Statistics and Applications {**ISSN 2452-7395 (online)**}
Volume 17, No. 2, 2019 (New Series), pp 87-98

Optimality of Binary Periodic Sequences of Odd Length and Its Applications to Finding Near Optimal 2-symbol Factorial Designs

Krishnasamy T. Arasu[1], Dursun A. Bulutoglu[2] and Jeffrey R. Hollon[3]
[1]*Riverside Research, 2640 Hibiscus Way, Beavercreek, OH 45431, USA*
[2]*Department of Mathematics and Statistics, Air Force Institute of Technology, WPAFB, OH 45433*
[3]*Applied Optimization Inc., 3040 Presidential Dr. Suite 100, Fairborn, OH 45324*

Received: April 24, 2019; Revised: October 01, 2019; Accepted: October 24, 2019

Abstract

We propose using periodic binary sequences with optimal correlation energy (CE) to generate near $E(s^2)$-optimal supersaturated designs (SSDs) and near D-optimal 2-symbol fractional factorial designs for the all main effects and the intercept model. We derive a lower bound for the CE of odd length periodic sequences and provide previously unknown odd length periodic sequences with optimal CE up to length 43.

Key words: Factorial designs; Supersaturated designs; Binary periodic sequences; Correlation energy.

1. Introduction

In this section, we provide the background material on binary periodic sequences, their periodic autocorrelations and correlation energy (CE). We also introduce the group ring notation for investigating periodic sequences. Such sequences with optimal CE generate near $E(s^2)$-optimal 2-symbol supersaturated designs (SSDs) and near D-optimal 2-symbol fractional factorial designs for the all main effects and the intercept model. In Section 2, we derive a lower bound for the CE of odd length periodic sequences. In Section 3, we provide previously unknown odd length periodic sequences with optimal CE up to length 43.

1.1 Sequences and Their Periodic Autocorrelations

A sequence $\mathbf{a} = (a_i)$, where $i = 0, 1, ..., v-1$, is called *periodic* with period (length) v if $a_i = a_{i+v}$ for all i. Such a sequence is also called a $\mathbb{Z}_v$-sequence. In this paper, we consider binary sequences consisting of entries in $\{-1, 1\}$. Let

$$C_{\mathbf{a},\mathbf{b}}(t) = \sum_{i=0}^{v-1} a_{(i+t) \bmod v} \bar{b}_i,$$

Corresponding Author: Dursun Bulutoglu
E-mail: dursun.bulutoglu@gmail.com

where $\bar{b}_i$ is the complex conjugate of b_i. Then $C_{\mathbf{a},\mathbf{b}}(t)$ is called the *periodic cross-correlation function* of $\mathbf{a}$ and $\mathbf{b}$, where the special case $C_{\mathbf{a},\mathbf{a}}(t)$ is called the *periodic autocorrelation function* (ACF) of $\mathbf{a}$. The sequence $\{C_{\mathbf{a},\mathbf{a}}(t)\}_{t=0}^{\infty}$ is again periodic with period (length) v. It is also easy to verify that $C_{\mathbf{a},\mathbf{a}}(t) = C_{\mathbf{a},\mathbf{a}}(-t)$. Hence, it suffices to find the autocorrelation coefficients $C_{\mathbf{a},\mathbf{a}}(t)$ for $t \in \{0, 1, \ldots, \lfloor v/2 \rfloor\}$. The ACF provides a measure of how much the original sequence differs from its translates. Define $D = \{0 \leq i \leq v-1 : a_i = 1\}$ and $d_D(t) = |(t+D) \cap D|$. Then $d_D(t) = |(t+D) \cap D|$ is called the *difference function* of $D \subseteq \mathbb{Z}_v$. It is easy to show that

$$C_{\mathbf{a},\mathbf{a}}(t) = v - 4(k - d_D(t)), \tag{1}$$

where $k = |D|$. From equation (1) we readily get $C_{\mathbf{a},\mathbf{a}}(t) \equiv v \pmod 4$. We call $C_{\mathbf{a},\mathbf{a}}(v)$, $C_{\mathbf{a},\mathbf{a}}(2v)$, $C_{\mathbf{a},\mathbf{a}}(3v), \ldots$ the *main lobes*, and the remaining $C_{\mathbf{a},\mathbf{a}}(i)$ *side lobes*. The value $\max_{g \in G} |C_{\mathbf{a},\mathbf{a}}(g)|$ is called the *peak side lobe*. A sequence $\mathbf{a}$ is said to have *good matched autocorrelation properties* if the peak side lobes in the autocorrelation are small and the sum of the squares of the side lobes in the autocorrelation is small.

Definition 1. *The periodic merit factor (PMF) of a sequence* ***a*** *is defined to be*

$$PMF = \frac{C^2_{\mathbf{a},\mathbf{a}}(0)}{\sum_{l=1}^{v-1} C^2_{\mathbf{a},\mathbf{a}}(l)}.$$

It is desirable to have a large PMF. Hence, our goal is to find sequences with the maximum PMF. Maximizing the PMF is analogous to maximizing the Golay merit factor, where the only difference is that the Golay merit factor is based on a sequence's aperiodic autocorrelation function Green and Green (2002).

Definition 2. *The correlation energy (CE) of a sequence* $\mathbf{a}$ *is defined by*

$$CE(\mathbf{a}) = \sum_{l=1}^{v-1} C^2_{\mathbf{a},\mathbf{a}}(l).$$

Maximizing the PMF of a $\{-1, 1\}$ sequence is equivalent to minimizing its CE. A sequence that minimizes the CE is called CE-optimal. We seek to identify what $\{-1, 1\}$ sequence(s) are CE-optimal. A $\{-1, 1\}$ CE-optimal sequence with $C_{\mathbf{a},\mathbf{a}}(i) = 0$ for $i = 1, 2, \ldots, v-1$ is called a *perfect sequence*. The only perfect sequence known is a row of the circulant Hadamard matrix of order 4. For $v \equiv 2 \pmod 4$ CE optimality is guaranteed to occur when $C_{\mathbf{a},\mathbf{a}}(i) = \pm 2$. For $v \equiv 1 \pmod 4$ and $v \equiv 3 \pmod 4$ CE optimality is likely to occur if each $C_{\mathbf{a},\mathbf{a}}(i)$ for $i = 1, 2, \ldots, v-1$ is in $\{-3, 1\}$ and $\{-1, 3\}$.

Next, we introduce the group ring notation that is needed in deriving our results.

Definition 3. *Let G be a finite group and R a ring, where $G = \{g_0, g_1, \ldots, g_{n-1}\}$. Then the group ring of G over R is the set denoted by $R[G]$ defined as:*

$$R[G] = \left\{ \sum_{g \in G} a_g g \mid a_g \in R \right\}.$$

When working with the group ring notation, multiplication and addition are defined in a way similar to those of polynomials. We further define the power of a group ring element in the following way.

Definition 4. *If $W = \sum_{g \in G} a_g g$ is an element of $R[G]$, and t is some integer, then*

$$\left(\sum_{g \in G} a_g g\right)^{(t)} = \sum_{g \in G} a_g g^t.$$

When we refer to a binary $\mathbb{Z}_v$-sequence, we mean a $\mathbb{Z}_v$-sequence with entries from $\{-1, 1\}$ or $\{0, 1\}$. By abuse of notation, we identify a $\mathbb{Z}_v$-sequence $\mathbf{a}$ with the group ring element $S = \sum_{g \in \mathbb{Z}_v} a_g g$. Also, for a $\{0, 1\}$ $\mathbb{Z}_v$-sequence, we identify the group ring element $\sum_{g \in A} g$ with the set A. In particular, $\mathbb{Z}_v$ identifies $\sum_{g \in \mathbb{Z}_v} g$. For a multiplicatively (additively) written cyclic group we use 1 (0) as the identity element. The group ring elements that correspond to $\mathbb{Z}_v$-sequences are used to calculate the autocorrelation of a $\mathbb{Z}_v$-sequence $\mathbf{a}$, where

$$C_{\mathbf{a},\mathbf{a}}(g) = \text{coefficient of } g \text{ in } SS^{(-1)}.$$

Let $A = \{d_0, d_1, \ldots, d_{k-1}\} \subseteq \mathbb{Z}_v$. For each $g \in \mathbb{Z}_v$ let a_g be the number of times g appears in A and $S = \sum_{g \in \mathbb{Z}_v} a_g g$. Then A is a (v, k, λ) *difference set* $\mathrm{DS}(v, k, \lambda)$ if

$$SS^{(-1)} = (k - \lambda)0 + \lambda \mathbb{Z}_v \ \in \ \mathbb{Z}[\mathbb{Z}_v],$$

and A is a $(v, k, \lambda, \lambda + 1)$ *almost difference set* $\mathrm{ADS}(v, k, \lambda, \lambda + 1)$ if

$$SS^{(-1)} = k0 + \lambda B + (\lambda + 1)(\mathbb{Z}_v - B - 0) \ \in \ \mathbb{Z}[\mathbb{Z}_v],$$

for some $B \subset \mathbb{Z}_v \setminus \{0\}$.

Remark 1. *Difference sets and almost difference sets are studied in the more general group theoretic context. The term "array" is used instead of the term "sequence" when the group in question is non-cyclic.*

For more on sequences, arrays, and their interplay with group developed combinatorial designs see Arasu (2011) and Arasu *et al.* (2019).

1.2 Using CE-optimal $\mathbb{Z}_v$-sequences to Construct Near D-optimal Designs and Near $E(s^2)$-optimal SSDs

The Hadamard maximum determinant problem seeks an $N \times N$ matrix of ± 1s with the largest possible determinant. Such matrices are called D-optimal matrices. An online source for this problem can be found at Orrick and Solomon (2018). Multiplying a row or a column of a matrix by -1 does not change its determinant. Hence, an $N \times N$ D-optimal design whose first column is the all 1s column always exists. The last $N - 1$ columns of an $N \times N$ D-optimal matrix whose first column is the all 1s column can be used as an N row, $N - 1$ column

2-symbol factorial design for estimating the all main effects and the intercept model. In fact, such a design minimizes the determinant of the variance-covariance matrix among all possible N row, $N-1$ column, 2-symbol factorial designs for the all main effects and the intercept model.

An N row, k factor, 2-symbol factorial design is called a *supersaturated design* (SSD) if $N < k+1$, i.e., if it does not have enough rows to estimate the all main effects and the intercept model. Most of the literature on SSDs assumes that each column in a 2-symbol SSD is *balanced*, i.e. has an equal number for 1s and -1s. However, recently Bulutoglu *et al.*Bulutoglu *et al.* (2019) considered $\{-1,1\}$ SSDs with a prespecified distribution of column sums. Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$ be an N row, m column, 2-symbol SSD with symbols from $\{-1,1\}$. Then the $E(s^2)$ value of $\mathbf{X}$ is defined as

$$E(s^2) = \frac{\sum_{i\neq j} s_{ij}^2}{m(m-1)},$$

where $s_{ij} = \mathbf{x}_i^T\mathbf{x}_j$ for $1 \leq i < j \leq m$. The $E(s^2)$ value criterion is used to compare two 2-symbol SSDs with the same number of rows and columns, where the SSD with the smaller $E(s^2)$ is more desirable Georgiou (2014). An SSD with the smallest possible $E(s^2)$ is called $E(s^2)$-optimal. For the best known $E(s^2)$-lower bounds of balanced SSDs, see Georgiou (2014) and Bulutoglu *et al.*Bulutoglu *et al.* (2019) for unbalanced SSDs.

A $\{-1,1\}$ $\mathbb{Z}_v$-sequence $\mathbf{a}$ of length v can be used to generate a $v \times v$ matrix of ± 1s by taking the sequence as the first column and obtaining the other columns as cyclic shifts. The resulting $v \times v$ matrix $\mathbf{A}$ is called the *corresponding design* to $\mathbf{a}$. Then $\mathbf{a}$ can either be used to generate a fractional factorial design or as an SSD. In the first case, a subset of A's rows are multiplied by -1 so that its first column is the all 1s column. The remaining $v-1$ columns of $\mathbf{A}$ constitute a v row, $v-1$ column factorial design that can be used for estimating the all main effects and the intercept model. When $\mathbf{A}$ is used as an SSD, the matrix $\mathbf{A}$ is taken as the v row, v column SSD, and the relation

$$\frac{CE(\mathbf{a})}{v-1} = E(s^2)$$

connects the $E(s^2)$ of $\mathbf{A}$ to $CE(\mathbf{a})$. Hence, a CE-optimal sequence $\mathbf{a}$ can be used to generate an SSD that is near $E(s^2)$-optimal.

It is well known that a $v\times v$ matrix $\mathbf{A}$ that has small $E(s^2)$ tends to have large determinant Bulutoglu *et al.* (2019). Hence, we expect that if $\mathbf{A}$ is used to generate a fractional factorial design, then the resulting factorial design will be near D-optimal for estimating the all main effects and the intercept model.

2. CE Lower Bounds

It is easy to see that $\mathbf{x}$ is a CE-optimal $\{0,1\}$ sequence if and only if $2\mathbf{x}-\mathbf{1}$ is a CE-optimal $\{-1,1\}$ sequence, where $\mathbf{1}$ is the all 1s sequence. We say that a $\{0,1\}$ sequence $\mathbf{x}$ *corresponds* to a $\{-1,1\}$ sequence $\mathbf{y}$ if $\mathbf{y} = 2\mathbf{x}-\mathbf{1}$. It is easy to see that the autocorrelation function of a $\{0,1\}$ sequence is $d_D(t)$. We will use the notation $C_{\mathbf{a},\mathbf{a}}(t)$ for $\{-1,1\}$ sequences.

If we refer to $C_{\mathbf{a},\mathbf{a}}(t)$ of a $\{0,1\}$ sequence, we mean the autocorrelation function of its corresponding $\{-1,1\}$ sequence. To find CE-optimal $\{0,1\}$ sequences, we first investigate $\{0,1\}$ sequences of odd length having at most three-valued $d_D(t)$s (at most two distinct non-trivial λs that differ by 1). For such a $\{0,1\}$ sequence $\mathbf{a}$ indexed by $\mathbb{Z}_v$ we have

$$SS^{(-1)} = k0 + \lambda B + (\lambda+1)(\mathbb{Z}_v - B - 0) \tag{2}$$

for some $B \subset \mathbb{Z}_v \setminus \{0\}$ with $|B| = t$. Then, by applying the principal character to both sides of equation (2), we get

$$\begin{aligned} \chi_0\left(SS^{(-1)}\right) =& \chi_0\left(k0 + \lambda B + (\lambda+1)(\mathbb{Z}_v - B - 0)\right) \\ |A|^2 =& k + \lambda|B| + (\lambda+1)(|\mathbb{Z}_v| - |B| - 1) \\ k^2 =& k + \lambda t + (\lambda+1)(v - t - 1). \end{aligned} \tag{3}$$

We refer to equation (3) as the group ring equation. By equation (1)

$$\lambda_1 = \lambda = \frac{4k - v + \gamma_1}{4} \quad \text{and} \quad \lambda_2 = \lambda + 1 = \frac{4k - v + \gamma_2}{4}, \tag{4}$$

where $\gamma_j = C_{\mathbf{a},\mathbf{a}}(i_j)$ for some i_j such that $i_j \neq v$ for $j \in \{1,2\}$. Observe that $\gamma_2 = \gamma_1 + 4$. Upon substituting the right hand side of (4) to the group ring equation we get

$$k^2 = k + \left(\frac{4k - v + \gamma_1}{4}\right) t + \left(\frac{4k - v + \gamma_1}{4} + 1\right)(v - t - 1).$$

We solve for $t(v,k,\gamma)$:

$$\begin{aligned} k^2 =& k + \left(\frac{4k - v + \gamma}{4}\right) t(v,k,\gamma) + \left(\frac{4k - v + \gamma}{4} + 1\right)(v - t(v,k,\gamma) - 1) \\ t(v,k,\gamma) =& k - k^2 + \left(\frac{4k - v + \gamma}{4} + 1\right)(v - 1) \\ t(v,k,\gamma) =& k - k^2 + \left(\frac{4k - v + \gamma + 4}{4}\right)(v - 1). \end{aligned} \tag{5}$$

Now, $\max(|\gamma_1|, |\gamma_2|)$ depends on $v \pmod 4$.

Case 1: $v \equiv 1 \pmod 4$, $\max(|\gamma_1|, |\gamma_2|) = \max(|\gamma_1|, |\gamma_1 + 4|) = |\gamma_1|$, where $\gamma_1 = -3$. Then

$$t(v,k,-3) = k - k^2 + \left(\frac{4k - v + 1}{4}\right)(v - 1).$$

In this case, minimizing the CE is equivalent to minimizing $t(v,k,-3)$. Solving for k when

$t(v, k, -3) = 0$, we get

$$\begin{aligned} 0 &= k - k^2 + \left(\frac{4k - v + 1}{4}\right)(v-1) \\ k^2 - k &= \left(\frac{4k - v + 1}{4}\right)(v-1) \\ 4k^2 - 4k &= (4k - v + 1)(v-1) \\ 4k^2 - 4vk + (v-1)^2 &= 0 \\ k &= \frac{4v \pm \sqrt{16v^2 - 16(v-1)^2}}{8} \\ k &= \frac{v \pm \sqrt{2v-1}}{2}. \end{aligned} \tag{6}$$

Hence, k should be rounded to the closest integer k^* to

$$\frac{v \pm \sqrt{2v-1}}{2}$$

so that $t(v, k^*, -3)$ is a nonnegative integer less than v.

Case 2: $v \equiv 3 \pmod 4$, $\max(|\gamma_1|, |\gamma_2|) = \max(|\gamma_2 - 4|, |\gamma_2|) = |\gamma_2|$, where $\gamma_2 = 3$. Thus, we have to minimize

$$\begin{aligned} t^*(v, k, 3) = v - t(v, k, 3) - 1 &= v - \left(k - k^2 + \left(\frac{4k - v + 3 + 4}{4}\right)(v-1)\right) - 1 \\ &= v - k + k^2 - \left(\frac{4k - v + 7}{4}\right)(v-1) - 1 \\ &= k^2 - k - \left(\frac{4k - v + 7}{4}\right)(v-1) - 1. \end{aligned}$$

Now $t^*(v, k, 3)$ has a minimum at

$$k = -\frac{-1 - (v-1)}{2} = \frac{v}{2}. \tag{7}$$

In light of equations (6) and (7), we determine what number of elements are required to minimize the CE for a given length. So, k should be rounded to the closest integer k^* to $v/2$ such that $t(k^*)$ is a nonnegative integer less than v. Based on the knowledge gained thus far, we provide the following table for the parameters of the odd length sequences up to length 49 that are CE-optimal when they exist.

Theorem 1. *When a sequence with the parameters in a row of Table 1 exists, then it is CE-optimal.*

Proof. The result is obvious for cases in which the correlation $|\gamma_j| = 3$ count is 0 or 2. For all the remaining cases, $v \equiv 1 \pmod 4$. So, for each of these cases the next best possibility with $C_{\mathbf{a},\mathbf{a}}(t') \notin \{-3, 1\}$ for some t' is when the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_2 = 5$ is 2 and the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_1 = 1$ is $v - 3$. For each v such that the correlation $|\gamma_j| = 3$ count is 4 or larger, the CE of a sequence with the frequency of $\gamma_2 = 5$ equal to 2 is larger or equal to the corresponding CE in Table 1. □

Table 1: The optimal parameters for odd length sequences up to length 49

v	k^*	Correlation $\|\gamma_j\| = 3$ count	CE
5	1	0	4
7	4	0	6
9	3	2	24
11	6	0	10
13	4	0	12
15	8	0	14
17	6	2	32
19	10	0	18
21	8	4	52
23	12	0	22
25	9	0	24
27	14	0	26
29	11	2	44
31	16	0	30
33	13	4	64
35	18	0	34
37	15	6	84
39	20	0	38
41	16	0	40
43	22	0	42
45	18	2	60
47	24	0	46
49	20	4	80

3. CE-optimal Sequences

We present Table 2 containing sequences with minimum CE (maximum periodic merit factor) by length, v. The remaining columns show the elements of the sequence that are 1, CE, number of 1s k in the sequence, and a column indicating if the sequence is optimal and has the parameters in Table 1. All lengths for which the column labeled "Conform?" is answered with "Y" are optimal as each has a set of parameters that appears in Table 1. If the column contains an "N", then the sequence listed is still optimal, it just does not have the parameters in Table 1. In other words, just because the equations indicate a particular parameter set as being optimal does not mean that such a sequence exists.

The length $v = 17$ is the shortest length for in which the Table 1 hypothetical optimal sequence fails to exist. The Table 1 hypothetical optimal sequence for this length should have $k = 6$ and a CE of 32. The indices of the 1 entries of such a sequence constitutes an ADS$(17, 6, 1, 2)$. However, a computer search revealed that this sequence does not exist. To search for a sequence with the next best possible CE, we must add an element to a hypothetical

ADS$(17, 6, 1, 2)$. For $k = 7$ the group ring equation

$$49 = 7 + 2t + 3(17 - t - 1)$$

gives $t = 6$, and a sequence with the next best possible CE is an ADS$(17, 7, 2, 3)$ with $CE = 64$. We found such a sequence by using a computer search.

For $v = 39$ and $v = 41$, the group ring equations and arguments for the CE-optimal sequences indicate that each of DS$(39, 20, 10)$ and DS$(41, 16, 6)$ is CE-optimal if it exists. Both of these are known not to exist by the Mann test Baumert and Gordon (2004). The nonexistence of DS$(39, 20, 10)$ requires that we look for a sequence with two different non-trivial autocorrelation values. The group ring equations give that the next best possible case is when $k = 18$, and from

$$18^2 = 18 + 8t + 9(39 - t - 1)$$

we get that $t = 36$. This particular sequence was not found by an exhaustive computer search. The next best case can be found by removing another element from the set making $k = 17$. Using the group ring equation,

$$17^2 = 17 + 7t + 8(39 - t - 1)$$

gives that $t = 32$. This sequence was found and is listed in Table 2.

A CE-optimal solution to the length $v = 41$ case has been open Luke and Schotten (2003). An optimal solution based on a DS$(41, 16, 6)$ is known not to exist Lander (1983). If we decrease k by 1 to $k = 15$, then the group ring equation

$$15^2 = 15 + 4t + 5(41 - t - 1)$$

implies $t = -10$. This is not possible. Thus, we must increase k by 1 to $k = 17$. When $k = 17$,

$$17^2 = 17 + 6t + 7(41 - t - 1)$$

giving $t = 8$. Such a sequence was found and is listed in Table 2. The autocorrelation of our length 41 $\{-1, 1\}$ sequence contains 8 -3s and 32 $+1$s. There is another distribution of autocorrelation values with 2 -3s, 2 $+5$s, 36 $+1$s, and $CE = 104$. A computer search proved that such a sequence does not exist.

A DS$(25, 9, 3)$ does not exist by the Mann test Baumert and Gordon (2004). A DS$(27, 14, 7)$ does not exist Lander (1983). We proved that an ADS$(29, 11, 3, 4)$ does not exist by an exhaustive computer search. In fact, to check the correctness of our exhaustive search implementation, we proved the existence/non-existence of each of the corresponding hypothetical difference set or almost difference set in each row of Table 1 by using our computer program. Hence, none of the hypothetical CE-optimal sequences in Table 1 for $v = 25, 27, 29$ exists. The solutions in Table 2 for the $v = 25, 27, 29$ cases are constructed similar to the $v = 17, 39, 41$ cases.

Theorem 2. *Each sequence in Table 2 is CE-optimal.*

Table 2: Optimal sequences of odd length up to length 43

v	Sequence elements	CE	k	Conform?
5	{0}	4	1	Y
7	{1,2,4}	6	3	Y
9	{0,1,3}	24	3	Y
11	{1,3,4,5,9}	10	5	Y
13	{0,1,5,11}	12	4	Y
15	{0,1,2,7,9,12,13}	14	7	Y
17	{0,1,2,3,5,8,12}	64	7	N
19	{1,4,5,6,7,9,11,16,17}	18	9	Y
21	{7,9,12,13,16,18,19,20}	52	8	Y
23	{1,2,3,4,6,8,9,12,13,16,18}	22	11	Y
25	{0,9,10,12,15,16,18,20,23,24}	72	10	N
27	{0,9,11,12,13,16,18,19,22,24,26}	74	11	N
29	{0,9,10,13,15,18,21,22,23,25,27,28}	92	12	N
31	{1,2,4,7,8,14,15,16,19,23,25,27,28,29,30}	30	15	Y
33	{0,9,13,14,15,19,21,22,24,26,29,30,32}	64	13	Y
35	{0,1,3,4,7,9,11,12,13,14,16,17,21,27,28,29,33}	34	17	Y
37	{0,6,12,14,17,19,23,24,27,28,31,33,34,35,36}	84	15	Y
39	{2,4,5,7,9,10,11,14,15,16,23,24,25,27,31,35,38}	86	17	N
41	{0,9,11,14,15,21,22,24,27,29,31,32,33,35,36,39,40}	104	17	N
43	{1,4,6,9,10,11,13,14,15,16,17,21,23,24,25,31,35,36,38,40,41}	42	21	Y

The CE-optimal sequences constructed in this paper can be used to construct ± 1 matrices with large determinants. These matrices can be used to construct fractional factorial designs that are near D-optimal for estimating the all main effects and the intercept model. In particular, the largest known determinant for ± 1 matrices of order 39 is given by Tamura in Tamura (2006) using group divisible designs. Tamura's record holding matrix has a determinant of $2^{43} \times 3^{36} \times 5$. While we can not beat this record, we come surprisingly close by using our optimal sequence of length 39. By creating a matrix whose first row is the sequence itself followed by each of the next 38 rows being a right-circulant shift of the previous, we generate a circulant matrix of order 39. This matrix has a determinant of $2^{36} \times 3^6 \times 5 \times 7 \times 29^3 \times 3331^3$. This is 95.7% of Tamura's determinant.

Acknowledgements

Dr. K. T. Arasu was supported by the U.S. Air Force Research Lab Summer Faculty Fellowship Program sponsored by the Air Force Office of Scientific Research.

The views expressed in this article are those of the authors, and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

References

Arasu, K. T. (2011). Sequences and arrays with desirable correlation properties. *NATO Science for Peace and Security Series - D: Information and Communication Security*, **29**, 136 – 171.

Arasu, K. T., Bulutoglu, D. A. and Hollon, J. R. (2019). Legendre G-array pairs and the theoretical unification of several G-array families. *Submitted*.

Baumert, L. D. and Gordon, D. M. (2004). On the existence of cyclic difference sets with small parameters. In A. J. V. D. Poorten and A. Stein, eds., *High Primes and Misdemeanours: Lectures in Honour of the 60th Birthday of in honor of Hugh Cowie Williams*, volume **41**, 61 – 68. Fields Institute Communications.

Bulutoglu, D., Chatterjee, K., Georgiou, S. D., Koukouvinos, C., Ryan, K. J. and Stylianou, S. (2019). Unbalanced two-symbol $E(s^2)$-optimal designs and D-optimal designs. *Submitted*.

Georgiou, S. D. (2014). Supersaturated designs: A review of their construction and analysis. *Journal of Statistical Planning and Inference*, **144**, 92–109.

Green, D. H. and Green, P. R. (2002). Polyphase power-residue sequences. *Proceedings of the Royal Society of London A Mathematical Physical and Engineering Sciences*, **459**, 817 – 827.

Lander, E. S. (1983). Difference sets. In *Symmetric Designs: An Algebraic Approach (London Mathematical Society Lecture Note Series)*, 120–187. Cambridge University Press, Cambridge.

Luke, H. D. and Schotten, H. D. (2003). Binary and Qudariphase Sequences With Optimal Autocorrelation Properties: A Survey. *IEEE Transactions on Information Theory*, **49**(**12**), 3271 – 3282.

Orrick, W. and Solomon, B. (2018). The Hadamard maximal determinant problem. `http://www.indiana.edu/~maxdet/`.

Tamura, H. (2006). D-Optimal Designs and Group Divisible Designs. *Journal of Combinatorial Designs*, **14**, 451–462.

Appendix

Proof of Theorem 2

The only cases that are not covered by Theorem 1 are when $v = 17, 25, 27, 29, 39, 41$. For each of these cases no solution with parameters in Theorem 1 exists. For a $\{-1, 1\}$ sequence $\mathbf{a}$ it is easy to show that

$$\sum_{t=0}^{v-1} C_{\mathbf{a},\mathbf{a}}(t) = (2k - v)^2.$$

Let $\mathbf{a}$ be a CE-optimal sequence. We first consider the cases when $C_{\mathbf{a},\mathbf{a}}(t) \in \{-3, -1, 1, 3\}$ for each nonnegative $t(v, k, \gamma^*)$, where

$$\gamma^* = \begin{cases} -3 \text{ if } v \equiv 1 \text{ (mod } 4), \\ 3 \text{ if } v \equiv 3 \text{ (mod } 4). \end{cases}$$

Let

$$\alpha(v, k, \gamma^*) = \text{number of } -3\text{s in } \{C_{\mathbf{a},\mathbf{a}}(t)\}_{t=0}^{v-1}.$$

Observe that $\alpha(v, k, \gamma^*)$ completely determines $\{C_{\mathbf{a},\mathbf{a}}(t)\}_{t=0}^{v-1}$. Then, for $v \equiv 1$ (mod 4)

$$\begin{aligned} \alpha + \beta &= v - 1 \\ -3\alpha + \beta &= (2k - v)^2 - v \end{aligned}$$

and

$$\alpha(v, k, -3) = \frac{(2v - 1) - (2k - v)^2}{4}.$$

For fixed $v > 0$, $\alpha(v, k, -3)$ is a quadratic function of k with a maximum at $k = v/2$. Similarly, for $v \equiv 3$ (mod 4)

$$\begin{aligned} \alpha + \beta &= v - 1 \\ 3\alpha - \beta &= (2k - v)^2 - v \end{aligned}$$

and

$$\alpha(v, k, 3) = \frac{(2k - v)^2 - 1}{4}.$$

For fixed $v > 0$, $\alpha(v, k, -3)$ is a quadratic function of k with a minimum at $k = v/2$.

In both $v \equiv 1$ (mod 4) and $v \equiv 3$ (mod 4) cases $\alpha(v, k, \gamma^*)$ must be a nonnegative integer as small as possible. Moreover, $t(v, k, \gamma^*)$ as in equation (5) must be a nonnegative integer. Also, since $C_{\mathbf{a},\mathbf{a}}(t) = C_{\mathbf{a},\mathbf{a}}(-t)$, $\alpha(v, k, \gamma^*)$ must be even. Then for each fixed $v \in \mathbb{Z}^{\geq 0}$, a solution to the integer nonlinear program (INLP)

$$\begin{array}{rl} \min\limits_{k} & \alpha(v, k, \gamma^*) \\ \text{subject to:} & t(v, k, \gamma^*) = y, \\ & v - 1 \geq \alpha(v, k, \gamma^*) = 2x \geq 0, \quad v - 1 \geq k \geq 2, \\ & v - 1 \geq y \geq 0 \quad x, y, k \in \mathbb{Z}, \\ & \text{a sequence } \mathbf{a} \text{ with } \{C_{\mathbf{a},\mathbf{a}}(t)\}_{t=0}^{v-1} \text{ determined by } \alpha(v, k, \gamma^*) \text{ exists} \end{array} \tag{8}$$

describes a CE-optimal sequence among all sequences with $\{C_{\mathbf{a},\mathbf{a}}(t)\}_{t=0}^{v-1} \in \{-3, 3, 1, -1\}$ for each t. Let k^* be a solution of the INLP (8). Then,

$$k^* = \begin{cases} \text{an integer farthest to } v/2 \text{ satisfying constraints of INLP (8) if } v \equiv 1 \text{ (mod } 4), \\ \text{an integer closest to } v/2 \text{ satisfying constraints of INLP (8) if } v \equiv 3 \text{ (mod } 4). \end{cases} \tag{9}$$

For $v \in \{17, 25, 27, 29, 39, 41\}$, each value of k in Table 2 satisfies condition (9), and the corresponding sequence is CE-optimal among all sequences $\mathbf{a}$ of the same length such that

$C_{\mathbf{a},\mathbf{a}}(t) \in \{-3,-1,1,3\}$ for $t = 1, 2, \ldots, v-1$. For cases $v = 17, 25, 29, 41$, the next best possibility with $C_{\mathbf{a},\mathbf{a}}(t') \notin \{-3, 1\}$ for some t' is when the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_2 = 5$ is 2 and the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_1 = 1$ is $v-3$. The CE resulting from this distribution of autocorrelations is smaller than the CE of the corresponding sequence in Table 2 only when $v = 41$. Hence, the length $17, 25, 29$ sequences listed in Table 2 are all CE-optimal. Then, for $v = 41$ the distribution of autocorrelations of a sequence with a smaller CE is given by 38 $+1$s and 2 $+5$s. However, by examining the group ring equation and using equations (4) we find that

$$k^2 = k + 38\left(\frac{4k-41+1}{4}\right) + 2\left(\frac{4k-41+5}{4}\right). \tag{10}$$

Equation (10) has no integer solutions. Thus, the length 41 sequence listed in Table 2 is CE-optimal with $CE = 104$.

For cases $v = 27, 39$, the next best possibility with $C_{\mathbf{a},\mathbf{a}}(t') \notin \{-1, 3\}$ for some t' is when the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_2 = -1$ is $v-3$ and the frequency of $C_{\mathbf{a},\mathbf{a}}(t) = \gamma_1 = -5$ is 2. The CE resulting from this distribution of autocorrelations is equal to the CE of the corresponding sequence listed in Table 2 for both $v = 27$ and $v = 39$ cases.