Statistica Sinica Preprint No: SS-2018-0267	
Title	Minimal Second Order Saturated Designs and Their Applications to Space-Filling Designs
Manuscript ID	SS-2018-0267
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202018.0267
Complete List of Authors	Boxin Tang Ching-Shui Cheng and Yuanzhen He
Corresponding Author	Boxin Tang
E-mail	boxint@sfu.ca

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Minimal Second Order Saturated Designs and Their Applications to Space-Filling Designs

May 17, 2020

Ching-Shui Cheng

cscheng@berkeley.edu, Academia Sinica and University of California, Berkeley

Yuanzhen He

heyuanzhen@bnu.edu.cn, Beijing Normal University

Boxin Tang

boxint@sfu.ca, Simon Fraser University

Abstract

Second order saturated (SOS) designs allow the estimation of a saturated model consisting of main effects and two-factor interactions. Apart from being useful in their own right, SOS designs have recently been applied to the construction of space-filling designs. This paper introduces the notion of minimal SOS designs to facilitate the study of SOS designs, and presents some results on the characterization and construction of minimal SOS designs. Both regular and nonregular minimal SOS designs are considered, and are applied to construct space-filling designs.

Key words and phrases: clear effect; maximal design; nonregular design; space-filling design; strong orthogonal array.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

1. Introduction

Second order saturated (SOS) designs allow the estimation of saturated models consisting of main effects and two-factor interactions (2fi's). They make the most efficient use of the degrees of freedom by allocating all of them to the estimation of factorial effects of the first two orders, which are the most important orders according to the principle of effect hierarchy. SOS designs were first introduced in Block and Mee (2003), and were also discussed in Chen and Cheng (2004) under the notion of estimation index. Constructions of nonregular SOS designs were explored by Cheng, Mee, and Yee (2008). For more details about SOS designs, see Mee (2009, Section 7.2) and Cheng (2014, Section 11.2).

SOS designs are important in their own right, but become more so owing to their utility in designing computer experiments. It is widely accepted that space-filling designs are appropriate choices for computer experiments (Santner, Williams, and Notz 2003). Among the methods available for constructing space-filling designs, that based on orthogonal arrays is very attractive because it provides designs that enjoy some guaranteed space-filling properties in low-dimensional projections. This line of research started with Latin hypercubes, which are orthogonal arrays of strength one, in McKay, Beckman, and Conover (1979), and continued with the work of Owen (1992) and Tang (1993). A significant enhancement to the idea is the recent introduction of strong orthogonal arrays (SOAs) by He and Tang (2013), Liu and Liu (2015), and He, Cheng, and Tang (2018). SOAs can be used to construct designs that have better space-filling properties than those constructed using ordinary orthogonal arrays. In the process of constructing SOAs us-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

ing regular 2^{m-p} designs, He, Cheng, and Tang (2018) found that all such SOAs can be constructed from SOS designs.

This paper examines applications of regular and nonregular SOS designs to the construction of SOAs by introducing the concept of minimal SOS designs. This concept is useful because, as shown later, all SOS designs can be generated from minimal ones. Furthermore, to produce SOAs that can accommodate more factors, one needs SOS designs with fewer factors. Section 2 reviews some material on SOAs of strength 2+ from He, Cheng, and Tang (2018), including how SOS designs can be used to construct SOAs. Section 3 first presents a characterizing result for regular minimal SOS designs using clear effects, and then shows that the four constructions in He, Cheng, and Tang (2018) all produce minimal SOS designs. This section imports several results from projective geometry and coding theory, using the equivalence of regular SOS designs to 1-saturating sets and duals of linear codes with covering radius 2. These results allow us to improve the bounds on the maximum number of factors in an SOA of strength 2+ obtained in He, Cheng, and Tang (2018). In Section 4, we discuss nonregular SOS designs. We first present extensions of the four constructions of regular SOS designs in He, Cheng, and Tang (2018) to nonregular designs, showing that they all give minimal SOS designs. We then show how to use these nonregular minimal SOS designs to construct SOAs of strength 2+. In addition to more flexible run sizes, SOAs constructed from nonregular SOS designs have other advantages, including possibly better three- and higher-dimensional projections. Furthermore, nonregular SOS designs provide more options for constructing SOAs

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

because there are often many more nonisomorphic nonregular designs than regular ones. Section 5 concludes the paper with a discussion.

2. Second Order Saturated Designs and Strong Orthogonal Arrays

A major focus of this study is two-level SOS designs. The following definition applies to both regular and nonregular designs.

Definition 1. *A two-level fractional factorial design with n runs and m factors is second order saturated (SOS) if it can be used to estimate all of the m main effects, together with at least one set of $n-1-m$ 2fi's (assuming that all other effects are negligible).*

Regular SOS designs were first considered by Block and Mee (2003). Under such a design, one can entertain a model with the largest number of 2fi's. Independently, Chen and Cheng (2004) defined the notion of estimation index. It is well known that each regular design is equivalent to a linear code; then the estimation index is the same as the covering radius of the dual code, and a design is SOS if and only if the estimation index is equal to 2.

Given a design of resolution IV, if one factor is added, the resulting design may have resolution III. A resolution IV design is called maximal if no factor can be added to maintain resolution IV. This concept is useful because all resolution IV designs can be obtained as projections of maximal resolution IV designs. It follows from a geometric result in Bruen, Haddad, and Wehlau (1998) that two-level designs of resolution IV are

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

maximal if and only if they have estimation index 2. Therefore, a resolution IV design is maximal if and only if it is SOS. An important byproduct of this result is that every two-level resolution IV design is a projection of a certain SOS design of resolution IV, a fact also observed by Block and Mee (2003). Thus, in addition to being able to entertain the largest number of 2fi's, another important practical value of SOS designs of resolution IV is that they can be used to generate all resolution IV designs of the same run size via projections. For example, there are three 32-run two-level SOS designs of resolution IV: a 2^{16-11}, a 2^{10-5}, and a 2^{9-4}. All other 32-run resolution IV designs can be obtained by deleting factors from one of these three designs.

Unexpectedly and interestingly, SOS designs have a third application: they can be used to construct strong orthogonal arrays. We use $\mathrm{OA}(n, m, s_1 \times \cdots \times s_m, t)$ to denote an orthogonal array of strength t in n runs for m factors, with its jth factor having s_j levels, $0, 1, \ldots, s_j - 1$. If $s_1 = \cdots = s_m = s$, the array is denoted as $\mathrm{OA}(n, m, s, t)$ for simplicity. Hedayat, Sloane, and Stufken (1999) and Dey and Mukerjee (1999) are two good general references for orthogonal arrays.

An $n \times m$ matrix with entries from $\{0, 1, \ldots, s^2 - 1\}$ is called an SOA of strength 2+ for n runs and m factors at s^2 levels if any subarray of two columns can be collapsed into an $\mathrm{OA}(n, 2, s^2 \times s, 2)$ and an $\mathrm{OA}(n, 2, s \times s^2, 2)$. We denote this array by $\mathrm{SOA}(n, m, s^2, 2+)$. Here, collapsing s^2 levels to s levels takes place according to $[a/s]$ for $a = 0, 1, \ldots, s^2 - 1$, where $[x]$ denotes the greatest integer not exceeding x. An s-level orthogonal array of strength two can be used to construct a Latin hypercube design that contains an equal

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

number of points in each cell of $s \times s$ grids in all two-dimensional projections. However, a Latin hypercube design constructed from an SOA$(n, m, s^2, 2+)$ has the better space-filling property that it contains the same number of points in each cell of finer $s \times s^2$ and $s^2 \times s$ grids in all two-dimensional projections.

Example 1. Displayed below are an SOA$(16, 10, 2^2, 2+)$ (left), and a Latin hypercube design constructed from it (right):

$$
\begin{pmatrix}
2 & 2 & 2 & 2 & 2 & 2 & 0 & 0 & 0 & 0 \\
2 & 2 & 0 & 2 & 0 & 0 & 1 & 2 & 2 & 2 \\
2 & 0 & 2 & 0 & 2 & 1 & 2 & 1 & 2 & 2 \\
2 & 0 & 0 & 0 & 0 & 3 & 3 & 3 & 0 & 0 \\
0 & 2 & 2 & 1 & 1 & 2 & 2 & 2 & 1 & 2 \\
0 & 2 & 0 & 1 & 3 & 0 & 3 & 0 & 3 & 0 \\
0 & 0 & 2 & 3 & 1 & 1 & 0 & 3 & 3 & 0 \\
0 & 0 & 0 & 3 & 3 & 3 & 1 & 1 & 1 & 2 \\
1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 & 1 \\
1 & 1 & 3 & 2 & 0 & 0 & 3 & 0 & 0 & 3 \\
1 & 3 & 1 & 0 & 2 & 1 & 0 & 3 & 0 & 3 \\
1 & 3 & 3 & 0 & 0 & 3 & 1 & 1 & 2 & 1 \\
3 & 1 & 1 & 1 & 1 & 2 & 0 & 0 & 3 & 3 \\
3 & 1 & 3 & 1 & 3 & 0 & 1 & 2 & 1 & 1 \\
3 & 3 & 1 & 3 & 1 & 1 & 2 & 1 & 1 & 1 \\
3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3
\end{pmatrix},
\quad
\begin{pmatrix}
8 & 9 & 10 & 11 & 8 & 10 & 0 & 2 & 3 & 3 \\
10 & 10 & 1 & 10 & 1 & 3 & 7 & 8 & 9 & 10 \\
9 & 3 & 11 & 3 & 11 & 7 & 10 & 6 & 11 & 9 \\
11 & 0 & 0 & 1 & 3 & 12 & 14 & 15 & 2 & 2 \\
2 & 8 & 8 & 5 & 4 & 11 & 8 & 10 & 6 & 11 \\
0 & 11 & 2 & 6 & 13 & 2 & 15 & 0 & 15 & 0 \\
1 & 2 & 9 & 12 & 5 & 4 & 2 & 14 & 12 & 1 \\
3 & 1 & 3 & 15 & 15 & 14 & 6 & 7 & 7 & 8 \\
7 & 7 & 4 & 8 & 10 & 9 & 11 & 9 & 10 & 7 \\
5 & 5 & 14 & 9 & 0 & 0 & 12 & 3 & 1 & 12 \\
6 & 15 & 6 & 2 & 9 & 5 & 1 & 12 & 0 & 13 \\
4 & 13 & 13 & 0 & 2 & 15 & 4 & 4 & 8 & 4 \\
13 & 4 & 5 & 7 & 7 & 8 & 3 & 1 & 13 & 14 \\
12 & 6 & 15 & 4 & 14 & 1 & 5 & 11 & 5 & 6 \\
14 & 12 & 7 & 13 & 6 & 6 & 9 & 5 & 4 & 5 \\
15 & 14 & 12 & 14 & 12 & 13 & 13 & 13 & 14 & 15
\end{pmatrix}.
$$

The SOA has the property that when the entries 0, 1, 2, and 3 in any column are replaced by 0, 0, 1, and 1 respectively, in the 16×2 matrix formed by the resulting column and any other original column, all eight ordered pairs of $\{0, 1\}$ and $\{0, 1, 2, 3\}$ appear equally often as rows. The Latin hypercube design on the right is obtained from the SOA by replacing the four occurrences of $i = 0, 1, 2, 3$, with a permutation of $4i, 4i + 1, 4i + 2$, and $4i + 3$, respectively. If we divide all entries by 16, and consider each row as a point, then we obtain 16 points in the 10-dimensional unit cube $[0, 1)^{10}$. This design exhibits

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

the uniformity property that there are two points in each cell of 4×2 and 2×4 grids in all two-dimensional projections. For a design constructed from a two-level orthogonal array of strength two, only stratification in 2×2 grids is guaranteed.

He, Cheng, and Tang (2018) gave the following result, which provides a complete characterization of SOAs of strength 2+, and also shows how they can be constructed from ordinary orthogonal arrays.

Lemma 1. *An SOA$(n, m, s^2, 2+)$, say D, exists if and only if there exist two arrays A and B, where $A = (a_1, \ldots, a_m)$ is an OA$(n, m, s, 2)$ and $B = (b_1, \ldots, b_m)$ is an OA$(n, m, s, 1)$, such that (a_j, a_k, b_k) is an orthogonal array of strength three for any $j \neq k$. The three arrays are linked through $D = sA + B$.*

Theorem 1 of He, Cheng, and Tang (2018) shows how Lemma 1 can be applied to two-level regular designs. As usual, we use $C = (c_{ij})_{n\times m}$, where $c_{ij} = \pm 1$, to represent a two-level factorial design of n runs for m factors. A regular saturated design S of 2^k runs for $2^k - 1$ factors can be obtained by first writing down a full factorial for k factors, and then adding all possible interaction columns. Then each regular 2^{m-p} design C with $p = m - k$ consists of m columns of S. The set of columns not in C, denoted by $\overline{C} = S \backslash C$, is called the complementary design of C. Being regular, S has the property that $ab \in S$ for any $a, b \in S, a \neq b$, where ab is the interaction column of a and b. If C is SOS, then the $2^k - 1$ degrees of freedom of $S = C \cup \overline{C}$ correspond to the main effects and a set of $2^k - 1 - m$ 2fi's of the m factors in C. (Our use of the union symbol technically corresponds to a matrix augmentation; the above $C \cup \overline{C}$ represents a matrix obtained by combining the column vectors of C with those of $\overline{C}$.) This gives a very simple description of C: each $d \in \overline{C}$ can be expressed as $d = ab$ for some $a, b \in C$. Let $\overline{C}$ be $(a_1, \ldots, a_{m'})$, where $m' = 2^k - 1 - m$ and $a_i = b_i c_i$, with $b_i, c_i \in C$, for all $i = 1, \ldots, m'$. As shown in the proof of Theorem 1 in He, Cheng, and Tang (2018), this implies that (a_j, a_k, b_k) is an orthogonal array of strength three for any $j \neq k$. Thus, Lemma 1 is applicable. Following this lemma, we can construct an SOA$(n, m', 2^2, 2+)$ by taking $A = \overline{C} = (a_1, \ldots, a_{m'})$ and $B = (b_1, \ldots, b_{m'})$. Then $D = A + B/2 + 3/2$ is a desired SOA$(n, m', 2^2, 2+)$. Note that because, for two-level designs, the two levels in Lemma 1 are represented by 0 and 1, to apply Lemma 1, we first need to transform -1 and 1 to 0 and 1, respectively. Thus, instead of $D = 2A + B$ as stated in Lemma 1, we should use $D = A + B/2 + 3/2$ here.

Example 2. Let (x_1, x_2, x_3, x_4) be a full 2^4 factorial in 16 runs. Then $C = (x_1, x_2, x_3, x_4, x_1x_2x_3x_4)$ is an SOS design. Note that this design has the defining relation $I = 12345$. It has resolution V, with the five main effects and ten 2fis distributed in the 15 alias sets, and hence is SOS. From the previous discussion, to construct an SOA using Lemma 1, we can let $A = \overline{C} = (x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1x_2x_3, x_1x_2x_4, x_1x_3x_4, x_2x_3x_4)$, and choose an appropriate $B = (b_1, \ldots, b_{10})$. Denote the ith column of A by a_i; then for all $i = 1, \ldots, 10$, any b_i such that both b_i and a_ib_i are columns of C will do. One choice is $B = (x_1, x_1, x_1, x_2, x_2, x_3, x_4, x_3, x_2, x_1)$. Then $D = A+B/2+3/2$ is the SOA$(16, 10, 4, 2+)$ as shown in Example 1.

3. Minimal SOS Designs and Results on Regular Factorials

3.1 SOS designs and their minimality

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Let C be an SOS design. Obviously, adding a factor to C still gives an SOS design. When a factor is deleted from C, the resulting design can be SOS, but is not necessarily so. If the design obtained by deleting one factor from C is still SOS, then we can continue the process of deleting a factor from the current SOS design until no factor can be deleted while maintaining an SOS design. At the end, we must obtain an SOS design such that if any factor is removed, the resulting design is no longer SOS.

Definition 2. *An SOS design is said to be minimal if the design resulting from deleting any factor is no longer SOS.*

Let (x_1, x_2, x_3) be a full 2^3 factorial in eight runs. Then $C = (x_1, x_2, x_3, x_1x_2x_3)$ is a minimal SOS design. The SOS design of 16 runs for five factors in Example 2 is also minimal.

The discussion prior to Definition 2 also explains why minimal SOS designs are useful, which we summarize in a lemma.

Lemma 2. *Any SOS design is either minimal or can be obtained by adding factors to a minimal SOS design.*

Lemma 2 states that all SOS designs can be constructed if all minimal SOS designs are available. Therefore, studies on SOS designs can be focused on minimal SOS designs. Furthermore, by Theorem 1 of He, Cheng, and Tang (2018), using a regular SOS design of n runs for m factors, one can construct an n-run SOA of strength 2+ for $n - 1 - m$ factors. In order to construct SOAs with more factors, we need SOS designs with fewer

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

factors. Thus, SOS designs with fewer factors are of interest.

The next two subsections are devoted to regular minimal SOS designs. We examine nonregular SOS designs in Section 4.

3.2 Characterization and construction of regular minimal SOS designs

A main effect or 2fi is said to be clear if it is not aliased with any other main effect or 2fi. Under the reasonable assumption that interactions of order three or higher are negligible, a clear effect is estimable regardless of other effects. For a detailed discussion on clear effects, refer to Cheng (2014, Chapter 10). The next result provides a complete characterization of a regular SOS design being minimal through clear effects.

Theorem 1. *Let* $C = (c_1, \ldots, c_m)$ *be a regular SOS design, with* c_j *denoting its* j*th column. Then* C *is a minimal SOS design if and only if, for any* $i = 1, \ldots, m$*, at least one of the following* m *effects is clear: the main effect* c_i *and all 2fi's* $c_i c_j$ *with* $j \neq i$*.*

Theorem 1 is obvious, and so not really in need of a proof. It simply recognizes the fact that the design given by deleting column c_i from C remains SOS if and only if none of c_i and $c_i c_j$, with $j \neq i$, are clear. Theorem 1 may be mathematically simple, but it is a very useful result as will be seen throughout the paper.

Corollary 1. *(i) If an SOS design has resolution IV or higher, then it must be minimal.*
(ii) A minimal SOS design of resolution III must have at least one clear 2fi.
(iii) For a minimal SOS design of $n = 2^k$ *runs and* m *factors, we must have that* $m \leq n/2$*.*

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Proof. Part (i) of Corollary 1 is true because all main effects are clear in a resolution IV or higher design. Part (ii) follows because some main effects cannot be clear in a resolution III design, meaning that some 2fi's have to be clear, from Theorem 1. Parts (i) and (ii) state that a minimal SOS design is either of resolution IV or higher, or has some clear 2fi's, both of which imply that $m \leq n/2$ (Cheng 2014, Corollary 9.6 and Theorem 10.7). The first result follows from Rao's bound, and the second was originally obtained by Chen and Hedayat (1998). This proves Corollary 1(iii). □

He, Cheng, and Tang (2018) presented four constructions of regular SOS designs. Using Theorem 1, we show that these SOS designs are actually minimal. Again, let S be the saturated design based on k independent factors, which we denote by $a_1, \ldots, a_{k_1}, b_1, \ldots, b_{k_2}$, where $k_1 \geq 2$, $k_2 \geq 2$, and $k_1 + k_2 = k$. Consider two subsets P and Q of S, where P consists of $a_1, \ldots, a_{k_1}$ and their interaction columns, and Q consists of $b_1, \ldots, b_{k_2}$ and their interaction columns. All four constructions build SOS designs using P and Q.

Construction 1: $C_1 = P \cup Q$.

Construction 2: $C_2 = (P \setminus \{a_1\}) \cup (Q \setminus \{b_1\}) \cup \{a_1 b_1\}$.

Construction 3: $C_3 = (P \setminus \{a_1\}) \cup (a_1 Q)$.

Construction 4: $C_4 = (b_1 P) \cup (a_1 Q \setminus \{a_1 b_1\})$.

Construction 1 gives a design with $n = 2^k$ runs and $2^{k_1} + 2^{k_2} - 2$ factors. Constructions 2–4 all produce designs with the same run size, but with $2^{k_1} + 2^{k_2} - 3$ factors, where $k_1 \geq 2$, $k_2 \geq 2$, and $k_1 + k_2 = k \geq 4$.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Theorem 2. *Designs C_1, C_2, C_3, and C_4 given by Constructions 1–4 are all minimal SOS designs.*

Proof. That C_1, C_2, C_3, and C_4 are all SOS is established in He, Cheng, and Tang (2018). Design C_4 has resolution IV, and is thus minimal, by Corollary 1(i). For design C_1, it is obvious that the 2fi pq is clear for any $p \in P$ and any $q \in Q$, implying that C_1 is minimal, by Theorem 1. Design C_2 is minimal because pq, a_1b_1p, and a_1b_1q are all clear for any $p \in P \setminus \{a_1\}$ and any $q \in Q \setminus \{b_1\}$. Design C_3 is minimal because pq is clear for any $p \in P \setminus \{a_1\}$ and any $q \in a_1Q$. $\square$

By Corollary 1, all SOS designs of resolution IV are minimal. On the other hand, as noted in Section 2, maximal designs of resolution IV are SOS. Therefore, we have the following simple result.

Corollary 2. All maximal resolution IV designs are minimal SOS designs.

Chen and Cheng (2006) examined the structures and constructions of maximal designs with $n/4+1$ or more factors. In contrast, the minimal SOS designs given by Constructions 1–4 all have fewer than $n/4+1$ factors unless $k_1 = 2$ or $k_2 = 2$. Finally, note that Constructions 1 and 4 were also given in Tang, Ma, Ingram, and Wang (2002) in their studies of clear 2fi's.

3.3 Imports from projective geometry and coding theory

It is well known that constructing a regular 2^{m-p} design amounts to choosing m points, one point for each factor, from the $2^{m-p}-1$ points in an $(m-p-1)$-dimensional projective

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

geometry based on the Galois field $GF(2) = \{0, 1\}$. Each line in this geometry contains three points. Then a 2fi can be identified with the third point on the line determined by the points corresponding to the two factors. A set A of points is called a 1-saturating set if every point in the complement of A is on the line determined by a certain pair of points in A. Thus, it is clear that regular SOS designs are equivalent to 1-saturating sets, and regular minimal SOS designs are equivalent to minimal 1-saturating sets. Regular SOS designs also have a coding theory connection, because the dual codes of linear codes with covering radius 2 are equivalent to 1-saturating sets. In this subsection, we import some results from projective geometry and coding theory.

Davydov, Marcugini, and Pambianco (2006) presented several construction methods for minimal 1-saturating sets. Their Constructions A and B are equivalent to our Constructions 1 and 2, respectively, in Section 3.2. In design language, we present one of their other methods because it gives SOS designs with fewer factors than our Constructions 1–4 do. As commented in the paragraph following Lemma 2, SOS designs with fewer factors are of interest because they result in SOAs with more factors.

Recall that our Construction 1 gives a minimal SOS design with $2^{k_1} + 2^{k_2} - 2$ factors, and Constructions 2–4 produce minimal SOS designs with $2^{k_1} + 2^{k_2} - 3$ factors, where $k_1 \geq 2$, $k_2 \geq 2$, and $k_1 + k_2 = k \geq 4$. However, if $k \geq 7$, a minimal SOS design with $2^{k_1} + 2^{k_2} - 4$ factors can be constructed for $k_1 \geq 3$, $k_2 \geq 4$, and $k_1 + k_2 = k$. This is Construction C of Davydov, Marcugini, and Pambianco (2006), which is presented below.

Let S, P, and Q be defined as in Section 3.2. Let $p_1p_2p_3 = q_1q_2q_3 = I$ be two defining words of length three, where $p_1, p_2, p_3 \in P$, $q_1, q_2, q_3 \in Q$, and I is the all-ones column. Take any $q_4 \in Q \setminus \{q_1, q_2, q_3\}$. Consider

$$C_5 = (P \setminus \{p_1, p_2, p_3\}) \cup (Q \setminus \{q_1, q_2, q_3, q_4\}) \cup \{p_1q_3, p_2q_3, p_3q_1, p_3q_2, p_3q_4\}.$$

Lemma 3. (Davydov, Marcugini, and Pambianco 2006) *Design C_5 is a minimal SOS design of $n = 2^k$ runs with $2^{k_1} + 2^{k_2} - 4$ factors, where $k_1 \geq 3$, $k_2 \geq 4$, and $k = k_1 + k_2 \geq 7$.*

However, for $k \geq 7$, SOS designs with even fewer factors can be constructed, as given in Theorems 1 and 2 of Gabidulin, Davydov, and Tombak (1991) in terms of duals of linear codes with covering radius 2. We summarize their results in a lemma.

Lemma 4. (Gabidulin, Davydov, and Tombak 1991) *For $k \geq 7$, an SOS design of $n = 2^k$ runs for m factors can be constructed, where*

$$m = \begin{cases} 5 \times 2^{w-2} - 1 & \text{if } k = 2w - 1, \\ 7 \times 2^{w-2} - 2 & \text{if } k = 2w. \end{cases}$$

To the best of our knowledge, whether or not the SOS designs given in Lemma 4 are minimal has not been established in the literature on coding theory and projective geometry. Here, we provide an affirmative answer to the question.

Proposition 1. *The SOS designs in Lemma 4 are minimal.*

The proof is rather lengthy and, thus, is given in the Appendix.

Example 3. For $k = 7$, and thus $w = 4$, the construction in Lemma 4 gives an SOS design with $n = 2^k = 128$ runs for $m = 5 \times 2^{w-2} - 1 = 19$ factors. The design matrix is

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

given by taking all linear combinations of the rows of

$$B = \begin{bmatrix} 000 & 1111 & 1111 & 1111 & 1111 \\ 011 & 0011 & 0011 & 0011 & 0000 \\ 101 & 0101 & 0101 & 0101 & 0000 \\ 000 & 0011 & 0101 & 0110 & 0011 \\ 000 & 0110 & 0011 & 0101 & 0101 \\ 000 & 0000 & 0000 & 1111 & 1111 \\ 000 & 0000 & 1111 & 0000 & 1111 \end{bmatrix},$$

with all calculations within $GF(2) = \{0, 1\}$. The resulting design can be converted to a version with familiar levels ± 1 by changing 0 to -1. According to Proposition 1, this SOS design is minimal. It is of resolution III, having one word of length three, which is given by the first three columns. In comparison, for $k = 7$ and $n = 128$, Construction 1 gives a minimal SOS design with 22 factors, Constructions 2–4 generate minimal SOS designs with 21 factors, and Lemma 3 gives a minimal SOS design with 20 factors. For $k = 8$, the construction in Lemma 4 gives an SOS design with 256 runs for 26 factors, which is also minimal, by Proposition 1.

Let m_k be the largest m for which an SOA$(2^k, m, 4, 2+)$ based on regular designs can exist. Using Constructions 2–4, He, Cheng, and Tang (2018) obtained a general lower bound on m_k. Lemma 4 offers an improvement for $k \geq 7$.

Proposition 2. *We have that, for $k \geq 7$,*

$$m_k \geq \begin{cases} 2^k - 5 \times 2^{w-2} & \text{if } k = 2w - 1, \\ 2^k - 7 \times 2^{w-2} + 1 & \text{if } k = 2w. \end{cases}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Davydov, Marcugini, and Pambianco (2006) provide a complete enumeration of all minimal 1-saturating sets in small geometries. Thus, their Table 1 classifies all regular minimal SOS designs for up to 64 runs and all regular minimal SOS designs with $m \leq 20$ for 128 runs. We give a summary of their results in Table 1 for the benefit of design researchers.

Table 1. Classification of all regular minimal SOS designs for $n \leq 64$ *and* $n = 128$ *with* $m \leq 20$.

n	m	III	IV	total
8	4	1	1	2
16	5	0	1	1
16	6	1	0	1
16	8	1	1	2
32	9	1	1	2
32	10	6	1	7
32	11	1	0	1
32	16	1	1	2
64	13	7	1	8
64	14	19	0	19
64	15	14	0	14
64	16	16	0	16
64	17	48	5	53
64	18	108	1	109
64	20	1	1	2
64	32	1	1	2
128	19	5	0	5
128	20	36	0	36

For given n and m, the last column of Table 1 gives the number of all regular minimal SOS designs; the third and fourth columns give the numbers of minimal SOS designs of resolution III and IV, respectively. For example, with $n = 64$ runs and $m = 13$ factors, there are exactly seven minimal SOS designs of resolution III, one minimal SOS design of

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

resolution IV, and eight minimal SOS designs in total.

Using the search results in Table 2 of Davydov, Marcugini, and Pambianco (2006), we obtain lower and upper bounds on m'_k, the size of the smallest regular minimal SOS design for $7 \le k \le 10$. These bounds on m'_k are then used to obtain upper and lower bounds on m_k, the greatest m such that an SOA$(2^k, m, 4, 2+)$ based on regular designs exists, because $m'_k + m_k = 2^k - 1$. Our Table 2 updates and expands Table 1 of He, Cheng, and Tang (2018). For completeness, we include information on both m'_k and m_k in Table 2. For $4 \le k \le 7$, the m'_k and m_k values are exact. For $8 \le k \le 10$, Table 2 gives the best known lower and upper bounds (e.g., $25 \le m'_8 \le 26$ and $229 \le m_8 \le 230$).

Table 2. The smallest number m'_k of factors for a regular minimal SOS design, and the largest number m_k of factors for an SOA of strength 2+ based on regular designs.

k	$n = 2^k$	m'_k	m_k
4	16	5	10
5	32	9	22
6	64	13	50
7	128	19	108
8	256	$[25, 26]$	$[229, 230]$
9	512	$[34, 39]$	$[472, 477]$
10	1024	$[47, 51]$	$[972, 976]$

4. Nonregular Minimal SOS Designs and Their Applications

4.1 Characterization and construction

Orthogonal arrays were introduced in Section 2. Throughout this subsection, the two levels in an OA$(n, m, 2, t)$ are denoted by ± 1, rather than 0 and 1. An OA$(n, m, 2, 2)$ is said

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

to be a nonregular design if it is not regular. For a review of nonregular designs, see Xu, Phoa, and Wong (2009). Prior to Sun and Wu (1993), nonregular designs were referred to as irregular (e.g., Addelman 1961). Hadamard matrices provide rich sources of nonregular designs, although not every nonregular design can be imbedded into a Hadamard matrix; see Sun, Li, and Ye (2008) and Bulutoglu and Kaziska (2009) for details. A main effect or 2fi is called clear in a nonregular design if it is orthogonal to all other main effects and 2fi's (Tang 2006).

Our general discussion on SOS designs and their minimality in Section 3.1 applies to nonregular and regular designs. Theorem 1 in Section 3 gives a complete characterization of a regular SOS design being minimal. A similar result holds for nonregular designs.

Theorem 3. *Let $C = (c_1, \ldots, c_m)$ be an SOS design, regular or nonregular. Then C is minimal if, for any $i = 1, \ldots, m$, at least one of the m effects c_i and $c_i c_j$ with $j \neq i$ is clear.*

Similarly to Theorem 1, Theorem 3 states that the condition of existence of certain clear effects is still sufficient for a nonregular SOS design to be minimal. Unlike Theorem 1, the necessity part cannot hold, in general, for nonregular designs. To illustrate this, consider the following example. There are exactly two nonequivalent OA$(12, 5, 2, 2)$s (Sun, Li, and Ye 2008). Using a computer, we can easily check that one is SOS and the other is not. The one that is SOS must also be minimal because there are not enough degrees of freedom for an OA$(12, 4, 2, 2)$ to be SOS. On the other hand, an OA$(12, 5, 2, 2)$ cannot have a clear effect, because otherwise the run size would be a multiple of 8 (Tang 2006).

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Proposition 1 of Tang (2006) relates to the existence of a clear 2fi; the same argument applies to a clear main effect.

Corollary 3. *If an OA$(n, m, 2, 3)$ is SOS, then it must be minimal.*

This result follows from Theorem 3, because all main effects are clear in an orthogonal array of strength three. Cheng, Mee, and Yee (2008) introduced some constructions of SOS OA$(n, m, 2, 3)$s, which are all minimal according to Corollary 3. These minimal SOS designs have relatively large numbers of factors; for example, their first construction gives $m = n/4 + 1$. When constructing SOAs, it is desirable to obtain minimal SOS designs with fewer factors, which we discuss next.

It turns out that the four constructions in Section 3 can all be adapted to the setting of nonregular designs. Let $H_{n_1} = (1_{n_1}, a_1, \ldots, a_{n_1-1})$ and $H_{n_2} = (1_{n_2}, b_1, \ldots, b_{n_2-1})$ be two Hadamard matrices of orders $n_1 \geq 4$ and $n_2 \geq 4$, respectively, where 1_{n_1} is a column vector of n_1 ones. Let $p_i = a_i \otimes 1_{n_2}$ for $i = 1, \ldots, n_1 - 1$, and $q_j = 1_{n_1} \otimes b_j$ for $j = 1, \ldots, n_2 - 1$. Furthermore, let $P = \{p_1, \ldots, p_{n_1-1}\}$ and $Q = \{q_1, \ldots, q_{n_2-1}\}$. Consider the following constructions:

Construction (i): $C_1 = P \cup Q$.

Construction (ii): $C_2 = (P \setminus \{p_1\}) \cup (Q \setminus \{q_1\}) \cup \{p_1 q_1\}$.

Construction (iii): $C_3 = (P \setminus \{p_1\}) \cup (p_1 Q)$.

Construction (iv): $C_4 = (q_1 P) \cup (p_1 Q \setminus \{p_1 q_1\})$.

All four designs have $n_1 n_2$ runs. Design C_1 has $n_1 + n_2 - 2$ factors, and designs C_2, C_3,

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

and C_4 have $n_1 + n_2 - 3$ factors.

Theorem 4. *Designs C_1, C_2, C_3, and C_4, given above, are minimal SOS designs.*

Proof. For brevity, we give proofs for Constructions (i) and (ii) only. The proofs for Constructions (iii) and (iv) use similar ideas, but are more tedious and complicated.

First, consider design $C_1 = P \cup Q$ from Construction (i). The Hadamard matrix $H_{n_1} \otimes H_{n_2}$ then consists of $1_{n_1 n_2}$, all main effect columns in P, all main effect columns in Q, and all interaction columns pq, where $p \in P$ and $q \in Q$. Therefore the design $C_1 = P \cup Q$ is SOS. Because any 2fi pq with $p \in P$ and $q \in Q$ is obviously clear, design C_1 is minimal, by Theorem 3.

Now, consider design C_2 from Construction (ii). We first decompose the set of $n_1 n_2$ columns in the Hadamard matrix $H_{n_1} \otimes H_{n_2}$ into a union of six disjoint subsets, as given by

$$H_{n_1} \otimes H_{n_2} = R_0 \cup R_1 \cup R_2 \cup R_3 \cup R_4 \cup R_5,$$

where $R_0 = \{1_{n_1 n_2}\}$, $R_1 = P$, $R_2 = Q$, $R_3 = p_1 Q$, $R_4 = q_1 P \setminus \{p_1 q_1\}$, and $R_5 = \{p_i q_j \mid i = 2, \ldots, n_1 - 1; j = 2, \ldots, n_2 - 1\}$. To prove that design $C_2 = (P \setminus \{p_1\}) \cup (Q \setminus \{q_1\}) \cup \{p_1 q_1\}$ is SOS, we need to show that for each R_j, where $j = 1, \ldots, 5$, we can choose a set of linearly independent main effects or 2fi's from design C_2 such that they span the same linear subspace as that spanned by the columns of R_j. Because every column in R_5 is a 2fi of design C_2, the job is done for R_5. Now, consider $R_1 = P$. If we can find a 2fi $p_i p_j$ of design C_2, where $2 \leq i < j \leq n_1 - 1$, such that it is not orthogonal to p_1, then the main effects $p_2, \ldots, p_{n_1 - 1}$ of C_2 plus this 2fi $p_i p_j$ are

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

linearly independent and, thus, span the linear subspace spanned by $R_1 = P$. Such a 2fi must exist; otherwise $p_2, \ldots, p_{n_1-1}, p_1p_2, \ldots, p_1p_{n_1-1}$ are mutually orthogonal within the linear subspace spanned by P, which is a contradiction. The same argument works for $R_2 = Q$. We now turn our attention to $R_3 = p_1Q$. Because the column vectors $p_1q_1, p_1q_1q_2, \ldots, p_1q_1q_{n_2-1}$ are mutually orthogonal and all belong to $L(R_3)$, the linear subspace spanned by the columns of $R_3 = p_1Q$, they therefore span $L(R_3)$. However, p_1q_1 is a main effect of design C_2, and $p_1q_1q_j$ for $j \geq 2$ is a 2fi between factor p_1q_1 and factor q_j of design C_2. This takes care of R_3. The same argument with a minor modification also works for R_4. We have thus proved that design C_2 is SOS. That C_2 is minimal follows from the fact that the 2fi of factor p_i and factor q_j is clear, for all $i \geq 2$ and $j \geq 2$, and that the main effect p_1q_1 is also clear. □

In the next subsection, we examine how to use the designs given by Constructions (i)–(iv) to construct SOAs of strength 2+.

Small orthogonal arrays have been completely enumerated by Sun, Li, and Ye (2008) and Schoen, Eendebak, and Nguyen (2010). Using these results, we conduct a complete search of OA$(n, m, 2, t)$s that are minimal SOS designs for $t = 2$ with $n = 12, 16$, and 20, and for $t = 3$ with $n = 16, 24, 32$, and 40. Table 3 presents a summary of all minimal SOS designs for these parameters. For a given strength t and pair (n, m), the last column of Table 3 gives the number N_{minsos} of OA$(n, m, 2, t)$s that are minimal SOS designs. For comparison, we also include in Table 3 the number N_{all} of all nonisomorphic designs and the number N_{sos} of all SOS designs. For example, there are 474 nonisomorphic

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

OA(20, 7, 2, 2)s, of which 339 arrays are SOS. Among the 339 SOS designs, 22 are minimal. Furthermore, Table 3 shows that there exist exactly three OA(16, 5, 2, 2)s that are minimal SOS designs. Comparing these results with those shown in Table 1, we conclude that one of these minimal SOS designs is regular, and the other two are nonregular.

Table 3. Classification of all OA$(n, m, 2, t)$s that are minimal SOS designs for $t = 2$ with $n = 12, 16$, and 20, and for $t = 3$ with $n = 16, 24, 32$, and 40, where N_{all}, N_{sos}, and N_{minsos} denote the number of all nonisomorphic designs, number of SOS designs, and number of minimal SOS designs, respectively.

t	(n, m)	N_{all}	N_{sos}	N_{minsos}
2	(12, 5)	2	1	1
2	(16, 5)	11	3	3
2	(16, 6)	27	14	2
2	(16, 8)	80	80	2
2	(20, 6)	75	15	15
2	(20, 7)	474	339	22
2	(20, 9)	2477	2466	1
3	(16, 5)	2	1	1
3	(16, 8)	1	1	1
3	(24, 12)	1	1	1
3	(32, 9)	34	6	6
3	(32, 10)	32	1	1
3	(32, 16)	5	5	5
3	(40, 20)	3	3	3

4.2 Applications to strong orthogonal arrays

In this subsection, we consider the problem of constructing SOA$(n, m, 4, 2+)$s. According to Lemma 1, we first need to find two arrays A and B, where $A = (a_1, \ldots, a_m)$ is an OA$(n, m, 2, 2)$ and $B = (b_1, \ldots, b_m)$ is an OA$(n, m, 2, 1)$, such that (a_j, a_k, b_k) has

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

strength three for any $j \neq k$. Then, as noted in the paragraph following Lemma 1, we obtain an SOA$(n, m, 4, 2+)$ using $D = A + B/2 + 3/2$.

He, Cheng, and Tang (2018) examined how to obtain A and B if their columns are selected from a saturated regular design. We now consider obtaining A and B by choosing their columns from a saturated orthogonal array, which can be nonregular.

Theorem 5. *Let S be an OA$(n, n-1, 2, 2)$. To construct an SOA$(n, m, 4, 2+)$ using $D = A + B/2 + 3/2$ with the columns of A and B selected from S, it is necessary and sufficient that, for any column $a \in A$, there exists a column $b \in S \setminus A$ such that ab is orthogonal to all columns in A.*

Theorem 5 extends Theorem 1 of He, Cheng, and Tang (2018), and includes the latter as a special case, because one can easily see that the condition for array A in Theorem 5 is equivalent to $S \setminus A$ being SOS if S is a regular saturated design. We omit the proof for Theorem 5 because it is very similar to that for Theorem 1 of He, Cheng, and Tang (2018).

Theorem 5 is actually constructive. Suppose that $A = (a_1, \ldots, a_m)$ satisfies the required condition in Theorem 5, meaning that, for any a_i, there exists a column in $S \setminus A$, say b_i, such that $a_i b_i$ is orthogonal to all a_j. Then we simply take $B = (b_1, \ldots, b_m)$.

Remark 1. When S is a nonregular saturated design, the condition for array A in Theorem 5 may not be equivalent to $S \setminus A$ being SOS. Further discussion on this issue will be given in Section 5. It is therefore not true that every nonregular SOS design can be used to construct an SOA$(n, m, 4, 2+)$. On the other hand, as shown below, almost all

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

SOS designs given by Constructions (i)–(iv) allow the construction of SOAs.

All minimal SOS designs obtained by Constructions (i)–(iii) can be used to construct SOA$(n, m, 4, 2+)$s. The minimal SOS designs obtained by Construction (iv) can also be used, provided that one of H_{n_1} and H_{n_2} is regular. Details are given as follows. Note that $S = H_{n_1} \otimes H_{n_2} \setminus \{1_{n_1 n_2}\}$.

Construction (i). Let $A_1 = S \setminus C_1$. All columns in A_1 have the form pq, where $p \in P$ and $q \in Q$. For any $a = pq \in A_1$, we take $b = p$.

Construction (ii). Let $A_2 = S \setminus C_2$. For $a = p_i q_j \in A_2$, where $i, j \geq 2$, we take $b = p_i$. For $a = p_1$, take $b = p_2$. For $a = q_1$, take $b = q_2$. For $a = p_1 q_j$, where $j \geq 2$, take $b = p_1 q_1$. For $a = q_1 p_i$, where $i \geq 2$, take $b = p_1 q_1$.

Construction (iii). Let $A_3 = S \setminus C_3$. For $a = p_i q_j$, where $i \geq 2, j \geq 1$, we take $b = p_1 q_j$. For $a = p_1$, take $b = p_2$. For $a = q_j$, where $j \geq 1$, take $b = p_1 q_{j'}$, where $j' \neq j$.

Construction (iv). Assume H_{n_1} is regular. Let $A_4 = S \setminus C_4$. For $a = p_i$, take $b = p_{i'} q_1$, where $i' \neq i$. For $a = q_j$, take $b = p_1 q_{j'}$, where $j' \neq j$. For $a = p_i q_j$, where $i \geq 2$ and $j \geq 2$, take $b = p_{i'} q_1$, where $p_{i'} = p_1 p_i$, which is possible because H_{n_1} is regular.

We can verify routinely that A_1, A_2, A_3, and A_4 satisfy the condition in Theorem 5. The above also shows how to obtain the corresponding B_1, B_2, B_3, and B_4. Using $D = A + B/2 + 3/2$, we can then obtain D_1, D_2, D_3, and D_4. We summarize these developments in the following theorem.

Theorem 6. *Design D_1 is an SOA$(n_1 n_2, m, 4, 2+)$ with $m = n_1 n_2 - n_1 - n_2 + 1$, and designs D_2, D_3, and D_4 are all SOA$(n_1 n_2, m, 4, 2+)$s with $m = n_1 n_2 - n_1 - n_2 + 2$.*

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Example 4. Take $n_1 = 4$ and $n_2 = 12$. Then D_1 is an SOA$(48, 33, 4, 2+)$, and D_2, D_3, and D_4 are SOA$(48, 34, 4, 2+)$s. If we take $n_1 = n_2 = 12$, then D_1 is an SOA$(144, 121, 4, 2+)$, and D_2 and D_3 are both SOA$(144, 122, 4, 2+)$s. Such run sizes cannot be attained using regular SOS designs. Note that D_4 is not available for $n_1 = n_2 = 12$ because H_{n_1} and H_{n_2} are both nonregular.

SOAs constructed from regular and nonregular designs differ in terms of their three-dimensional space-filling properties. The distribution of the design points in the eight cells when projected onto three dimensions and viewed on a $2 \times 2 \times 2$ grid is determined by the three corresponding columns of the array A in Lemma 1. This array of three columns has only two possible structures for regular A, but a lot more for nonregular A.

Example 5. Hall (1961) identified five nonisomorphic Hadamard matrices of order 16, denoted by H_I, H_{II}, H_{III}, H_{IV}, and H_V, respectively, where only H_I is regular. Denote a submatrix consisting of the $j_1, j_2, \ldots, j_m$th columns of H_i by $H_i(j_1, j_2, \ldots, j_m)$, where $i = I, II, III, IV, V$. Then it can be verified that $C_1 = H_I(7, 11, 12, 13, 14, 15)$, $C_2 = H_I(6, 7, 9, 11, 13, 14)$, and $C_3 = H_{II}(1, 2, 3, 7, 11, 15)$ are all SOS. Furthermore, their complementary designs $A_1 = H_I(1, 2, 3, 4, 5, 6, 8, 9, 10)$, $A_2 = H_I(1, 2, 3, 4, 5, 8, 10, 12, 15)$, and $A_3 = H_{II}(4, 5, 6, 8, 9, 10, 12, 13, 14)$ all satisfy the condition of Theorem 5. One can construct SOA$(16, 9, 4, 2+)$s $D_i = A_i + B_i/2 + 3/2$, for $i = 1, 2, 3$, by choosing $B_1 = H_I(12, 12, 12, 11, 11, 11, 7, 7, 7)$, $B_2 = H_I(6, 9, 13, 9, 11, 6, 7, 7, 6)$, and $B_3 = H_{II}(3, 2, 1, 3, 2, 1, 3, 2, 1)$. Because A_1 has seven defining words of length three, for D_1, there are seven three-dimensional projections in which there are points in only four of the eight cells in a $2 \times 2 \times 2$ grid. For

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

D_2, there are six such three-dimensional projections. Because of the nonregular structure of A_3, for all three-dimensional projections of D_3, there is at least one point in each cell. The numbers of four-dimensional projections in which only eight of the sixteen cells in a $2 \times 2 \times 2 \times 2$ grid are occupied are 51, 45, and 9 for D_1, D_2, and D_3, respectively. This comparison shows that D_3, which is constructed from a nonregular SOS design, has better coverage than D_1 and D_2, which are constructed from regular SOS designs.

5. Discussion

This study conducts a comprehensive investigation of SOS designs and their minimality, focusing on their usefulness in constructing strong orthogonal arrays. In both regular and nonregular cases, we establish characterizing results for SOS designs to be minimal, and provide some construction results for minimal SOS designs. In the case of regular designs, results from projective geometry and coding theory allow us to construct SOAs of strength 2+ with more factors (see Proposition 2 and Table 2) than those in He, Cheng, and Tang (2018). The nonregular counterparts of the four constructions in He, Cheng, and Tang (2018) allow us to construct four families of SOAs of strength 2+.

In the case of regular designs, Grynkiewicz and Lev (2010) studied the structures and sizes of large 1-saturating sets. Although we are more concerned with SOS designs with small numbers of factors, one of their results is relevant here. They show that while the largest minimal SOS design has $m = n/2$, the second largest minimal SOS design must have $m = 5n/16$, provided that n is is sufficiently large. From Table 1, we see that this is already true for $n = 64$. The results in Table 2 of Davydov, Marcugini, and Pambianco

(2006) confirm the statement.

An important unresolved problem is whether the constructions in Lemmas 3 and 4 can be adapted to nonregular designs. Because the construction for the design in Lemma 4 relies heavily on the regular structure, it does not seem possible to generalize this to nonregular settings. However, in Lemma 3, as long as both H_{n_1} and H_{n_2} contain a defining word of length three, there is a nonregular counterpart for the construction. The questions then are whether the resulting design is minimal SOS, and whether it can be used to construct an SOA of strength 2+. We leave these questions to future research.

The construction of SOAs in Theorems 5 and 6 raises an intriguing question, which is at least of technical interest: what is the relationship between the condition in Theorem 5 and the property of being SOS? We know that they are equivalent in the case of regular designs. In the nonregular case, the following result sheds some light on the issue.

Lemma 5. *Let C, an $OA(n, m, 2, 2)$, be an orthogonal SOS design, meaning that there exists a set A of $n-1-m$ mutually orthogonal 2fi's that are also orthogonal to the main effects. Then array A satisfies the condition in Theorem 5.*

The proof is straightforward by taking $S = A \cup C$, which is an $\mathrm{OA}(n, n-1, 2, 2)$. Lemma 5 seems to suggest that the condition in Theorem 5 is stronger than being SOS. A proof or a counterexample is worth seeking.

Acknowledgments. The research of Boxin Tang was supported by the Natural Sciences and Engineering Research Council of Canada. The authors thank Richard Groe-

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

newald for computational assistance with Example 3 and two OA$(12, 5, 2, 2)$s in Section 4.1. The authors are most grateful to Professor Eric Schoen for providing them with an OA package for the nonisomorphic designs in Table 3.

Appendix: Proof of Proposition 1

The constructions were given in Theorems 1 and 2 of Gabidulin, Davydov and Tombak (1991). Our notation in this appendix is different from the main body of our paper. Instead, we use the notation in Gabidulin, Davydov and Tombak (1991). This is just to make presentation easier. Let $e_0 = 0, e_1, \ldots, e_B$ be the elements of $GF(2^b)$, where $B = 2^b - 1$. Further let $(e_i)^b$ denote the column vector that is the binary b-bit representation of e_i. Define two matrices E^b and $F^{2b}(e_j)$ as follows

$$E^b = \left[\begin{array}{cccc} (e_0)^b & (e_1)^b & \cdots & (e_B)^b \end{array} \right],$$

$$F^{2b}(e_j) = \left[\begin{array}{cccc} (e_0)^b & (e_1)^b & \cdots & (e_B)^b \\ (e_0)^b & (e_1^{-1}e_j)^b & \cdots & (e_B^{-1}e_j)^b \end{array} \right].$$

Let E_0^b be E^b with the first column deleted. We use $P^b(e_i)$ to denote a matrix with the same column $(e_i)^b$ repeated where the number of repetitions is determined by context.

Lemma A. (Gabidulin, Davydov and Tombak 1991, Theorem 1) Let the parity check matrix of a code be

$$B^{2m-1} = [N \; D \; Q \; M \; G],$$

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

where $2m-1=r\geq 7$ and matrices N, D, Q, M and G are, respectively, given by

$$\begin{bmatrix} 0 \cdots 0 \\ E_0^{m-2} \\ \\ P^m(e_0) \\ \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ F^{2(m-2)}(w_1) \\ 0 \cdots 0 \\ 0 \cdots 0 \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ F^{2(m-2)}(w_2) \\ 0 \cdots 0 \\ 1 \cdots 1 \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ F^{2(m-2)}(w_3) \\ 1 \cdots 1 \\ 0 \cdots 0 \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ P^{m-2}(e_0) \\ E^{m-2} \\ 1 \cdots 1 \\ 1 \cdots 1 \end{bmatrix},$$

where $w_1, w_2, w_3 \in GF(2^{m-2})$; $w_1, w_2 \neq 0$, $w_1 \neq w_2$, $w_1 + w_2 = w_3$. Then this code has covering radius $R=2$.

Let C be the design generated by taking linear combinations of the rows of B^{2m-1} with coefficients from $GF(2)=\{0,1\}$. According to Lemma A, design C is SOS as the code has covering radius $R=2$.

Proposition 1a. *The array C, generated by B^{2m-1}, is a minimal SOS design.*

Proof. We already know that C is SOS. To show its minimality, we use our Theorem 1 in Section 3. As design C has entries from $GF(2)=\{0,1\}$, a main effect column c_i is clear if $c_i \neq c_j$ and $c_i \neq c_j + c_k$ for any other columns c_j and c_k, and a 2fi c_ic_j is clear if $c_i + c_j \neq c_k$ and $c_i + c_j \neq c_k + c_l$ for any other columns c_k and c_l. Equivalently, one can verify these properties for the columns of B^{2m-1}. We will prove that each column in $[D, Q, M, G]$ is clear, and each column in N has a clear 2fi.

Suppose that a column $\vec{d}$ in D is not clear. Then there must exist two other columns x and y in B^{2m-1} such that $\vec{d} = x + y$, where $x \neq \vec{d}$ and $y \neq \vec{d}$. By examining the first row and last two rows of B^{2m-1}, we see that $\vec{d} = x + y$ is possible only if x is a column in N and y a column in D. Suppose that neither $\vec{d}$ nor y is the first column of D. Since $\vec{d}$,

x and y have form

$$\vec{d} = \begin{bmatrix} 1 \\ (e_i)^{m-2} \\ (e_i^{-1}w_1)^{m-2} \\ 0 \\ 0 \end{bmatrix}, \quad x = \begin{bmatrix} 0 \\ (e_j)^{m-2} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad y = \begin{bmatrix} 1 \\ (e_k)^{m-2} \\ (e_k^{-1}w_1)^{m-2} \\ 0 \\ 0 \end{bmatrix},$$

it is impossible to have $\vec{d} = x + y$ unless $(e_i^{-1}w_1)^{m-2} = (e_k^{-1}w_1)^{m-2}$. But $(e_i^{-1}w_1)^{m-2} = (e_k^{-1}w_1)^{m-2}$ implies $e_i = e_k$, which leads to $\vec{d} = y$. This is a contradiction. It is also obvious that $\vec{d} = x + y$ cannot hold if one of $\vec{d}$ and y is the first column of D. Therefore, $\vec{d}$ must be clear. Similarly, one can prove that any column in Q, M or G is also clear.

For any column $\vec{n}$ of N, we will show that $\vec{n} + \vec{d}$ is clear where $\vec{d}$ is any except the first column of D. Since $\vec{d}$ is clear, we only need to prove $\vec{n} + \vec{d} \neq x + y$ for any other two columns x, y in B^{2m-1}. Again by examining the first row and the last two rows of B^{2m-1}, the only possible scenario for $\vec{n} + \vec{d} = x + y$ is that x is a column of N and y a column of D. An argument very similar to that in the last paragraph shows that if $\vec{n} + \vec{d} = x + y$, then we must have $\vec{n} = x$ and $\vec{d} = y$. This shows that $\vec{n} + \vec{d}$ is clear, and thus each column in N has a clear 2fi. The proof is completed. □

Let D_1 and B_1^{2m-1} be the matrices of D and B^{2m-1} with their first column $(1, 0, \ldots, 0)^T$ deleted, respectively.

Lemma B. (Gabidulin, Davydov and Tombak 1991, Theorem 2) *Let the parity check matrix of a code be*

$$T^{2m} = [Z \ Y],$$

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

where $2m = r \geq 8$ and

$$Z = \begin{bmatrix} 0 \cdots 0 \\ B_1^{2m-1} \end{bmatrix}, \quad Y = \begin{bmatrix} 1 \cdots 1 \\ E^{m-1} \\ P^m(e_i) \end{bmatrix},$$

where $i \in \{0, \ldots, 2^m - 1\}$. Then the code has covering radius $R = 2$.

Let N^*, D_1^*, Q^*, M^* and G^* denote the matrices of N, D_1, Q, M and G with an added head row $(0, \ldots, 0)$. Partition Y into $Y = [Y_1, Y_2]$ such that the second row of Y_1 is the all-zeros row, and the second row of Y_2 is the all-ones row. Then T^{2m} can be partitioned into seven submatrices as $T^{2m} = [N^*, D_1^*, Q^*, M^*, G^*, Y_1, Y_2]$ with $N^*, D_1^*, Q^*, M^*, G^*, Y_1, Y_2$ given, respectively, by

$$\begin{bmatrix} 0 \cdots 0 \\ 0 \cdots 0 \\ E_0^{m-2} \\ \\ P^m(e_0) \end{bmatrix}, \begin{bmatrix} 0 \cdots 0 \\ 1 \cdots 1 \\ F_0^{2(m-2)}(w_1) \\ 0 \cdots 0 \\ 0 \cdots 0 \end{bmatrix}, \begin{bmatrix} 0 \cdots 0 \\ 1 \cdots 1 \\ F^{2(m-2)}(w_2) \\ 0 \cdots 0 \\ 1 \cdots 1 \end{bmatrix}, \begin{bmatrix} 0 \cdots 0 \\ 1 \cdots 1 \\ F^{2(m-2)}(w_3) \\ 1 \cdots 1 \\ 0 \cdots 0 \end{bmatrix},$$

$$\begin{bmatrix} 0 \cdots 0 \\ 1 \cdots 1 \\ P^{m-2}(e_0) \\ E^{m-2} \\ 1 \cdots 1 \\ 1 \cdots 1 \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ 0 \cdots 0 \\ E^{m-2} \\ \\ P^m(e_i) \end{bmatrix}, \begin{bmatrix} 1 \cdots 1 \\ 1 \cdots 1 \\ E^{m-2} \\ \\ P^m(e_i) \end{bmatrix},$$

where $F_0^{2(m-2)}(w_1)$ is obtained from $F^{2(m-2)}(w_1)$ by deleting its first column which consists of all zeros.

Proposition 1b. *The array C', generated by T^{2m}, is a minimal SOS design.*

Proof. We are going to prove that each column in $[D_1^*, Q^*, M^*, G^*]$ is clear and each column in $[N^*, Y_1, Y_2]$ has a clear 2fi.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Since any column of D is clear in B^{2m-1} (from the proof of Proposition 1a), we have that if a column $\vec{d}$ in D_1^* has $\vec{d} = x_1 + x_2$ for two other columns x_1, x_2 from T^{2m}, then at least one of x_1 and x_2 must come from Y. Examining the first two rows of T^{2m}, we see that the only possible case is that x_1 is from Y_1 and x_2 is from Y_2. Now if we take a look at the $(m+1)$th to $(2m-2)$th rows of D_1^*, Y_1 and Y_2, we see that it is impossible for $\vec{d} = x_1 + x_2$ to hold. This is because the $(m+1)$th to $(2m-2)$th entries of $x_1 + x_2$ are all zeros whereas the corresponding entries of $\vec{d}$ are those of $(e_i^{-1} w_1)^{m-2}$, which cannot be all zeros. We have thus established that any column $\vec{d}$ in D_1^* is clear. According to the entries in the first and the last two rows of T^{2m}, one can prove that each column in $[Q^*, M^*, G^*]$ is also clear.

In a manner almost identical to that of showing a column in D_1^* is clear as given above, we can show that the 2fi $\vec{n} + \vec{d}$ is clear for any column $\vec{n}$ from N^* and any column $\vec{d}$ from D_1^*.

For any column $\vec{y}$ from Y_1 or Y_2, and any column $\vec{d}$ from D_1^*, we can show that $\vec{y} + \vec{d}$ is clear. The arguments are, though a bit more involved, also very similar to the above. We omit the details. This completes the proof. $\square$

References

Addelman, S. (1961). Irregular fractions of the 2^n factorial experiments. *Technometrics*, **3**, 479-496.

Block, R.M. and Mee, R.W. (2003). Second order saturated resolution IV designs. *J. Stat. Theory Appl.*, **2**, 96-112.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Bruen, A., Haddad, L. and Wehlau, L. (1998). Binary codes and caps. *J. Combin. Designs*, **6**, 275-284.

Bulutoglu, D.A. and Kaziska, D.M. (2009). A counterexample to Beder's conjectures about Hadamard matrices. *Journal of Statistical Planning and Inference*, **139**, 3381-3383.

Chen, H. H. and Cheng, C.S. (2004). Aberration, estimation capacity and estimation index. *Statistica Sinica*, **14**, 203-215.

Chen, H. H. and Cheng, C.S. (2006). Doubling and projection: a method of constructing two-level designs of resolution IV. *The Annals of Statistics*, **34**, 546-558.

Chen, H. and Hedayat, A.S. (1998). 2^{n-m} designs with resolution III or IV containing clear two-factor interactions. *Journal of Statistical Planning and Inference*, **75**, 147-158.

Cheng, C.S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments.* CRC Press.

Cheng, C.S., Mee, R.W., and Yee, O. (2008). Second order saturated orthogonal arrays of strength three. *Statistica Sinica*, **18**, 105-119.

Davydov, A.A., Marcugini, S. and Pambianco, F. (2006). Minimal 1-saturating sets and complete caps in binary projective spaces. *Journal of Combinatorial Theory, Series A*, **113**, 647-663.

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans.* New York: Wiley.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Gabidulin, E.M., Davydov, A.A. and Tombak, L.M. (1991). Linear codes with covering radius 2 and other new covering codes. *IEEE Transactions on Information Theory*, **37**, 219-224.

Grynkiewicz, D.J. and Lev, V.F. (2010). 1-saturating sets, caps, and doubling-critical sets in binary spaces. *SIAM Journal on Discrete Mathematics*, **24**, 169-190.

Hall, M, J. (1961). Hadamard matrix of order 16. Jet Propulsion Laboratory, Research Summary, **1**, 21-36.

He, Y., Cheng, C.S. and Tang, B. (2018). Strong orthogonal arrays of strength two plus. *Annals of Statistics*, **46**, 457-468.

He, Y. and Tang, B. (2013). Strong orthogonal arrays and associated Latin hypercubes for computer experiments. *Biometrika*, **100**, 254-260.

Hedayat, A.S., Sloane, N.J.A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications.* Springer, New York.

Liu, H. and Liu, M.Q. (2015). Column-orthogonal strong orthogonal arrays and sliced strong orthogonal arrays. *Statistica Sinica*, **25**, 1713-1734.

McKay, M.D., Beckman, R.J. and Conover, W.J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, **21**, 239-245.

Mee, R. (2009). *A Comprehensive Guide to Factorial Two-Level Experimentation.* Springer Science & Business Media.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Owen, A.B. (1992). Orthogonal arrays for computer experiments, integration, and visualization. *Statistica Sinica*, **2**, 439-452.

Santner, T.J., Williams, B.J. and Notz, W.I. (2003). *The Design and Analysis of Computer Experiments.* New York: Springer.

Schoen, E.D., Eendebak, P.T. and Nguyen, M.V.M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs*, **18**, 123-140.

Sun, D.X., Li, W., and Ye, K.Q. (2008). Algorithmic construction of catalogs of non-isomorphic two-level orthogonal designs for economic run sizes. *Statistics and Applications*, **6**, 141-155.

Sun, D.X. and Wu, C.F.J. (1993). Statistical properties of Hadamard matrices of order 16. In Kuo, W., editor, *Quality Through Engineering Design*, pages 169-179. New York: Elsevier.

Tang, B. (1993). Orthogonal array-based Latin hypercubes. *Journal of the American Statistical Association*, **88**, 1392-1397.

Tang, B. (2006). Orthogonal arrays robust to nonnegligible two-factor interactions. *Biometrika*, **93**, 137-146.

Tang, B., Ma, F., Ingram, D. and Wang, H. (2002). Bounds on the maximum number of clear two-factor interactions for 2^{m-p} designs of resolution III and IV. *Canadian Journal of Statistics*, **30**, 127-136.

Statistica Sinica: Preprint
doi:10.5705/ss.202018.0267

Xu, H., Phoa, F.K. and Wong, W.K. (2009). Recent developments in nonregular fractional factorial designs. *Statistics Surveys*, **3**, 18-46.