s-1-i+j}$ and $\underline{a}' = \mathbb{L}(a)$, then $a'_i = a_{i+1} = a_{s-1-i-1+j} = a'_{s-1-i+(j-2)}$ so we can assume $j = 0$ or 1 and the three cases above exhaust the possibilities.

Theorem 8. *Let $\underline{a}$ be a base sequence of period s, $\underline{e}$ be a shift sequence of length T, and $\underline{u} = IL(\underline{a}, \underline{e})$. Then $u_i = u_{sT-1-i}$ if and only if $a_i = a_{s-1-i+j}$ and $e_i = -e_{T-1-i}$.*

Proof. First we show the relationship between the entries of the shift sequence; that is, that $e_i = -e_{T-1-i}$. Consider the ith column of the interleaved sequence in matrix form. The entry a_0 appears in the column in position $s - i$. Because $\underline{u}$ is palindromic, the entry a_0 must then appear in column $T - 1 - i$ in position e_i. That forces the top entry in that column, e_{T-1-i}, to be equal to $s - e_i$. We equate them, then use the fact that the addition is performed modulo s:

$$s - e_i = e_{T-1-i}$$

$$e_i = -e_{T-1-i}$$

Now Since $\underline{u}$ is palindromic, we know that $u_k = u_{sT-1-k+j}$, for all k where $0 \leq k < sT$. Consider the first column of U that is not all zeroes, that is, the first non-∞ entry in the shift sequence. Let this be column i. The entries in this column are a_{e_i+n} where i is fixed and $0 \leq n \leq s - 1$.

First we convert the $\underline{a}$ form to $\underline{u}$ form. Because $\underline{u}$ is palindromic, we know that

$$\begin{aligned} a_{e_i+n} &= u_{nT+i} \\ &= u_{sT-1-(nT+i)} \\ &= u_{(s-n)T-1-i}. \end{aligned}$$

To be a valid value of u_{xT+y}, we must have that $0 \leq x \leq s$ and $0 \leq y \leq T-1$. Since $-1-i+j < 0$, we add T to bring this value into the correct range and obtain

$$u_{(s-n-1)T+(T-1-i)} = a_{e_{(T-1-i)}+(s-1-n)}.$$

We substitute our value for e_i:

$$\begin{aligned} &= a_{-e_i+(s-1-n)} \\ &= a_{s-1-n-e_i}. \end{aligned}$$

Therefore $\underline{a}$ is palindromic, since we have that $a_k = a_{(s-1-k)}$, where $k = e_i + n$.

Conversely, assume we have base sequence $\underline{a}$ and shift sequence $\underline{e}$ such that $a_i = a_{s-1-i+j}$ and $e_i = -e_{T-1-i}$. We start with any u_{sT-1-i}, where $i = yT + z$, and show that it is equal to u_i (with a possible j).

$$\begin{aligned} u_{sT-1-i} &= u_{sT-1-yT-z} \\ &= u_{(s-y)T-1-z} \\ &= a_{e_{-1-z}+s-y} \\ &= a_{s-1-e_{-1-z}-s+y+j} \\ &= a_{-e_{-1-z}-1+y+j} \\ &= a_{e_{T-1-(-1-z)}-1+y+j} \\ &= a_{e_{T+z}-1+y+j} \\ &= a_{e_z+1-1+y+j} \\ &= a_{e_{z+j}+y} \\ &= u_{(y+j)T+z} \end{aligned}$$

So, in the case that $j = 0$, this is simply $u_{yT+z} = u_i$ as required. In the case that

$j = 1$ for the base sequence $\underline{a}$, we have that $u_{sT-1-i} = u_{(y+1)T+z} = u_{i+T}$. This fits our definition of palindromic, including the adjustment factor of T. Therefore $\underline{u}$ is palindromic as required. □

Example 3. Let $\underline{a} = (0, 1, 2, 1)$ and $\underline{e} = [\infty, 0, 1, 3, 0, \infty]$.

Because we have that $a_1 = a_3$, $a_2 = a_2$, and $a_0 = a_4$, we have that $\underline{a}$ is palindromic with $j = 1$:

$$a_0 = a_{s-1-0+1} = a_4,$$

$$a_1 = a_{s-1-1+1} = a_3,$$

$$a_2 = a_{s-1-2+1} = a_2.$$

The interleaved sequence is $\underline{u} = 00110001201002112001021 0 \ldots$. We can test any entries we like to see that it is palindromic; $u_0 = u_{29} = u_5$, $u_7 = u_2 2$, and so on. We see that $\underline{u}$ is in fact palindromic, with $u_i = u_{sT-1-i+T}$.

3.3.4 Coverage

The property of t-coverage generalizes the t-tuple property. Instead of taking consecutive values, we can take any set of positions.

Definition 6. A sequence $\underline{a}$ has *t-coverage* for an integer t if, given any non-zero t-tuple in $\mathbb{F}_q$ and a set of t positions i_j for $0 \leq j \leq t-1$, there exists a starting position k in $\underline{a}$ such that $(a_{k+i_1}, a_{k+i_2}, \ldots, a_{k+i_{t-1}})$ is the tuple.

Example 4. The sequence $\underline{a} = (0, 0, 1, 0, 1, 1, 1, \ldots)$ has 2-coverage. The non-zero 2-tuples are (0,1), (1,0), and (1,1). Given any two positions, say $i_0 = 0$ and $i_1 = 2$, we can find all three tuples at some starting point. For (0, 1) the starting point is $k = 0$, since $a_0 = 0$ and $a_2 = 1$. For (1, 0) the starting point is $k = 5$, since $a_5 = 1$ and $a_7 = 0$. For (1, 1) the starting point is $k = 2$, since $a_2 = 1$ and $a_4 = 1$.

Returning to the pattern we noticed in section 3.1.1, it seems we should look for shift sequences that have some kind of property where the differences between entries are evenly distributed. *Starter vectors* were introduced by Meagher and Stevens [8], also for the purpose of constructing covering arrays using a different method than we use here.

Definition 7. A *starter vector* is a sequence $\underline{e} = [e_0, e_1, \ldots, e_{T-1}]$ of elements taken from $\mathbb{Z}_s \cup \{\infty\}$, where for every $g \in \mathbb{Z}_T$ (the *gap*) and every $d \in \mathbb{Z}_s \cup \{\infty, -\infty\}$ (the *difference*), there exists an i such that:

$$d = e_{i+g} - e_i \pmod{s}$$

where the subscripts are added modulo T, and for any $z \in \mathbb{Z}$, we use the convention that $\infty - z = \infty$ and $z - \infty = -\infty$.

We build on this concept with two changes. First, rather than simply a representative for each i, we want each set to have entries equally distributed, as much as possible. Second, we need to accommodate the rotation of shift sequences as we defined it earlier.

Definition 8. A *balanced-difference near-starter vector* is a sequence $\underline{e} = [e_0, e_1, \ldots, e_{T-1}]$ of elements taken from $\mathbb{Z}_s \cup \{\infty\}$, with the following conditions. Consider the sets d_i, $1 \leq i \leq k-1$, where

$$d_i = \{(e_{j+i} - e_j) \bmod s : 0 \leq j < T - i\} \cup \{(e_{j+i} + 1 - e_j) \bmod s, T - 1 \leq j < T\}$$

where any difference is omitted if $e_j = \infty$ or $e_{j+i} = \infty$, and the subscripts are taken modulo T. Each set of d_i must contain every element from 0 to $s-1$ at least once, and the number of each element cannot vary by more than one from each other element.

Example 5. For a base sequence of period 7 and a shift of length 9, the vector

$[\infty, 0, 0, 5, 0, 4, 3, 6, 0]$ is a balanced-difference near-starter vector. Considering consecutive elements ($i = 1$), we have that $d_1 = (-\infty, 0, 5, 2, 4, 6, 3, 1, \infty)$. Considering a gap of 2, $d_2 = (-\infty, 5, 0, 6, 3, 2, 4, \infty, 1)$. We continue in this way until we are considering a gap of 8, where $d_8 = (-\infty, \infty, 1, 3, 6, 4, 2, 5, 0)$. In all cases the elements 0 through 6 appear in the sets d_i. Here is a closer look at some of the calculations when $i = 6$, for clarity.

When $j = 0$: $i + j < T$, so we calculate $e_6 - e_0 = 3 - \infty = -\infty$.

When $j = 1$: $i + j < T$, so we calculate $e_7 - e_1 = 6 - 0 = 6$.

When $j = 2$: $i + j < T$, so we calculate $e_8 - e_2 = 0 - 0 = 0$.

When $j = 3$: $i + j = T$, so we calculate $e_9 + 1 - e_3 = \infty + 1 - 5 = \infty$.

When $j = 4$: $e_1 + 1 - e_4 = 0 + 1 - 0 = 1$.

As in the t-tuple property discussed in Section 3.3.2, we are interested in finding interleaved sequences that have greater t-coverage than the underlying base sequence. The decomposition of m-sequences tells us that an interleaved sequence with $q^n - 1$ elements can be decomposed into a base sequence of length $q^m - 1$ as long as m is a proper factor of n. The interleaved sequence has the t-tuple property for $t = n$, and the base sequence has it for $t = m$. Since the m-sequence can be decomposed, we can reverse the process.

For example, for a binary base m-sequence $\underline{a} = (0, 0, 1)$ of period 3 and a shift sequence $\underline{e}$ of length 5, we know that there must exist at least one shift sequence that creates a binary m-sequence of period 15. Since it is an m-sequence, we know it has the n-tuple property. From the balance property discussed in Section 3.3.1, we know that we need $3/3 = 1$ shift of ∞ in the shift sequence. An exhaustive search of all possible shift sequences finds two of length 5 over $\mathbb{Z}_3 \cup \{\infty\}$ with one ∞ that fit the bill. They are $[\infty, 0, 0, 2, 0]$ and $[\infty, 0, 1, 0, 0]$, which are reverses of each other by Section 3.2.2. They both fit the criteria to be balanced-difference near-starter vectors.

4 Arrays from Interleaved Sequences

4.1 Coverage Property

Earlier we discussed the *t-coverage* property of a sequence, where we take a t-set of positions and consider each entry as we offset the t positions by a different value.

As the name implies, we are interested in the coverage property for the possibility of creating covering arrays from these sequences. We can create a covering array by considering the *subinterval array* of the sequence.

Definition 9. Let $\underline{a}$ be a sequence of period s, and let $C_i^s(\underline{a})$ be the subinterval of $\underline{a}$ of length s beginning in position i. The subinterval array of $\underline{a}$ of length n is an $s \times s$ array A^s where

$$A^s = \begin{bmatrix} C_0^s(\underline{a}) \\ C_1^s(\underline{a}) \\ \vdots \\ C_s^s(\underline{a}) \end{bmatrix}.$$

If the sequence $\underline{a}$ has t-coverage for some t, then the array A^s has strength t.

Theorem 9. *Let $\underline{a}$ be an m-sequence with the 2-tuple property, and let $\underline{e}$ be a balanced-difference near-starter vector with at least one entry of ∞. Then the interleaved sequence $\underline{u} = IL(\underline{a}, \underline{e})$ has 2-coverage, with one exception: for each pair of relative positions $i, i+j$ for which the coverage property fails in $\underline{a}$, there will be a pair of relative positions $i, i+jT$ uncovered in $\underline{u}$.*

Proof. To have the 2-coverage property, we need to show that for every pair of relative positions j and $j+k$ in the interleaved sequence and every tuple in $\{00, 01, 10, 11\}$, there exists a position i such that (u_j, u_{j+k}) form the desired tuple.

First consider the case where k is not a multiple of T. We can express $k = wT + r$, where $r, w \in \mathbb{Z}$ such that $0 < r < T$. Pick any non-zero tuple; it must occur

somewhere in $\underline{a}$ since it has the 2-tuple property. Let the positions be a_l and a_{l+1}. (They must be consecutive positions because $\underline{a}$ is only assumed to have the 2-tuple property, not 2-coverage.)

Our choice of shift sequence ensures that there is a pair of shift sequence entries e_m, e_{m+r} such that $e_{m+r} - e_m = w$. This gives us the tuple a_l, a_{l+1} in positions e_m, e_{m+r}. Since we can do this for any tuple, we have the coverage property in this case.

Now consider when k is a multiple of T; say $k = oT$ for some integer o. This means we are considering relative positions $i, i+oT$ in $\underline{u}$, which correspond to positions $i, i+o$ in $\underline{a}$. If these relative positions are covered in $\underline{a}$, they will be covered in $\underline{u}$; but if they are not covered in $\underline{a}$, they cannot be covered in $\underline{u}$. These establish the uncovered columns in $\underline{u}$.

The existence of the shift of ∞ provides for the zero tuple in all cases. If k is a multiple of T, any pair in that column is the zero tuple. If k is not a multiple of T, since every column of $\underline{u}$ has at least one zero entry, we simply choose any zero entry in a column r positions from the shift. □

Example 6. Consider the base sequence $\underline{a} = (0, 0, 1, 1 \ldots)$, which is an m-sequence slightly modified to have the 2-tuple property, but not 2-coverage. To see this, note that the relative positions i and $i+2$ give the tuples 01 and 10 twice in each period, but never 00 or 11. The shift sequence $\underline{e} = [\infty, 0, 0, 2, 1, 2]$ is a balanced-difference near-starter vector for the case where $s = 4$ and $T = 6$. When we create $\underline{u} = IL(\underline{a}, \underline{e})$, we find a sequence of period 24 with 2-coverage everywhere except the relative positions i and $i + 12$. These occur 12 times, meaning that the interleaved sequence has 264 pairs of columns covered out of a possible $\binom{24}{2} = 276$ pairs.

number of 2-sets covered	number of shifts with this coverage	smallest shift (lex least)	largest shift (lex greatest)
66	9	$[\infty, 0, 0, 0]$	$[\infty, 0, 2, 2]$
54	2	[0, 0, 1, 2]	[0, 0, 2, 0]
48	4	$[\infty, \infty, 0, 0]$	[0, 0, 1, 0]
42	1	[0, 0, 0, 2]	
36	3	[0, 0, 0, 0]	[0, 2, 0, 2]
30	1	$[\infty, 0, \infty, 0]$	
24	2	$[\infty, 0, \infty, 2]$	[0, 0, 2, 2]
12	1	$[\infty, \infty, \infty, 0]$	
0	2	$[\infty, \infty, \infty, \infty]$	[0, 1, 2, 0]

Table 5: *Base sequence (0,1,1). Comparison of coverage all possible shift sequences of length 4.*

4.2 Experimental Results

A simple computer program makes it easy for us to examine the coverage properties of various pairs of base sequences and shift sequences.

First we would like to understand the typical pattern of coverages for a fixed base sequence and all representative shift sequences. Table 5 shows an example for the base sequence (0, 1, 1) and all shifts of length 4. The number of 2-sets in the interleaved sequence is $\binom{12}{2} = 66$. Because the base sequence $(0, 1, 1)$ has 2-coverage, we expect that many shift sequences maintain the property of 2-coverage in the interleaved sequence, and the chart confirms that - we have that nine shift sequences have full coverage of all 66 2-sets. For balance, we can also calculate $\frac{T-1}{s+1} = 3/4$, and we see in the chart that shift sequences with one infinity seem to perform best, as expected. From this point on, we consider only shift sequences with the correct number of shifts of ∞.

A more interesting question is: can we achieve 3-coverage in the interleaved sequence with 2-coverage in the base sequence? Our experimental results show that when the length of the shift sequence is long enough, we can get close to complete coverage in the interleaved sequence. To illustrate this, the next two tables show the

shift seq. length	3-sets covered	3-sets total	best shift seq.	no. missing	%
4	168	220	$[\infty, 0, 2, 0]$	52	24
5	360	455	$[\infty, 0, 0, 1, 0]$	95	21
6	720	816	$[\infty, 0, 0, 0, 2, 0]$	96	12
7	1281	1330	$[\infty, \infty, 0, 0, 1, 0, 0]$	49	4
8	2016	2024	$[\infty, \infty, 0, 0, 1, 2, 1, 1]$	8	0.4
9	2916	2925	$[\infty, \infty, 0, 0, 0, 2, 1, 1, 2]$	9	0.3

Table 6: *Base sequence (0,1,1). Comparison of 3-coverage of various lengths of shift sequences.*

shift seq. length	3-sets covered	3-sets total	best shift seq.	no. missing	%
3	264	276	$[0, 0, 0]$	12	4
4	480	496	$[\infty, 0, 0, 1]$	16	3
5	760	780	$[\infty, 0, 0, 0, 1]$	20	2.6
6	1104	1238	$[\infty, \infty, 0, 0, 0, 1]$	24	1.9

Table 7: *Base sequence (0,1,1,2,0,2,1,1). Comparison of 2-coverage of various lengths of shift sequences.*

3-coverage of interleaved sequences created from two different base sequences with 2-coverage. Table 6 uses the base sequence (0,1,1), and Table 7 uses the sequence (0, 1, 1, 2, 0, 2, 1, 1) over $\mathbb{F}_3$.

In Table 6, something interesting happens when we get to shift sequences of length 8. It is worth digressing a little to understand whether this is best possible, and whether the coverage can be repaired in some way. For example, when creating covering arrays from m-sequences, it is common to add a row of zero elements to the final covering array to compensate for the missing zero-tuple in the base sequence.

The eight 3-sets that are not covered occur in the pattern $(i, i+8, i+16)$, where the addition is performed modulo 16. Because the length of the shift sequence is 8, this means that we are looking at 3-sets created from a single column in the interleaved array. These columns are all 3 elements long (the length of the base sequence), and so they are either some shift of the base sequence, or a zero column coming from

the shifts of ∞. This easily explains why the 3-sets are not covered, since the base sequence does not have 3-coverage. To say it another way, the only possible 3-tuples that can be covered are 000, 011, 110, and 101, corresponding to the four different columns of the interleaved matrix.

It is not as clear why these eight 3-sets are the only ones not covered. Further work involves a proof of the existence of a shift sequence with the property that only T sets are uncovered. It may be helpful that the number of missing t-sets increases linearly with the size of T, where the number of t-sets increases as a polynomial of degree t.

A possible repair could be made by adding four rows to the final array to cover the missing tuples; however, the best possible CA(3,24,3) is already known to have 21 rows, and ours would have 28. In this case it is not a fruitful strategy, but when considering 4-coverage, it may be helpful. This is one of the next steps to consider in this direction of research.

4.2.1 Constructing shift sequences

Now we turn to creating and testing balanced-difference near-starter vectors. A computer search produces small examples readily. Our parameters are s and T, where T is the length of the shift sequence and s is the size of the group used for the entries of the shift sequence. We omit the cases where $s = 1$ or $T < 4$ as being too small to be interesting. Table 8 shows some examples.

Our final question is how well these balanced-difference near-starter vectors work as shift sequences. For small values of s and T, it is feasible to check all possible shift sequences and determine which have the best coverage. In Table 9 we compare

	$s = 2$	$s = 3$	$s = 4$
$T = 4$	$[\infty, 0, 0, 1]$	n/a	n/a
$T = 5$	$[\infty, 0, 0, 0, 1]$	$[\infty, 0, 0, 2, 0]$	n/a
$T = 6$	$[\infty, \infty, 0, 0, 0, 1]$	$[\infty, 0, 0, 0, 2, 0]$	$[\infty, 0, 0, 2, 1, 2]$
$T = 7$	$[\infty, \infty, 0, 1, 0, 0, 0]$	$[\infty, \infty, 0, 0, 1, 2, 1]$	$[\infty, 0, 0, 0, 2, 1, 2]$
$T = 8$	$[\infty, \infty, 0, 0, 0, 1, 0, 0]$	$[\infty, \infty, 0, 0, 0, 1, 2, 1]$	$[\infty, 0, 0, 0, 1, 3, 0, 3]$
$T = 9$	$[\infty, \infty, 0, 0, 0, 1, 0, 1, 1]$	$[\infty, \infty, 0, 0, 1, 1, 0, 1, 0]$	$[\infty, \infty, 0, 0, 0, 1, 3, 2, 1]$

	$s = 5$	$s = 6$	$s = 7$
$T = 8$	n/a	$[\infty, 0, 0, 5, 2, 0, 2, 3]$	n/a
$T = 9$	$[\infty, 0, 0, 0, 1, 3, 0, 4, 2]$	n/a	$[\infty, 0, 0, 5, 0, 4, 3, 6, 0]$
$T = 10$	$[\infty, 0, \infty, 0, 0, 0, 4, 2, 3, 0]$	$[\infty, 0, 0, 1, 0, 3, 5, 1, 5, 4]$	
$T = 11$	$[\infty, \infty, 0, 0, 0, 1, 3, 0, 3, 2, 0]$		

Table 8: *Examples of balanced-difference near-starter vectors for small values of s and T.*

shift seq. length	3-sets total	maximum 3-sets covered	3-sets covered by BDNS	difference
5	455	420	420	0
6	816	720	720	0
7	1330	1281	1281	0
8	2024	2016	2016	0
9	2925	2916	2916	0

Table 9: *Base sequence (0,1,1). Comparison of best possible coverage to coverage generated by balanced-difference near-starter vectors.*

the coverage generated by the balanced-difference near-starter vectors. We find that the balanced-difference near-starter vectors perform as well as best possible shift sequences of each possible length. More investigation is needed to see whether this is true for other base sequences and shift sequence lengths.

5 Conclusion

5.1 Results

Covering arrays have been studied for at least 25 years. The area of finite fields has given us many methods of construction. Heuristic methods such as simulated annealing have also found many specific examples. By studying interleaved sequences, we hope to find another method of construction that can be easily implemented.

We began by establishing an equivalence relation on shift sequences, to reduce the number of shift sequences needing to be tested.

We reviewed the balance, tuple, and coverage properties. For m-sequences, we established the conditions for the interleaved sequence to inherit the balance properties from the base sequence. Theorem 5 states that for an m-sequence with period s and a shift sequence of length T, the interleaved sequence will satisfy the balance property if and only if the number of ∞ entries in the shift sequence is exactly $(T-1)/(s+1)$, regardless of the finite field the m-sequence elements are taken from. In future work, we aim for a more general result for all types of sequences.

For the tuple property, we showed one simple condition for an interleaved sequence to inherit the property from the base sequence. Theorem 6 states that if a shift sequence contains t consecutive entries, the t-tuple property of the base sequence will be maintained. For the coverage property, our experimental results suggest that it is often possible to find a shift sequence that actually increases the coverage of the interleaved sequence, except for a few sets of columns.

We examined the possible periods of the interleaved sequence, and proved the somewhat surprising result that the period will be maximal (that is, equal to sT, the product of the period of the base sequence and the length of the shift sequence) if some simple conditions are met. Theorem 7 shows that if we factor $sT = ln$, we can have interleaved sequences of period l as long as $n|T$ and $\gcd(n, s) = 1$. Corollary 1

gives the simple result that if it is possible to choose s and T such that $sT|s^T$, any interleaved sequence created will have maximal period.

Finally, we show the precise conditions for an interleaved sequence to be palindromic. Naturally the base sequence must be palindromic, but we also show the precise form the shift sequence must take; it needs to be reversed, but each element must also be negated.

5.2 Open Questions

We have observed that our covering arrays have 'good' coverage by some measurement, but not full coverage. There are at least two ways we can compensate for the lack of coverage. The first is to repair the matrix by adding rows that add the missing t-sets. The second is to view these covering arrays as *partial m-covering arrays*, as introduced by Bonis et. al [6].

Our experimental results suggest that balanced-difference near-starter vectors preserve the coverage property when used as shift sequences. Further investigation is required to either prove the assertion, or determine for which cases the result holds.

In this thesis, we have focused on m-sequences as base sequences. A natural avenue for investigation is to consider different types of sequences, and determining which properties are preserved in interleaved sequences. Examples to consider are twin prime sequences, Legendre sequences, and sextic sequences.

The main focus of ongoing research is establishing the exact results for coverage of an interleaved sequence, based on the coverage of the base sequence and the construction of the shift sequence. Being able to build a base sequence of length s and a shift sequence of length T, rather than directly having to build a sequence of length sT, may prove to be a faster method of obtaining sequences.

References

[1] C. J. Colbourn and J. H. Dinitz. *CRC Handbook of Combinatorial Designs*. CRC Press, 1996.

[2] L. Euler. *Recherches sur une nouvelle espect de quarres magiques*. 1782.

[3] R. Fisher. *The Design of experiments*. Oliver, 1951.

[4] R. Games. Crosscorrelation of m-sequences and gmw-sequences with the same primitive polynomial. *Discrete Applied Mathematics*, 12:139–146, 1985.

[5] S. W. Golomb and G. Gong. *Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar*. Cambridge University Press, 2005.

[6] A. de Bonis K. Sarkar, C. J. Colbourn and U. Vaccaro. Partial covering arays: Algorithms and asymptotics. *Theory of Computing Systems*, pages 1–20, 2017.

[7] R. Lidl and H. Neiderreiter. *Finite Fields*. Addison-Wesley, 1983.

[8] K. Meagher and B. Stevens. Group construction of covering arrays. *Journal of Combinatorial Designs*, 13:70–77, 2005.

[9] B. Stevens and E. Mendelsohn. Efficient software testing protocols. *Proc. of Center for Advanced Studies Conf. (Cascon '98)*, pages 270–293, 1998.

[10] F. Swetz. *Legacy of the Luoshu*. A K Peters/CRC Press, 2008.

[11] G. Tarry. Le problème de 36 officiers. *Compte Rendu de l'Association Française pour l'Avancement des Sciences*, 1:122–123, 1900.

[12] Wikipedia contributors. Magic square — Wikipedia, the free encyclopedia. `https://en.wikipedia.org/w/index.php?title=Magic_square&oldid=849716469`, 2018. [Online; accessed 13-July-2018].