Berliner Hochschule für Technik

Fachbereich II
Mathematik - Physik - Chemie

01/2021

Ulrike Grömping

A unified implementation of stratum (aka strong) orthogonal arrays

Eine einheitliche Implementierung von Stratum (alias Strong) Orthogonal Arrays (englischsprachig)

Reports in Mathematics, Physics and Chemistry
Berichte aus der Mathematik, Physik und Chemie

ISSN (print): 2190-3913

Reports in Mathematics, Physics and Chemistry

Berichte aus der Mathematik, Physik und Chemie

The reports are freely available via the Internet:
http://www1.bht-berlin.de/FB_II/reports/welcome.htm

01/2021, October 2021

A unified implementation of stratum (aka strong) orthogonal arrays
Eine einheitliche Implementierung von Stratum (alias Strong) Orthogonal Arrays (englischsprachig)

Update history
Feb 17, 2022: A few small clarifications/corrections on pp. 11,15,17,18,21,22,27; noteworthy: corrected correlation information on p.15, replaced S with A in fourth bullet on p.22
Feb 25, 2022: correct formula for upper bound for m_k for HCT is much more complicated than previously alleged and has been omitted, which leads to rewording in the introduction (substantially -> slightly) and in Section 5.2.1; a few corrections/clarifications to Table 5

Editorial notice / Impressum

Published by / Herausgeber:
Fachbereich II
Beuth Hochschule für Technik Berlin
Luxemburger Str. 10
D-13353 Berlin
Internet: https://www.bht-berlin.de/ii

ISSN (print): 2190-3913

A unified implementation of stratum (aka strong) orthogonal arrays

Ulrike Grömping

Berliner Hochschule für Technik

Abstract

In recent years, several authors proposed constructions of so-called "strong orthogonal arrays" (SOAs). The approaches and notations taken in the different proposals vary widely. This paper sets out to explain SOAs and their constructions, taking a unified notation and basing everything on a simple set of equations. Besides providing a unified overview, some constructions are improved, e.g. by guaranteeing orthogonal columns where the original constructions pay no attention to column orthogonality. As an aside, it is argued that "stratum orthogonal arrays" is a better long version for the acronym SOAs.

1 Introduction

He and Tang (2013, 2014) introduced so-called "Strong Orthogonal Arrays" (SOAs) and proposed their use for the construction of Latin Hypercube Designs (LHDs) for computer experiments. Subsequent authors built on their work and introduced various variants, among them Liu and Liu (2015), He, Cheng and Tang (2018), Zhou and Tang (2019), Shi and Tang (2020), and most recently Li, Liu and Yang (2021). This author is attracted by the concept, but considers its label as misleading: by a stretch of concept, SOAs can be seen as orthogonal arrays, but as very *weak* ones (low OA strength) only. This paper will explicitly distinguish between OA strength and SOA strength, because these are related but different concepts. This improves the clarity of presentation and supports an understanding of the use of OAs in the role of SOAs. In order to avoid confusing use of the adjective "strong", this paper will use the acronym SOAs, but connect it with the long form "Stratum Orthogonal Arrays". The rationale behind that expression: when collapsed to strata, SOAs become strong(er) orthogonal arrays.

Arrays with many levels per column, such as LHDs, are primarily used for computer experiments with quantitative variables. Their most important property is their "space-filling" behavior, which can be measured in a variety of ways. Many constructions are based on orthogonal arrays, e.g. Tang's (1993) proposal of expanding the levels of an OA, Ye's (1998) proposal to obtain an LHD with orthogonal columns from a construction based on regular fractional factorial 2-level columns, or Xiao and Xu's (2018) maximin distance level expansion (MDLE) arrays. SOAs provide another and very structured way of expanding the levels of OAs. Key benefits of SOAs are their more refined and controlled low-dimensional stratification behavior, and the limited effort with which their space-filling properties can be improved during construction by level permutation, using an approach proposed by Weng (2014).

The stratification properties of a strength 3 SOA are illustrated in Figure 1, based on a small SOA with three 27-level columns in 27 runs (this particular SOA is an LHD). The three plots in the top row show that there are 27 strata in 3D: separated by three coarsened levels of X_3 (first nine, middle nine, and last nine levels), coarsened levels of X_1 and X_2 produce nine strata with one element each for each X_3 stratum. The bottom row shows the two ways to stratify the $X_1 \times X_2$ space into 27 strata, and the two 1D stratifications for X_1 and X_2 (each row contains exactly one element, each column contains exactly one element).

This paper provides a unifying overview of diverse SOA constructions that have been proposed in recent years. It has been written with a clear focus on practically feasible constructions. An R implementation of the constructions presented in the paper is currently underway. All constructions are presented based on simple matrix equations. The unifying view on a diverse set of recent articles revealed a few opportunities for improvements:

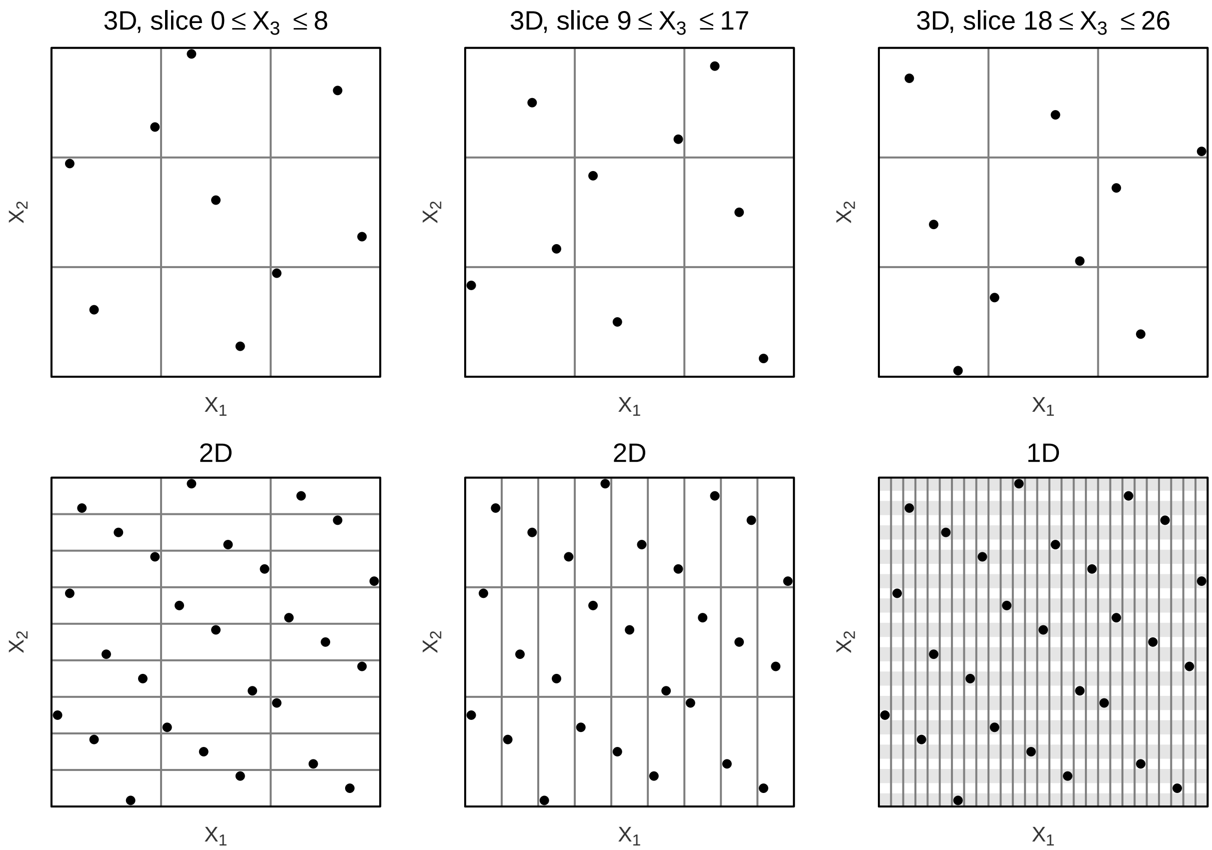

Figure 1: Illustration of stratification properties for an SOA(27,3,27,3). Each stratum contains exactly one element. The figures in the top row show the 3D stratification of $X_1 \times X_2 \times X_3$ into $3 \cdot 3 \cdot 3 = 27$ strata. The bottom row shows the stratification of $X_1 \times X_2$ into $3 \cdot 9 = 27$ strata (left) or $9 \cdot 3 = 27$ strata (middle), and the stratification of X_1 (horizontal) and X_2 (vertical) into 27 1D strata each (right).

- The constructions by Zhou and Tang (2019) and Li et al. (2021) are improved regarding their space filling properties by providing a slight modification to a key matrix in those constructions; this modification also improves the chance for obtaining better stratification properties.
- The construction for Shi and Tang's (2020) Family 3 is improved to achieve orthogonal columns.
- For the constructions by He et al. (2018), a bipartite pair matching algorithm for matching columns between two matrices guarantees orthogonal columns where orthogonality is compatible with the requested balance properties for the requested number of columns.
- The upper bound for the achievable number of columns in one of the He et al. constructions is slightly tightened, but is believed to be still quite lenient.

The different constructions for SOAs combine many established concepts of experimental design theory, so that a self-contained presentation requires a substantial amount of basic facts (Section 2). Section 3 presents and illustrates the early constructions, details the practically relevant classes of SOAs, provides the equations used for the constructions in this paper and states necessary and sufficient requirements for obtaining the different classes, as far as they can be stated in general terms. Sections 4 and 5 provide further specific constructions in the unified notation of this paper. Section 6 gives an overview of the constructions and their properties in terms of run sizes, numbers of factors and quality criteria. The discussion gives an overall assessment and an outlook at future developments, and several appendices connect the original constructions to the unified notation of this paper or provide start values and examples for recursive constructions.

2 Notation and basic facts

$\lfloor\cdot\rfloor$ and $\lceil\cdot\rceil$ denote the floor and ceiling functions, respectively.

2.1 Matrix notation

Matrices and vectors are denoted with bold face capital or lower case letters, respectively. $\mathbf{1}_n$ and $\mathbf{0}_n$ denote a column vector of n identical elements (1 or 0), $\top$ denotes the transpose of a matrix or vector. Vectors with single digit integer elements are parsimoniously written as a string of integers, e.g. $2 \cdot \mathbf{1}_5 = 22222$. $\otimes$ denotes the Kronecker product. The $n \times m$ matrix $\mathbf{X}$ is written as

$$\mathbf{X} = (x_{i,j})_{i=1:n,j=1:m} = \begin{pmatrix} x_{11} & \dots & x_{1m} \\ \vdots & & \vdots \\ x_{n1} & \dots & x_{nm} \end{pmatrix} = (\mathbf{x}_1, \dots, \mathbf{x}_m) = \begin{pmatrix} \mathbf{x}^{(1)} \\ \vdots \\ \mathbf{x}^{(n)} \end{pmatrix}.$$

For a matrix $\mathbf{M}$ with c columns, $\mathbf{M}_{c_1:c_2}$ denotes the sub matrix of columns c_1 to c_2. The function *cyc* denotes a cyclic permutation of the columns, i.e. $cyc(\mathbf{M}) = (\mathbf{m}_2, \dots, \mathbf{m}_c, \mathbf{m}_1)$.

2.2 Galois fields

A Galois field $GF(s)$ (see e.g. Appendix A of Hedayat et al. 1999) is a finite field over the elements $\{\alpha_0, \alpha_1, \dots, \alpha_{s-1}\}$; Galois fields exist whenever s is a prime or an integer power of a prime. Galois fields come with addition (neutral element α_0) and multiplication (neutral element α_1); for prime s, one can choose $\{\alpha_0, \alpha_1, \dots, \alpha_{s-1}\} = \{0, 1, \dots, s-1\}$ with $\mod s$ arithmetic. For non-prime prime powers, this paper also denotes the elements of the Galois field with the numbers $\{0, 1, \dots, s-1\}$, but uses suitable addition and multiplication tables that fulfill the requirements for a field; Tables 1 and 2 show the respective tables for prime powers 4, 8 and 9. Addition modulo a prime or non-prime s, as well as addition w.r.t. a Galois field $GF(s)$, will be denoted as $+_s$, multiplication as $\cdot_s$.

2.3 Orthogonal arrays

An OA is a rectangular table of symbols that typically stand for the levels of an experimental factor. In this paper, the columns of an OA stand for the factors, the rows for the level combinations used in experimental runs. An OA$(n, m, s_1^{m_1} \dots s_k^{m_k}, t)$ has n rows and m columns. m_1 columns have s_1 levels, $\dots$,m_k columns have s_k levels, $m_1 + \dots + m_k = m$. The OA's strength is t, which means that any combination of t columns indexed by $i_1 \dots i_t$ has all $s(i_1) \cdot_{\dots} \cdot s(i_t)$ level combinations the same number of times, where the function $s()$ returns the number of levels for the respective column. SOAs are typically

Table 1: Addition tables for GF(4), GF(8) and GF(9)

	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

	0	1	2	3	4	5	6	7
0	0	1	2	3	4	5	6	7
1	1	0	3	2	5	4	7	6
2	2	3	0	1	6	7	4	5
3	3	2	1	0	7	6	5	4
4	4	5	6	7	0	1	2	3
5	5	4	7	6	1	0	3	2
6	6	7	4	5	2	3	0	1
7	7	6	5	4	3	2	1	0

	0	1	2	3	4	5	6	7	8
0	0	1	2	3	4	5	6	7	8
1	1	2	0	4	5	3	7	8	6
2	2	0	1	5	3	4	8	6	7
3	3	4	5	6	7	8	0	1	2
4	4	5	3	7	8	6	1	2	0
5	5	3	4	8	6	7	2	0	1
6	6	7	8	0	1	2	3	4	5
7	7	8	6	1	2	0	4	5	3
8	8	6	7	2	0	1	5	3	4

Table 2: Multiplication tables for GF(4), GF(8) and GF(9)

	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	3	1
3	0	3	1	2

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	5	7	1	3
3	0	3	6	5	1	2	7	4
4	0	4	5	1	7	3	2	6
5	0	5	7	2	3	6	4	1
6	0	6	1	7	2	4	3	5
7	0	7	3	4	6	1	5	2

	0	1	2	3	4	5	6	7	8
0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8
2	0	2	1	6	8	7	3	5	4
3	0	3	6	7	1	4	5	8	2
4	0	4	8	1	5	6	2	3	7
5	0	5	7	4	6	2	8	1	3
6	0	6	3	5	2	8	7	4	1
7	0	7	5	8	3	1	4	2	6
8	0	8	4	2	7	3	1	6	5

based on *symmetric* OAs, i.e. OAs with $s_1 = \cdots = s_k = s$; such OAs are denoted by $\mathrm{OA}(n, m, s, t)$. Non-symmetric OAs are also called *asymmetric* or *mixed level*. For an $\mathrm{OA}(n, m, s, t)$, the strength implies that $n = \lambda s^t$ for an integer λ that is called the index of the OA. At this point, note that the expression "strength" will have to be used in two different meanings in this paper about SOAs: "OA strength" will always be explicitly referred to as such, whereas the mere use of the word "strength" outside of this section always refers to "SOA strength", which is a different concept (see Definition 3.1).

OAs typically have small numbers of levels; many of the usual symmetric OAs have $s = 2$, $s = 3$ or at most $s = 4$, and mixed level OAs often have the majority of their factors at 2 or 3 levels, possibly with a few exceptions. There are various construction algorithms for symmetric OAs; these are, for example, explained in Hedayat, Sloane and Stufken (1999). They can typically construct OAs whose number of levels s is a power of a prime (e.g. 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, ...). Among the construction algorithms, *regular* fractional factorial OAs take a specific role: An OA is called "regular", if its columns can be obtained as linear combinations of some basic columns. The construction of such arrays for prime numbers of levels can be based on modulo arithmetic, whereas powers of a prime (e.g. 4, 8, 9) require Galois field arithmetic.

The imbalance of an OA is often measured by the so-called generalized word length pattern (GWLP), which is a way to measure the confounding of j factor interactions with the intercept for $j = 0, 1, \dots, m$. The elements of the GWLP are denoted as $A_0, A_1, A_2, \dots, A_m$, with $A_0 = 1$ (intercept perfectly confounded with itself). The entries A_j are zero for all $j \le t$, which indicates the balance implied by the strength. The smallest j for which $A_j > 0$ is called the resolution of the OA (i.e. the resolution is $t + 1$). Xu and Wu (2001) introduced the GWLP and the ranking criterion "generalized minimum aberration" (GMA), which works as follows: a design with higher OA strength (or higher resolution) is better; for two designs with resolution R, the design with the smaller A_R is better; if A_R is the same, the design with the smaller A_{R+1} is better, and so forth. A GMA design is an overall best design according to this criterion (it is not necessarily unique).

SOAs are based on OAs with at least strength 2 and are themselves strength 1 OAs, i.e. they have resolution II (resolution is denoted as a Roman numeral). Their A_2 value can thus be used to measure their imbalance in terms of the GWLP; however, since SOAs are typically created for quantitative variables, the GWLP is not necessarily a suitable metric for assessing their quality.

2.3.1 Regular saturated strength 2 fractions

Several constructions of SOAs make use of regular saturated strength 2 $\mathrm{OA}(s^k, (s^k - 1)/(s - 1), s, 2)$ that can be obtained for s-level columns with s a prime or prime power, using the so-called Rao-Hamming construction (see e.g. Section 3.4 of Hedayat et al.): It consists of a set of k basic vectors on $GF(s)^{s^k}$ and all their linear combinations, where uniqueness is achieved by using only those coefficient vectors from $GF(s)^k$ for which the first non-zero element is 1. For example, an $\mathrm{OA}(3^2, (3^2 - 1)/(3 - 1), 3, 2) = \mathrm{OA}(9, 4, 3, 2)$ is obtained from the two basic vectors 012012012 and 000111222 and their two interactions with coefficients (1,1) and (1,2), which are the columns 012120201 and 012201120.

2.3.2 Yates matrix for regular 2-level fractions

Regular fractional factorial 2-level OAs are particularly well-understood (see e.g. Mee 2009). Note that they are often discussed in $-1/+1$ coding with multiplication instead of 0/1 coding with $+_2$ (remember that $+_2$ denotes addition modulo 2). The two approaches are equivalent, and this paper uses the latter because of consistency with the cases for $s \neq 2$.

Large catalogues of non-isomorphic regular 2-level fractions are available. These can be parsimoniously specified using the so-called Yates matrix column numbers, whose fascinating systematics is now explained. A Yates matrix of degree k is the $2^k \times (2^k - 1)$ matrix for all effects in a full factorial design with k basic factors. Starting from two basic factors, the principle of recursive construction is shown below:

$$Y(2) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}, \quad Y(3) = \begin{pmatrix} Y(2) & \mathbf{0}_4 & Y(2) \\ Y(2) & \mathbf{1}_4 & Y(2) +_2 1 \end{pmatrix}, \quad Y(4) = \begin{pmatrix} Y(3) & \mathbf{0}_8 & Y(3) \\ Y(3) & \mathbf{1}_8 & Y(3) +_2 1 \end{pmatrix}, \quad \dots$$

If $\mathbf{e}_1, \mathbf{e}_2, \dots$ denote the basic factors (fast changing first), the Yates matrix column numbers have a systematic structure, and their binary representations indicate which effects are captured by them: $\mathbf{e}_j$ is in column 2^{j-1}, e.g. for $k = 4$, the four basic factors are in columns 1, 2, 4, and 8 (binary representations: 1=0001,2=0010,4=0100,8=1000). Further column numbers also indicate which interaction effect is represented by the column in a full factorial model, for example column 11 (binary representation 11=1011, with further leading zeroes for $k > 4$) captures the three-factor interaction of $\mathbf{e}_1$, $\mathbf{e}_2$ and $\mathbf{e}_4$. The structure of the Yates matrix implies that the first $2^u - 1$ columns capture the effects of basic factors $\mathbf{e}_1, \ldots, \mathbf{e}_u$ and all their interactions. Likewise, the $2^{k-u} - 1$ Yates matrix columns numbered with multiples of 2^u capture the effects of the last $k - u$ basic factors and all their interactions. For example, for $k = 5$ and $u = 3$, columns 1 to 7 capture all effects related to the first three basic factors, whereas the three columns numbered with multiples of 8 (8, 16, 24) capture the design with the last $5 - 3 = 2$ basic columns. Computationally, instead of the recursive construction, if a function for creating a full factorial in k columns is available that allows to specify the order of columns (first or last factor changing fastest), the Yates matrix for k factors can be obtained by multiplying the full factorial in the k basic factors (fastest first) with the transpose of that same full factorial (slowest first) (calculations $\mod 2$), and omitting the column of zeroes in the first position.

2.4 Latin hypercube designs and space filling criteria

Projections of a symmetric s-level OA of OA strength t onto a 1D, 2D, $\ldots, t$D marginal space have exactly $s, s^2, \ldots, s^t$ distinct points. For the typically small s, there are thus very few distinct points in lower-dimensional projections. The benefit of OAs lies in their implicit replication that allows the estimation of low order interaction effects and random error. As soon as random error becomes irrelevant, like in computer experiments, the small number of points in low order projections becomes a disadvantage. Therefore, LHDs have been proposed especially for computer experiments with quantitative factors, for which changing the levels is often easy and good exploration of the entire experimental space is desired.

An LHD for m quantitative variables in n runs is an OA$(n, m, n, 1)$, i.e., each column has as many levels as there are rows (=runs). We write LHD(n, m). Note that some authors write about latin hypercube *designs*, others about latin hypercube *samples*; both are two slightly different versions of the same concept. LHDs are usually specified in terms of integer levels. In latin hypercube *sampling*, one often works with real numbers in $[0, 1]^m$, where the integer numbers from the LHD approach correspond to intervals within $[0, 1]$. This paper works with LHDs, but the difference is unsubstantial since it is straightforward to go back and forth between LHD and latin hypercube sample.

There is a large amount of literature on quality criteria for LHDs. It is commonly agreed that LHDs should be "space-filling", i.e. fill the m-dimensional space as well as possible. Popular metrics to assess their space-filling properties are the minimum interpoint distance (which should be as large as possible, maximin distance) or the maximum interpoint distance (which should be as small as possible, minimax distance), or various criteria that measure discrepancy from uniformity in some way (and should be small). This paper uses the maximin distance criterion, as well as the so-called ϕ_p criterion which is commonly used in the SOA literature:

$$\phi_p(\mathbf{X}) = \left(\sum_{\{i,j\} \subset \{1,\ldots,n\}, i \neq j} d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})^{-p} \right)^{1/p},$$

where $d()$ is a suitable distance function, e.g. Minkowski with $q = 2$ (euclidean) or $q = 1$ (manhattan) and $x^{(i)}$ is the observation vector for the ith unit, i.e. the ith row of the matrix $\mathbf{X}$ that represents the runs. For large p, minimizing ϕ_p is known to be a good substitute for the maximin distance criterion. $d(c \cdot \mathbf{x}^{(i)}, c \cdot \mathbf{x}^{(j)}) = c \cdot d(\mathbf{x}^{(i)}, \mathbf{x}^{(j)})$, and $\phi_p(c\mathbf{X}) = \phi_p(\mathbf{X})/c$, so that normalized versions d^*_{ij} and ϕ^*_p for range $[0, 1]$ can be easily obtained, which is helpful when comparing arrays with different numbers of levels.

2.5 Orthogonal columns

For an OA, "orthogonal" refers to combinatorial balance which is invariant to level coding. When talking about orthogonal columns, "orthogonal" refers to geometric orthogonality in n-dimensional space, which can be measured by correlation between columns: two columns are geometrically orthogonal if their correlation is zero. Combinatorial orthogonality implies geometric orthogonality, but the reverse is not true. Geometric orthogonality is of interest for quantitative experimental variables only, and it is heavily dependant on level coding. LHDs with orthogonal columns have been discussed in the literature (e.g. Ye 1998). The benefit of geometric orthogonality is that estimated coefficients in simple main effect linear regression do not change, regardless whether one does or does not include other columns in the model. The expression "3-orthogonality", which was introduced by Bingham, Sitter and Tang (2009), describes a stronger orthogonality property: 3-orthogonal arrays guarantee that columns are not only orthogonal to each other and to the constant column, but also to products of pairs of other columns and to squares of other columns. This effectively means that main effect estimates are uncorrelated with the estimates of second order effects (i.e. of pairwise column products or squared columns). 3-orthogonality thus prevents misleading conclusions on main effects from neglecting relevant second order effects, and makes estimation of second order models more efficient. To the author's knowledge, Ye (1998) was the first author to propose 3-orthogonal LHDs; his construction produces 2^k or $2^k + 1$ runs with $2k - 2$ columns (but does not yield SOAs).

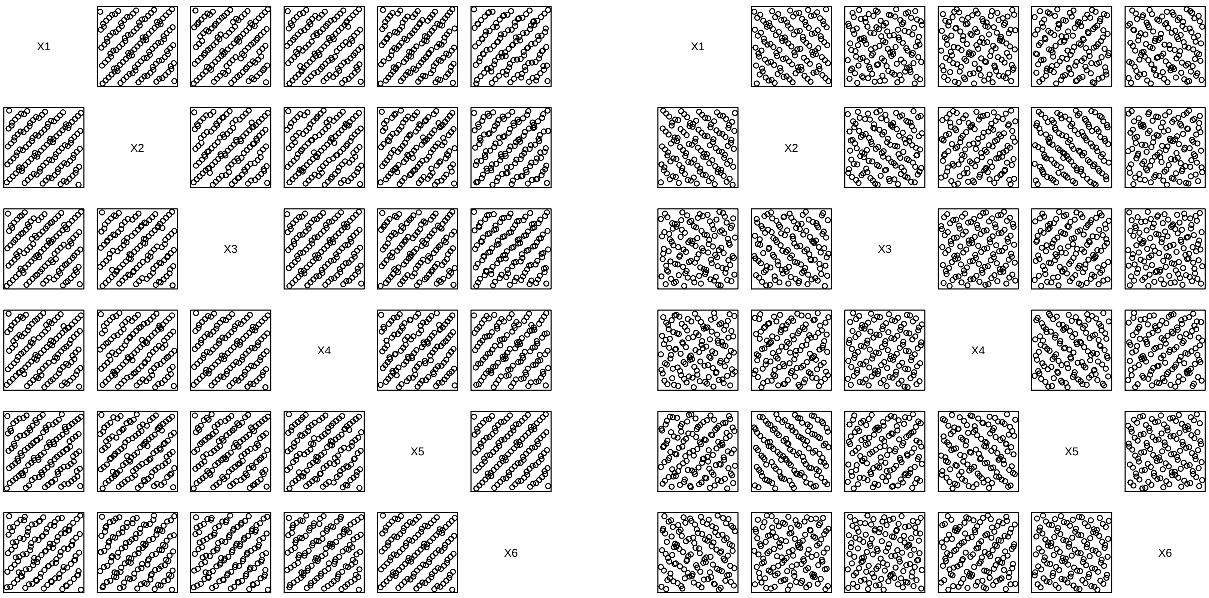

Figure 2: 2D projections of unoptimized (left) and optimized (right) SOA in 125 runs with six orthogonal columns at 125 levels each

Note that orthogonality or 3-orthogonality does not imply space filling; thus, it is advisable to consider additional space-filling criteria, because the default constructions can exhibit strong patterns that leave large holes unfilled. For example, an unoptimized orthogonal 125 run SOA seems to arrange the design points in parallel diagonal stripes that are less space-filling than would be desirable, while even one round of optimization towards lowering ϕ_p substantially improves this behaviour as can be seen from the 2D projections in Figure 2, and by comparing the ϕ_p values (0.0395 reduced to 0.013 by the optimization). Note, however, that the right-hand side figure also shows systematic holes in several of the projections.

2.6 Expanding levels

An early proposal for obtaining LHDs by expanding the levels of OAs was by Tang (1993). Two simple techniques to expand an OA(n, m, s, t) to a symmetric OA with OA strength $t' \geq 1$ and with columns in $s \cdot \ell$ levels are presented in Sections 2.6.1 and 2.6.2. Level expansion of OAs is easy to implement. The quality of the resulting array strongly depends on the level orderings within the initial OA, and also within replacements. Depending on the size of ℓ, the space to optimize over can be huge.

2.6.1 Expand within the OA

This type of expansion returns an array with the same number of rows as the ingoing OA. It can be applied, if ℓ divides n/s. An OA$(n, m, s \cdot \ell, t')$ can be obtained by expanding the levels of each column in the following way: allocate

- new levels $0, \ldots, \ell - 1$ to the n/s runs with old level 0,

- new levels $\ell, \ldots, 2\ell - 1$ to the n/s runs with old level 1,
- ...,
- and new levels $(s-1)\ell, \ldots, s\ell - 1$ to runs with old level $s - 1$,

conducting all allocations in a balanced way.

Xiao and Xu (2018) proposed an algorithm for optimizing this type of expansion, which they called MDLE for "maximin distance level expansion". In particular, they showed that it is beneficial to start from a GMA OA, which has itself been optimized for maximin distance by level permutations, before applying level expansion (the maximization of the minimum distance can be omitted for 2-level starting OAs). Xiao and Xu proposed to use a threshold acceptance (TA) algorithm. Their approach is computationally more demanding than the optimization proposed by Weng (2014, see Section 2.8), but also yields better results. Since Xiao and Xu did not impose any structural constraints on the level expansion, the optimization space for MDLE is larger than that for SOAs. MDLE designs will not be covered in this paper.

2.6.2 Expand levels by expanding each row

This type of expanding an $\mathrm{OA}(n, m, s, t)$ returns an $\mathrm{OA}(n \cdot \ell, m, s \cdot \ell, t')$ with $t' \geq 1$. There are no requirements for $\ell \geq 2$, except that it is an integer. In the $\mathrm{OA}(n, m, s, t)$, insert a vector

- $0, \ldots, \ell - 1$ instead of positions with level 0,
- $\ell, \ldots, 2\ell - 1$ instead of positions with level 1,
- $\ldots,$
- $(s-1)\ell, \ldots, s\ell - 1$ instead of positions with level $s - 1$.

Of course, level permutation can also greatly affect the properties of arrays obtained by this type of expansion.

2.7 Collapsing levels

Collapsing the levels $0, \ldots, s^v - 1$ of a column in s^v levels into only s^u levels $0, \ldots, s^u - 1$, $u < v$, can be simply done with the formula $x_{s^u} = \lfloor x_{s^v}/(s^{v-u}) \rfloor$, where $\lfloor \cdot \rfloor$ denotes the floor function.

If an array was obtained by level expansion, collapsing its levels (again) to s levels either recovers the original array (in case of Section 2.6.1) or a replicate of the original array (in case of Section 2.6.2). In either case, the collapsed array inherits its balance properties from the original array.

For SOAs with columns in s^v levels, stratifications into collapsed columns are often considered, e.g. 2D stratifications of s^3 level columns into $s^2 \times s$ or $s \times s^2$, 3D stratifications of s^4 level columns into $s^2 \times s \times s$ or $s \times s^2 \times s$ or $s \times s \times s^2$. In order to avoid both repetitive writing and complex notation, this paper uses the single representative with exponents sorted from largest to smallest to include all stratifications with the same multi-set of numbers of levels, e.g. "all $s^2 \times s \times s$ stratifications" includes all the 3D stratifications that were listed above for s^4 level factors.

2.8 Optimization by level permutations

As was mentioned before, while combinatorial properties are invariant to the level coding of array columns, column orthogonality or space filling properties heavily depend on level coding. Column orthogonality is typically guaranteed by a construction mechanism (see Sections 3.1.2, 4.1 and 5.1); space filling properties can be optimized by adequate level permutations that do not destroy structural requirements.

A brute force method for level permutations would conduct all combinations of conceivable non-distinct level permutations and select the best outcome. For many situations, such an approach is prohibitive. Weng (2014) suggested to proceed in a reduced version that is adopted in this paper and is now explained. This section assumes the quality criterion ϕ_p for space filling.

Let ν denote the number of permutation applications; ν could e.g. be $3m$, if each of the m columns of three m-column matrices with s levels each is subjected to separate level permutation, like in Equation (3) of the next section. The total number of permutations for a brute force search would be $(s!)^\nu$. In Weng's approach, the number of actual permutations that have to be conducted is far smaller:

a) Start with a random pattern Π_0 of ν level permutations.

b) Obtain all the one-neighbours of Π_0, in the sense that only a single permutation of the ν level permutations is modified versus Π_0.
c) Assess ϕ_p for Π_0 and all its one-neighbors.
d) If Π_0 is already best in step c, proceed to step e. Else restart step b with the best pattern of level permutations as the new Π_0.
e) Inspect the current pattern of level permutations and all its two-neighbors, in the sense that two permutations are modified ($\binom{\nu}{2}$ such two-neighbours).
f) If Π_0 is best in step e, declare it the winner. Else restart step b with the best pattern of level permutations from step e as the new Π_0.

Weng applied her approach to the He and Tang (2013) constructions and gave examples for which this reduced method came close to the optima found in previous brute-force searches, and she also emphasized that omission of the step with two-neighbours does not lead to satisfactory results. When applying her approach to other constructions, care must be taken that the structural requirements of a construction are not destroyed by level permutations.

3 SOAs

This section provides a formal look at SOAs and their properties, starting with the definition of an (O)SOA of strength t:

Definition 3.1 (SOA and OSOA)**.** Let $\mathbf{D}$ denote an OA$(n, m, s^t, 1)$.

(i) $\mathbf{D}$ is an SOA of strength t, denoted as SOA(n, m, s^t, t) $(m \geq t)$, if and only if all j-dimensional $s^{u_1} \times \cdots \times s^{u_j}$ projections for columns $i_1, \ldots, i_j$, $1 \leq j \leq t$ produce s^t equally-sized strata, where the i_ℓth column is collapsed to s^{u_ℓ} levels, $s^{u_\ell} \geq 1$, and $\sum_1^j u_\ell = t$.
(ii) An SOA(n, m, s^t, t) whose correlation matrix is the m-dimensional identity matrix is called an OSOA(n, m, s^t, t).

Figure 1 visualized an SOA$(27, 3, 3^3, 3)$, i.e. $s = 3$ and $t = 3$. For $j = 1$, each 1D projection onto 3^3 levels (i.e. the consideration of a single column without level coarsening) has exactly $3^3 = 27$ equally-sized strata; for $j = 2$, each 2D $3^2 \times 3^1$ or $3^1 \times 3^2$ projection (i.e. consideration of two columns at a time with one column coarsened to nine levels and the other to three levels) has exactly 27 equally sized strata; and the 3D $3^1 \times 3^1 \times 3^1$ projection (i.e. three columns, each coarsened to three levels) also has 27 equally-sized strata.

This paper presents all SOA constructions in the form of simple equations: Let $\mathbf{A}_1$ to $\mathbf{A}_k$ denote a set of OA$(n, m, s, 1)$, such that

$$\mathbf{D} = s^{k-1}\mathbf{A}_1 + \cdots + s\mathbf{A}_{k-1} + \mathbf{A}_k \tag{1}$$

is an OA$(n, m, s^k, 1)$; the "such that" implies assumptions for $\mathbf{A}_1$ to $\mathbf{A}_k$ that are not inspected for the general form of Equation (1). The following lemma makes clear that a strong $\mathbf{A}_1$ in this type of equation is generally beneficial for the balance properties of $\mathbf{D}$. Its proof is a trivial consequence of the fact that the equation is a way to produce a level expansion for $\mathbf{A}_1$ and is therefore omitted.

Lemma 3.1. *If an OA$(n, m, s^k, 1)$ named $\mathbf{D}$ has arisen from Equation (1), where $\mathbf{A}_1$ has strength t, collapsing the columns of $\mathbf{D}$ to s levels each by the formula $\lfloor x/s^{k-1} \rfloor$ yields an OA(n, m, s, t).*

This paper mainly considers constructing SOAs in s^2 levels from

$$\mathbf{D} = s\mathbf{A} + \mathbf{B} \tag{2}$$

and SOAs in s^3 levels from

$$\mathbf{D} = s^2\mathbf{A} + s\mathbf{B} + \mathbf{C}. \tag{3}$$

Lemma 3.2 (recast from He and Tang 2013)**.** *An SOA$(n, m, s^2, 2)$ $\mathbf{D}$ exists if and only if $n \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$ can be found such that $\mathbf{A}$ is an OA$(n, m, s, 2)$, and all pairs $(\mathbf{a}_\ell, \mathbf{b}_\ell)$ are OA$(n, 2, s, 2)$. These arrays are related by Equation (2).*

Table 3: Constructions by He and Tang 2013 of an SOA(n, m', s^t, t) based on an $n \times m$ matrix $\mathbf{V}$, which is an OA(n, m, s, t). Notations were explained in Section 2.1.

t	$\mathbf{D} =$	Matrices
2	$s\mathbf{A} + \mathbf{B}$	$\mathbf{A} = \mathbf{V}, \mathbf{B} = cyc(\mathbf{A})$
3	$s^2\mathbf{A} + s\mathbf{B} + \mathbf{C}$	$\mathbf{A} = \mathbf{V}_{1:(m-1)}, \mathbf{B} = (\mathbf{v}_m, \dots, \mathbf{v}_m), \mathbf{C} = cyc(\mathbf{A})$
4	$s^3\mathbf{A}_1 + s^2\mathbf{A}_2 + s\mathbf{A}_3 + \mathbf{A}_4$	$\mathbf{A}_1 = \mathbf{V}_{1:m'}, \mathbf{A}_2 = \mathbf{V}_{(m'+1):(2m')}, \mathbf{A}_3 = cyc(\mathbf{A}_2), \mathbf{A}_4 = cyc(\mathbf{A}_1)$
5	$s^4\mathbf{A}_1 + s^3\mathbf{A}_2 + s^2\mathbf{A}_3 + s\mathbf{A}_4 + \mathbf{A}_5$	$\mathbf{A}_1 = \mathbf{V}_{1:m'}, \mathbf{A}_2 = \mathbf{V}_{(m'+1):(2m')}, \mathbf{A}_3 = \mathbf{1}_{m'}^\top \otimes \mathbf{v}_m,$ $\mathbf{A}_4 = cyc(\mathbf{A}_2), \mathbf{A}_5 = cyc(\mathbf{A}_1)$

Lemma 3.3 (Shi and Tang, quoting He and Tang 2013)**.** *An SOA$(n, m, s^3, 3)$ $\mathbf{D}$ exists if and only if $n \times m$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ can be found such that $\mathbf{A}$ is an OA$(n, m, s, 3)$, and all triples $(\mathbf{a}_\ell, \mathbf{a}_j, \mathbf{b}_j)$, $(\mathbf{a}_\ell, \mathbf{b}_\ell, \mathbf{c}_\ell)$ are OA$(n, 3, s, 3)$, $\ell \neq j$. These arrays are related by Equation (3).*

Similar statements can also be made for larger strengths.

The following trivial lemma shows that it is generally easy to assign a column $\mathbf{c}_\ell$ for given columns $\mathbf{a}_\ell$ and $\mathbf{b}_\ell$ such that strength 3 is achieved, as long as an SOA is based on 2-level OAs whose columns are taken from a saturated regular fractional factorial $\mathbf{S}$.

Lemma 3.4. *For $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ chosen as columns from a saturated fractional factorial 2-level design $\mathbf{S}$, the matrix $\mathbf{C}$ fulfills the assumptions of Lemma 3.3, iff $\mathbf{c}_\ell$ does not coincide with any of $\mathbf{a}_\ell$, $\mathbf{b}_\ell$ or $\mathbf{a}_\ell +_2 \mathbf{b}_\ell$.*

The next lemma will be used for improving a construction to yield orthogonal columns.

Lemma 3.5. *Let $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ be OA$(n, m, s, 2)$ such that their combination into a single matrix $(\mathbf{A} \vdots \mathbf{B} \vdots \mathbf{C})$ is an OA$(n, 3m, s, 2)$. Then Equation (3) yields a matrix $\mathbf{D}$ with orthogonal columns.*

Proof. $\mathbf{D} = s^2\mathbf{A} + s\mathbf{B} + \mathbf{C}$. The correlation matrix of $\mathbf{D}$ can be obtained with the usual rules for calculating correlation matrices of sums and products with scalars. The strong assumptions of the lemma imply the requested orthogonality property. □

The following subsection presents the early constructions of strength 2 to strength 4 or 5 (O)SOAs. Later sections will provide constructions for refined strength classifications.

3.1 Early constructions for (O)SOAs

3.1.1 SOAs of strengths 2 to 5 by He and Tang (2013)

He and Tang (2013) introduced SOAs, based on so-called generalized orthogonal arrays (GOAs), for strengths $t = 2, \dots, 5$. Their specific constructions are presented here in the form of Equation (1) instead, which is straightforward, because it only requires a re-grouping of the columns of the matrices from Equation (1): the GOA matrices $\mathbf{B}_1^{GOA}, \dots, \mathbf{B}_m^{GOA}$ for m SOA columns hold the first columns of the matrices $\mathbf{A}_j$ in $\mathbf{B}_1^{GOA}$, the second columns in $\mathbf{B}_2^{GOA}$, and so forth, obtaining m matrices with t columns each for a strength t SOA with s^t level columns. The formulae in Table 3 state constructions for an SOA(n, m', s^t, t) from an OA(n, m, s, t), which is named $\mathbf{V}$. The number of columns m' obtainable from the m columns of $\mathbf{V}$ can be calculated as

$$m'(m, t) = \begin{cases} \lfloor 2m/t \rfloor & t \text{ even} \\ \lfloor 2(m-1)/(t-1) \rfloor & t \text{ odd} \end{cases}. \qquad (4)$$

The relevant specific cases are $m'(m, 2) = m$, $m'(m, 3) = m - 1$, $m'(m, 4) = \lfloor m/2 \rfloor$ and $m'(m, 5) = \lfloor (m-1)/2 \rfloor$. The matrices in the construction equations have m' columns each. In each construction, $\mathbf{V}$ denotes an OA(n, m, s, t). Table 3 summarizes the constructions; the rows for $t = 2$ and $t = 3$ correspond to Equations (2) and (3); for the sake of avoiding several subscript levels, these equations use $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ instead of $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$.

He and Tang (2013) themselves did not consider the use of level permutations in SOA construction. As Weng's (2014) algorithm greatly improves space filling of the resulting SOAs, it is recommended to apply it to this construction.

3.1.2 OSOAs of strengths 2 to 4 by Liu and Liu

Liu and Liu presented constructions for OSOAs. They chose signed levels centered at zero, which is quite common for the orthogonal column case. Nevertheless, for keeping the same notation for all constructions, we will continue to use $0, \ldots, s-1$ for the OA levels and $0, \ldots, s^t-1$ for the levels of the OSOA. Strength $t = 4$ already requires a large number of rows per column, so that designs with larger strengths are not considered. The constructions correspond to Theorem 2 of Liu and Liu for even strength and to Theorem 4 of Liu and Liu for odd strength. They use the equations of Table 3, based on the matrices provided below. The connections of those matrices to Liu and Liu's exposition are detailed in the appendix.

Lemma 3.6 (Liu and Liu 2015 Theorems 2 and 4)**.** *The subsequent t-specific constructions ($t = 2, 3, 4$) based on an OA(n, m, s, t) named $\mathbf{V}$ create an OSOA(n, m', s^t, t), where m' is provided in the constructions. If $t > 2$ and $m' > 2$, the columns are 3-orthogonal.*

Note that the OA $\mathbf{V}$ used in the constructions need not be regular, and s need not be a prime power.

The following definition simplifies the presentation of the constructions:

Definition 3.2. Let $\mathbf{M}$ be an $n \times m$ matrix with elements from $\{0, \ldots, s-1\}$, m even. The function $\mathcal{S}$ returns an $n \times m$ matrix with the ℓth column given as

$$\mathcal{S}(\mathbf{M})_\ell = \begin{cases} \mathbf{m}_{\ell+1} & \ell \text{ odd} \\ s-1-\mathbf{m}_{\ell-1} & \ell \text{ even} \end{cases}, \quad \ell = 1, \ldots, m.$$

If $\mathbf{V}$ is an OA$(n, m, s, 2)$, the $m' = 2\lfloor m/2 \rfloor$ columns of the matrix $\mathbf{A}$ are obtained as

$$\mathbf{a}_\ell = \begin{cases} \mathbf{v}_{\ell+1} & \ell \text{ odd} \\ \mathbf{v}_{\ell-1} & \ell \text{ even} \end{cases}, \quad \ell = 1, \ldots, m', \tag{5}$$

and $\mathbf{B} = \mathcal{S}(\mathbf{A})$. Using $\mathbf{A} = \mathbf{V}_{1:m'}$ would yield a different but comparable construction.

If $\mathbf{V}$ is an OA$(n, m, s, 3)$, the first $\widetilde{m} = 2\lfloor m/4 \rfloor$ columns of the matrices $\mathbf{A}$ and $\mathbf{B}$ are obtained as

$$\mathbf{a}_\ell = \begin{cases} \mathbf{v}_{2\ell+1} & \ell \text{ odd} \\ \mathbf{v}_{2\ell-3} & \ell \text{ even} \end{cases}, \quad \mathbf{b}_\ell = \mathbf{v}_{2\ell}, \quad \ell = 1, \ldots, \widetilde{m}, \tag{6}$$

and $\mathbf{C} = \mathcal{S}(\mathbf{A})$. If $m - 2\widetilde{m} < 3$, $m' = \widetilde{m}$. Otherwise, $m' = \widetilde{m} + 1$, and the additional column can be obtained as follows:

$$\mathbf{a}_{m'} = \mathbf{v}_m, \quad \mathbf{b}_{m'} = \mathbf{v}_{m-1}, \quad \mathbf{c}_{m'} = \mathbf{v}_{m-2}. \tag{7}$$

If $\mathbf{V}$ is an OA$(n, m, s, 4)$, the $m' = 2\lfloor m/4 \rfloor$ columns of the matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ are obtained as

$$\mathbf{a}_{1;\ell} = \begin{cases} \mathbf{v}_{2\ell+2} & \ell \text{ odd} \\ \mathbf{v}_{2\ell-3} & \ell \text{ even} \end{cases}, \quad \mathbf{a}_{2;\ell} = \begin{cases} \mathbf{v}_{2\ell+1} & \ell \text{ odd} \\ \mathbf{v}_{2\ell-2} & \ell \text{ even} \end{cases}, \tag{8}$$

together with $\mathbf{A}_3 = \mathcal{S}(\mathbf{A}_2)$ and $\mathbf{A}_4 = \mathcal{S}(\mathbf{A}_1)$.

Hence, among the t matrices for a strength t construction, the last $\lfloor t/2 \rfloor$ matrices are obtained from the first $\lfloor t/2 \rfloor$ matrices by applying function $\mathcal{S}$. The permutation approach by Weng (2014) can thus be applied independently to the columns of the first $\lceil t/2 \rceil$ matrices. This can be implemented by independently permuting the levels of the columns of $\mathbf{V}$. (Level permutations seem to be less powerful for this construction than for some others.)

For strength 2, a wide variety of OSOAs can be constructed with this technique. The constructions for larger strengths require relatively many runs per column. For example, one can obtain an OSOA$(64, 16, 8, 3)$

Table 4: Number of equally-sized strata for low-dimensional projections of refined classes of (O)SOAs

	strength	1D	2D	3D	4D
s^2 levels	2	s^2	s^2	$--$	$--$
	2+	s^2	s^3	$--$	$--$
	3−	s^2	s^3	s^3	$--$
s^3 levels	2*	s^3	s^3	$--$	$--$
	3	s^3	s^3	s^3	$--$
	3+	s^3	s^4	s^4	$--$
s^4 levels	4	s^4	s^4	s^4	s^4

from an OA$(64, 32, 2, 3)$, an OSOA$(81, 4, 27, 3)$ from an OA$(81, 10, 3, 3)$, an OSOA$(64, 4, 16, 4)$ from an OA$(64, 8, 2, 4)$ or an OSOA$(256, 8, 16, 4)$ from an OA$(256, 17, 2, 4)$. The considerable benefit of the constructions for $t > 2$ is that they produce 3-orthogonal columns, which implies that main effects are uncorrelated to second order effects in linear regression models (see Section 2.5).

3.2 Classes of SOAs

We saw general results for strength 2 and 3 SOAs and early constructions by He and Tang (2013) for SOAs with strengths 2 to 5 and by Liu and Liu (2015) for OSOAs with strengths 2 to 4. In the following, this paper considers constructions for strength 3 and four more refined classes of SOAs, all of which provide less balance than strength 4 but more balance than strength 2. This is because strength 4 or higher is usually prohibitive in terms of run size, while strength 2 without further balance criteria is often not satisfactory.

Table 4 summarizes the stratification properties for the classes of (O)SOAs considered in the following, namely (O)SOAs of strengths 2+, 3−, 2*, 3, or 3+; the border cases of strengths 2 and 4 are included for reference. 1D projections consider a single column, 2D projections two columns collapsed to s^a and s^b levels, with $a + b$ equal to the exponent in the table entry, 3D projections three columns collapsed to s^a, s^b and s^c levels with $a + b + c$ equal to the exponent in the table entry, and so forth. Note that the expression "strength 3+" is coined here (see Definition 3.4), in obvious analogy to the expression "strength 2+" used by He, Cheng and Tang (2018).

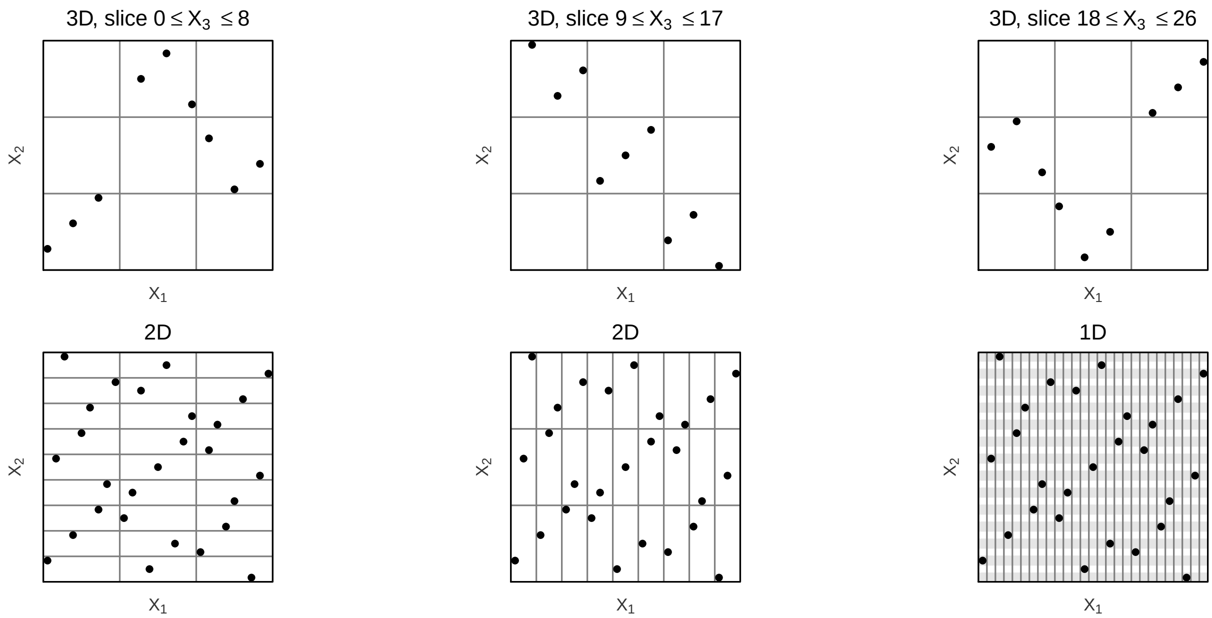

Figure 3: Illustration of stratification properties for the first three columns of the unoptimized OSOA(27,4,27,2*) from the Li et al. construction.

Figure 1 illustrated strength 3 stratification properties, using an SOA$(27, 3, 3^3, 3)$. As an SOA$(n, m, s^3, 2^*)$ shares the properties of an SOA$(n, m, s^3, 3)$ for 2D and 1D projections and does not provide stratification guarantees for 3D projections, some, many or all 3D projections for a strength 2^* SOA may look like the top row of Figure 3 (compared to the top row of Figure 1). SOA$(n, m, s^2, 3-)$ stratify like Figure 1, except for having fewer levels so that the bottom right plot would have three points in each of nine rows and columns, respectively, and the points would also be less dispersed in the other plots. An SOA$(n, m, s^2, 2+)$ is not only coarser than the SOA shown in Figure 1 but may also have 3D stratification behavior similar to Figure 3.

Remark. One might argue that the introduction of strengths $3-$ and 2^* is somewhat redundant, because it would suffice to communicate the numbers of levels with the strengths 3+, 3 or 2+: strengths 3+, 3 and 2^* indicate s^3 levels, strengths $3-$ and 2+ indicate s^2 levels for each column, i.e. the pairs $(3+, s^3)$, $(3, s^3)$, $(2+, s^3)$, $(3, s^2)$ and $(2+, s^2)$ would be sufficient. Nevertheless, we use the established notation from the literature, and its natural extension by 3+.

OSOAs have orthogonal columns. SOAs without the “O” tend to have correlated columns, but whenever space filling behavior is optimized in some way, correlations are typically not too severe, so that it may be acceptable to use non-orthogonal SOAs, as long as they achieve better space filling properties. For example, the optimized 27 run SOA from Figure 1 has X_2 uncorrelated with the other two columns, and the correlation of X_1 with X_3 is approximately -0.0495 (while the unoptimized version had correlation almost 0.2 for all pairs).

Lemmata 3.2 and 3.3 stated existence and construction hints for strength 2 and strength 3 (O)SOAs. He, Cheng and Tang (2018) ascertained the following construction for strength 2+ (their proposition 1):

Lemma 3.7 (He, Cheng and Tang 2018)**.** *An SOA$(n, m, s^2, 2+)$ exists if and only if $n \times m$ matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *can be found such that* $\mathbf{A}$ *is an OA$(n, m, s, 2)$,* $\mathbf{B}$ *is an OA$(n, m, s, 1)$, and all triples* $(\mathbf{a}_\ell, \mathbf{a}_j, \mathbf{b}_j)$ *are OA$(n, 3, s, 3)$ for $\ell \neq j$. These arrays are linked through Equation (2).*

Zhou and Tang (2019) gave conditions for obtaining an OSOA of strength 2+ (their Theorem 1 and Remark 1):

Lemma 3.8 (Zhou Tang 2019)**.** *If the matrix* $\mathbf{B}$ *in Lemma 3.7 is an OA$(n, m, s, 2)$ or a column-orthogonal OA$(n, m, s, 1)$, Equation (2) yields an OSOA$(n, m, s^2, 2+)$.*

In the light of this lemma, it is advisable to bring $\mathbf{B}$ as close to strength 2 as possible for strength 2+ SOAs, in order to achieve column orthogonality, where possible. Zhou and Tang furthermore gave conditions for making the SOA obtained from Equation (2) achieve strength $3-$ (their Lemma 1):

Lemma 3.9 (Zhou Tang 2019)**.** *If the matrix* $\mathbf{A}$ *in Lemma 3.7 is an OA$(n, m, s, 3)$, Equation (2) yields an SOA$(n, m, s^2, 3-)$.*

Li, Liu and Yang (2021) stated their rules for strength 2^* w.r.t. their specific construction only. In general terms, strength 2^* is attained, whenever the conditions of Lemma 3.3 are fulfilled, except for weakening the requirement for $\mathbf{A}$ to strength 2 instead of strength 3.

Shi and Tang (2020) introduced Greek letters for distinguishing three different types of strength 4 properties of the strength 3+ SOAs, with properties α and γ pertaining to 2D projections, property β to 3D projections:

Definition 3.3 (Properties α, β, γ, Shi and Tang 2020)**.**

- property α: all $s^2 \times s^2$ stratifications in 2D yield s^4 equally-sized strata.
- property β: all $s^2 \times s \times s$ stratifications in 3D yield s^4 equally-sized strata. (9)
- property γ: all $s^3 \times s$ stratifications in 2D yield s^4 equally-sized strata.

Definition 3.4 (Strength 3+)**.** A strength 3 SOA has strength 3+ iff it fulfills all three properties of Definition 3.3.

The following results of this section hold for $s = 2$ only. Shi and Tang (2020) provided necessary and sufficient conditions under which properties α, β and γ of Definition 3.3 are fulfilled (their Proposition 1):

Lemma 3.10 (Shi and Tang 2020)**.** *Let* $\mathbf{D} = s^2\mathbf{A} + s\mathbf{B} + \mathbf{C}$ *an SOA*$(n, m, 8, 3)$*,* $n = 2^k$*, and let* $\mathbf{A}$*,* $\mathbf{B}$ *and* $\mathbf{C}$ *be chosen from the saturated regular OA*$(n, 2^k - 1, 2, 2)$*. The properties (9) are obtained under the following conditions:*

(i) $\mathbf{D}$ *projects into* s^4 *equally-sized* $s^2 \times s^2$ *strata (property* α*)*
iff $(\mathbf{a}_\ell, \mathbf{b}_\ell, \mathbf{a}_j, \mathbf{b}_j)$ *has strength 4 for all* $\ell \neq j$*.*
(ii) $\mathbf{D}$ *projects into* s^4 *equally-sized* $s^2 \times s \times s$ *strata (property* β*),*
iff $(\mathbf{a}_\ell, \mathbf{a}_j, \mathbf{a}_u, \mathbf{b}_u)$ *has strength 4 for all triples* (ℓ, j, u) *with distinct elements.*
(iii) $\mathbf{D}$ *projects into* s^4 *equally-sized* $s^3 \times s$ *strata,*
iff $(\mathbf{a}_\ell, \mathbf{a}_j, \mathbf{b}_j, \mathbf{c}_j)$ *has strength 4 for all* $\ell \neq j$ *(property* γ*).*

Let us briefly consider the performance of the strength 3 construction of He and Tang (2013; see Table 3) with a regular fractional factorial 2-level matrix $\mathbf{V}$ w.r.t. the criteria: Part (i) of the lemma implies that this construction cannot produce SOAs with property α, because all columns of $\mathbf{B}$ are identical (up to level permutation). Part (ii) implies that the construction produces an SOA with property β iff all quadruples of columns of $\mathbf{V}$ that contain column $\mathbf{v}_m$ have strength 4. Part (iii) implies that the construction cannot produce an SOA with property γ for all quadruples $(\mathbf{a}_\ell, \mathbf{a}_j, \mathbf{b}_j, \mathbf{c}_j)$, because $\mathbf{a}_\ell = \mathbf{c}_j$ for some pairs (ℓ, j).

3.3 OAs as OSOAs

We now take a brief look at using OAs as SOAs. First, note that any OA$(n, m, \ell, 2)$ has orthogonal columns. In the following, even though there are no underlying s-level construction matrices (at least not at first view), we consider OAs with s^k levels relative to the underlying s, when considering their strengths as SOAs.

(i) Any OA$(n, m, s^2, 2)$ is per construction also an OSOA$(n, m, s^2, 2+)$, and it additionally fulfills property α.
(ii) An OA$(n, m, s^2, 2)$ may be an OSOA$(n, m, s^2, 3-)$ or even an OSOA$(n, m, s^2, 3+)$ (see example below).
(iii) An OA$(n, m, s^2, 3)$ is an OSOA$(n, m, s^2, 3-)$ and additionally fulfills property α (and further balance properties for which no labels have been defined in the SOA literature).
(iv) Any OA$(n, m, s^3, 2)$ is per construction also an OSOA$(n, m, s^3, 2^*)$ that fulfills properties α and γ.
(v) Any OA$(n, m, s^3, 3)$ is per construction also an OSOA$(n, m, s^3, 3+)$.

For example, an OA$(81, 10, 9, 2)$ with $s = 3$ is an OSOA$(81, 10, 9, 2+)$ and fulfills property α; similarly, an OA$(729, 28, 27, 2)$ with $s = 3$ is an OSOA$(729, 28, 27, 2^*)$. The 16 8-level columns from the OA$(128, 8^{16}16^1, 2)$ from the Kuhfeld (2010) collection with $s = 2$ are an OSOA$(128, 16, 8, 3+)$ $(s = 2)$ and are thus an example for (ii). An OA$(729, 10, 9, 3)$ with $s = 3$ is an OSOA$(729, 10, 9, 3-)$ with property α (and further balance properties for which no labels have been defined in the SOA literature).

The literature provides many constructions for symmetric OAs whose numbers of levels are powers of a prime, for example by Bose (1947), Bush (1952), Bose and Bush (1952), or Addelman and Kempthorne (1961). These are typically not stated in terms of matrix equations. Points (ii), (iv) and (v) above imply that all OA(n, m, s^2, t) can be represented by Equation (2) and all OA(n, m, s^3, t) can be represented by Equation (3), where $t = 2$ or $t = 3$. It is straightforward to obtain those matrices from the OA. This observation is interesting, but not of immediate use.

4 Constructions for further (O)SOAs in s^3 levels

Equation (3) is based on the symmetric s level OAs $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. Except for $\mathbf{A}$, their OA strength may be as low as 1. Any construction for (O)SOAs in s^3 level columns can be presented using this universal equation. The strength 3 constructions by He and Tang (2013) and Liu and Liu (2015) were already covered in Section 3.1. This section covers the construction of OSOAs of strength 2^* or 3 by Li, Liu and Yang (2021) as well as the construction of strength 3 or 3+ 8-level SOAs by Shi and Tang (2020).

4.1 Li, Liu and Yang's construction of OSOAs of strength 2^* or 3

Li et al. (2021) proposed a procedural algorithm, based on two OA$(n, m, s, 2)$ called $\mathbf{A}$ and $\mathbf{B}$. The constructions are related to the construction by Liu and Liu (2015; see Section 3.1.2). A key difference is that Li et al. considered separate matrices $\mathbf{A}$ and $\mathbf{B}$ and made more lenient assumptions on $\mathbf{B}$, rather than taking the columns of both these matrices from a single OA $\mathbf{V}$ that is subjected to strong assumptions.

For odd m, the last column of both matrices $\mathbf{A}$ and $\mathbf{B}$ is omitted, so that one can require m to be even. The algorithm constructs OSOA$(n, m, s^3, \text{strength})$ of strengths 2^* or 3. It yields

- strength 3 for $m = n/2 - 2$ columns with 8 levels, where n is a multiple of 8, based on doubled Hadamard matrices,
- strength 2^* for $m' = 2\lfloor m/2 \rfloor$ columns with s^3 levels, based on an arbitrary OA$(n/s, m, s, 2)$, $\mathbf{V}$, say. This has the special case, where $n = s^k$ and $\mathbf{V}$ is the regular saturated OA$(n/s, (s^{k-1}-1)/(s-1), s, 2)$.
- strength 3, if the $\mathbf{V}$ from the previous bullet has strength 3.

Proposition 4.1. *Li et al.'s (2021) algorithm based on the $n \times m$ matrices $\mathbf{A}$ and $\mathbf{B}$ can be restated as follows with $m' = 2 \cdot \lfloor m/2 \rfloor$:*

$$\mathbf{D} = s^2\mathbf{A}_{1:m'} + s\mathbf{B}_{1:m'} + \mathbf{C}$$

with the columns of the $n \times m'$ matrix $\mathbf{C}$ obtained as

$$\mathbf{c}_\ell = \begin{cases} \mathbf{a}_{\ell+1} & \ell \text{ odd} \\ s - 1 - \mathbf{a}_{\ell-1} & \ell \text{ even} \end{cases}, \tag{10}$$

where $s - 1 - \mathbf{a}_{\ell-1}$ indicates a reversal of the levels of column $\mathbf{a}_{\ell-1}$.

This representation will be used here, because it fits in nicely with Equation (3) and related results. Note that Equation (10) implies that $\mathbf{C} = \mathcal{S}(\mathbf{A}_{1:m'})$ with $\mathcal{S}$ from Definition 3.2, like for the Liu and Liu (2015) construction for strength 3. Appendix B contains the proof for the proposition.

Depending on the properties of $\mathbf{A}$ and $\mathbf{B}$, the construction generates OSOA$(n, m', s^3, 2^*)$ or OSOA$(n, m', s^3, 3)$, where $m' = 2\lfloor m/2 \rfloor$. Note that one does not need to assume that $\mathbf{A}$ and $\mathbf{B}$ are subsets of columns from a saturated regular OA. Li et al. provided the following general results:

Lemma 4.1 (Li et al. 2021)**.** *Let $\mathbf{D} = s^2\mathbf{A} + s\mathbf{B} + \mathbf{C}$ with $\mathbf{C}$ chosen according to Equation (10), and let $\mathbf{A}$ and $\mathbf{B}$ be OA$(n, m, s, 2)$.*

(i) If all three-column sets $(\mathbf{a}_i, \mathbf{a}_j, \mathbf{b}_j)$, $i \neq j$, are OA$(n, 3, s, 3)$, $\mathbf{D}$ is an OSOA$(n, m, s^3, 2^*)$ (Theorem 2 in Li et al.).

(ii) If in addition to (i) $\mathbf{A}$ is an OA$(n, m, s, 3)$, $\mathbf{D}$ is an OSOA$(n, m, s^3, 3)$ (Theorem 3 in Li et al.).

4.1.1 Obtain an OSOA from a general OA

Let $\mathbf{V}$ be an OA$(n/s, m, s, 2)$, and $m' = 2\lfloor m/2 \rfloor$. Li et al.'s construction of $n \times m'$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ for obtaining an OSOA$(n, m', s, 2^*)$ via Equation (3) can be stated as follows:

$$\mathbf{A} = (\mathbf{V}^\top, \mathbf{V}^\top +_s 1, \ldots, \mathbf{V}^\top +_s (s-1))^\top, \quad \mathbf{B} = (\mathbf{V}^\top, \mathbf{V}^\top, \ldots, \mathbf{V}^\top)^\top, \tag{11}$$

with $\mathbf{C}$ again obtained from Equation (10). According to Lemma 4.1, $\mathbf{D}$ generally has strength 2^*. If $\mathbf{V}$ has OA strength 3, the OSOA $\mathbf{D}$ also has strength 3. Section 4.1.3 will provide a modification for $\mathbf{A}$ that preserves the benefits of Equation (11) and improves the chances for good space-filling and for obtaining strength 3 even if $\mathbf{V}$ only has OA strength 2.

Li et al. emphasized that the construction (11) does not require s to be a prime power. Thus, it can, e.g., be used for constructing an OSOA$(432, 6, 216, 2^*)$ from the symmetric 6-level portion of the OA$(72, 16, 2^5 3^3 4^1 6^7, 2)$ from Warren Kuhfeld's collection.

4.1.2 Strength 3 constructions from a strength 2 OA

Strength 3 for the OSOA $\mathbf{D}$ does not require that the OA $\mathbf{V}$ in Equation (11) has OA strength 3: it is sufficient that the matrix $\mathbf{A}$ has OA strength 3. This is for example achieved for $s = 2$, whenever $\mathbf{V}$ is the 0/1 version of a Hadamard matrix without the constant column, or more generally, whenever the foldover method is able to turn a strength 2 OA $\mathbf{V}$ into OA strength 3. Thus, strength 3 OSOAs with 8-level columns in n runs can be obtained for up to $n/2 - 2$ columns from a Hadamard matrix construction (see Section 3.2 of Li et al. 2021).

For $s > 2$, the foldover principle no longer holds. Nevertheless, the matrix $\mathbf{A}$ from Equation (11) can achieve OA strength 3 in some cases, and whether or not that happened can at least be diagnosed post-hoc. The next section proposes a modification to the construction of $\mathbf{A}$ that increases the chances of obtaining OA strength 3 for $\mathbf{A}$ from a $\mathbf{V}$ that only had OA strength 2.

4.1.3 Modification of construction (11)

For construction (11), the level permutations by Weng (2014) amount to separate permutations for the levels of $\mathbf{A}$ and $\mathbf{B}$, possibly with an additional permutation of the blocks within matrix $\mathbf{A}$; the construction of matrix $\mathbf{C}$ from $\mathbf{A}$ must always follow Equation (10). A slight generalization of construction (11) permits more efficient level permutations. The key idea is to add column-specific constants in the construction of $\mathbf{A}$:

$$\mathbf{A} = \mathbf{B} +_s \mathbf{M}, \quad \mathbf{B} = (\mathbf{V}^\top, \mathbf{V}^\top, \ldots, \mathbf{V}^\top)^\top, \tag{12}$$

where $\mathbf{M}$ is a matrix that consists of columns $(\pi_{1\ell}\mathbf{1}_{n/s}^\top, \ldots, \pi_{s\ell}\mathbf{1}_{n/s}^\top)^\top$, $\ell = 1, \ldots, m'$, with $\boldsymbol{\pi}_\ell = (\pi_{1\ell}, \ldots, \pi_{s\ell})^\top$ denoting a permutation of $(0, \ldots, s-1)$. This increased freedom in constructing the columns of $\mathbf{A}$, combined with level permutation applied to the columns of $\mathbf{B}$ and $\mathbf{M}$, yields better chances for good space filling without destroying orthogonality. Figure 2 showed an OSOA$(125, 6, 125, 2^*)$ for which the optimization substantially improved the 2D space-filling behavior.

If $\mathbf{V}$ already had OA strength 3, that strength is preserved, like in construction (11), and if $\mathbf{V}$ has OA strength 2, a beneficial pattern of permutations in $\mathbf{M}$ can cause the OA strength of $\mathbf{A}$ to become 3. For example, as an OA$(81, 9, 3, 3)$ exists, using the 9 3-level columns from the mixed-level OA$(27, 3^9 9^1, 2)$ offers the opportunity to achieve a strength 3 OSOA, as was found by trying different seeds until optimization of space filling via level permutation yielded strength 3. In the examples that were inspected, the strength 3 SOAs did not exhibit close to optimal space filling in terms of the ϕ_p value, but rather a value around the upper quartile of ϕ_p values. Thus, higher strength and better space filling in terms of ϕ_p seem to be conflicting targets. This matter has not been explored in depth.

4.2 Shi and Tang's strength 3 SOAs with additional balance properties

Shi and Tang (2020) constructed SOAs from a $2^k \times (2^k - 1)$ saturated regular strength 2 fraction $\mathbf{S}$. The $n \times m'$ matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ ($n = 2^k$) for Equation (3) have columns from that $\mathbf{S}$. Shi and Tang treated 2-level fractions in the $-1/+1$ coding with multiplication that is often used for 2-level fractions. Here, we will use the equivalent 0/1 encoding with $+_2$. The additional balance properties α, β and γ that were introduced by Shi and Tang where already presented in Section 3.2. Emphasis is on the construction of matrices $\mathbf{A}$ and $\mathbf{B}$. The matrix $\mathbf{C}$ can always be obtained according to Lemma 3.4.

4.2.1 $5n/16$ 8-level columns with property α

This section presents Shi and Tang's first family of SOAs that exists for $n \geq 16$ and is based on the following recursive construction.

Lemma 4.2. *Let $\mathbf{A}_k$ and $\mathbf{B}_k$ fulfill all conditions for obtaining an SOA$(2^k, m, 8, 3)$ with property α through Equation (3). Then the matrices $\mathbf{A}_{k+2}$ and $\mathbf{B}_{k+2}$ constructed by*

$$\mathbf{A}_{k+2} = \begin{pmatrix} \mathbf{A}_k & \mathbf{A}_k & \mathbf{A}_k & \mathbf{A}_k \\ \mathbf{A}_k & 1 +_2 \mathbf{A}_k & \mathbf{A}_k & 1 +_2 \mathbf{A}_k \\ \mathbf{A}_k & \mathbf{A}_k & 1 +_2 \mathbf{A}_k & 1 +_2 \mathbf{A}_k \\ \mathbf{A}_k & 1 +_2 \mathbf{A}_k & 1 +_2 \mathbf{A}_k & \mathbf{A}_k \end{pmatrix}$$

and

$$\mathbf{B}_{k+2} = \begin{pmatrix} \mathbf{B}_k & \mathbf{B}_k & \mathbf{B}_k & \mathbf{B}_k \\ \mathbf{B}_k & \mathbf{B}_k & 1 +_2 \mathbf{B}_k & 1 +_2 \mathbf{B}_k \\ \mathbf{B}_k & 1 +_2 \mathbf{B}_k & 1 +_2 \mathbf{B}_k & \mathbf{B}_k \\ \mathbf{B}_k & 1 +_2 \mathbf{B}_k & \mathbf{B}_k & 1 +_2 \mathbf{B}_k \end{pmatrix}.$$

fulfill all conditions for obtaining an SOA$(2^{k+2}, 4m, 8, 3)$ *with property* α *through Equation (3).*

Lemma 4.2 states the recursive construction rule by Shi and Tang in the notation of this paper. Once there are start values for even and for odd k, one can recursively construct all designs for larger values of k. Start values are available for $k = 4$ and $k = 7$, and the case $k = 5$ can be treated separately, so that SOAs with property α are available for $k \geq 4$. Start values are provided in Appendix C.

The recursive construction of Lemma 4.2 is somewhat inconvenient, and it is not straightforward to state it as a non-recursive general formula. However, it is straightforward to state update rules in terms of Yates matrix column numbers, which simplifies considerations and computations. The construction of Lemma 4.2 means that the Yates matrix columns from the smaller design remain in place, and further Yates matrix columns are added according to the following proposition.

Proposition 4.2. *Let* $Y_{\mathbf{A}}$ *and* $Y_{\mathbf{B}}$ *denote the tuples of Yates matrix column numbers of matrices* $\mathbf{A}$ *and* $\mathbf{B}$*, and let* $\mathbf{A}_k$ *and* $\mathbf{B}_k$ *denote the matrices from constructions for* $n = 2^k$*. Then, the matrices for* $n = 2^{k+2}$ *can be obtained with the following Yates matrix tuples:*

$$Y_{\mathbf{A}_{k+2}} = (Y_{\mathbf{A}_k}, Y_{\mathbf{A}_k} + 2^k, Y_{\mathbf{A}_k} + 2^{k+1}, Y_{\mathbf{A}_k} + 2^k + 2^{k+1}),$$
$$Y_{\mathbf{B}_{k+2}} = (Y_{\mathbf{B}_k}, Y_{\mathbf{B}_k} + 2^{k+1}, Y_{\mathbf{B}_k} + 2^k + 2^{k+1}, Y_{\mathbf{B}_k} + 2^k).$$

According to Lemma 3.10 (i), the matrix $\mathbf{C}$ is irrelevant for obtaining property α. With the start values given in Appendix C, it can be observed that the Yates matrix column numbers from Proposition 4.2 do not contain multiples of 16 (excluding the single special case $k = 5$). This also holds for $Y_{\mathbf{A}+_2\mathbf{B}}$. According to Lemma 3.4, it is therefore adequate (though by no means necessary) to choose $\mathbf{C}$ as a matrix with all columns equal to one of those Yates matrix columns. (For $k = 4$, there are no such Yates matrix columns; Appendix C suggests one of many possible solutions for that case.)

4.2.2 Strength 3+ SOAs in $n/4-1$ 8 level columns, or one more column without property γ

The construction of this section is again a recast of Shi and Tang's constructions in different notation. For $n = 2^k$, Shi and Tang proposed an SOA$(n, n/4, 8, 3)$ with properties α and β (their family 2) and an SOA$(n, n/4 - 1, 8, 3+)$ (their family 3). The two constructions are closely related and are therefore presented together.

The starting point is a saturated regular OA$(n/4, n/4 - 1, 2, 2)$ called $\mathbf{X}$ that is based on $k - 2$ basic vectors. Let $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_{n/4-1})$ be a reshuffling of the columns of $\mathbf{X}$ such that

- $\mathbf{x}_\ell$ and $\mathbf{y}_\ell$ are different for $\ell = 1, \ldots, m$
- for $\mathbf{Z} = \mathbf{X} +_2 \mathbf{Y}$, the triple $(\mathbf{x}_\ell, \mathbf{y}_\ell, \mathbf{z}_\ell)$ has strength at least 2 for $\ell = 1, \ldots, m$.

Shi and Tang proved that such a $\mathbf{Y}$ can be found. The start vectors for $k = 2$ and $k = 3$ for the recursive construction are given in Lemma 4.3 in terms of Yates matrix column numbers, and the subsequent proposition provides the recursive construction:

Lemma 4.3 (restated from Shi and Tang 2020)**.** *Let* $Y_{\mathbf{M}}$ *denote the vector of Yates matrix column numbers for a matrix* $\mathbf{M}$*. The start values of the Shi and Tang construction for their families 2 and 3 are given as follows:*
For $k = 2$*,* $Y_{\mathbf{X}} = 123$*,* $Y_{\mathbf{Y}} = 231$*, and* $Y_{\mathbf{Z}} = 312$*.*
For $k = 3$*,* $Y_{\mathbf{X}} = 1234567$*,* $Y_{\mathbf{Y}} = 7521643$ *and* $Y_{\mathbf{Z}} = 6715324$*.*

Proposition 4.3 (restated from Shi and Tang 2020)**.** *Let* $Y_{\mathbf{X}_k}$*,* $Y_{\mathbf{Y}_k}$ *and* $Y_{\mathbf{Z}_k}$ *denote the tuples of Yates matrix column numbers for* $2^k \times (2^k - 1)$ *matrices* $\mathbf{X}_k$*,* $\mathbf{Y}_k$ *and* $\mathbf{Z}_k$*, such that* $\mathbf{x}_\ell +_2 \mathbf{y}_\ell = \mathbf{z}_\ell$ *and* $(\mathbf{x}_\ell, \mathbf{y}_\ell, \mathbf{z}_\ell)$

have at least strength 2, $\ell = 1\ldots, 2^k - 1$*. Then,* $2^{k+2} \times (2^{k+2} - 1)$ *matrices* $\mathbf{X}_{k+2}$*,* $\mathbf{Y}_{k+2}$ *and* $\mathbf{Z}_{k+2}$ *with the same properties can be obtained using the following Yates matrix tuples:*

$$
\begin{aligned}
Y_{\mathbf{X}_{k+2}} &= (Y_{\mathbf{X}_k}, 2^k, Y_{\mathbf{X}_k} + 2^k, 2^{k+1}, Y_{\mathbf{X}_k} + 2^{k+1}, 2^k + 2^{k+1}, Y_{\mathbf{X}_k} + 2^k + 2^{k+1}),\\
Y_{\mathbf{Y}_{k+2}} &= (Y_{\mathbf{Y}_k}, 2^{k+1}, Y_{\mathbf{Y}_k} + 2^{k+1}, 2^k + 2^{k+1}, Y_{\mathbf{Y}_k} + 2^k + 2^{k+1}, 2^k, Y_{\mathbf{Y}_k} + 2^k),\\
Y_{\mathbf{Z}_{k+2}} &= (Y_{\mathbf{Z}_k}, 2^k + 2^{k+1}, Y_{\mathbf{Z}_k} + 2^k + 2^{k+1}, 2^k, Y_{\mathbf{Z}_k} + 2^k, 2^{k+1}, Y_{\mathbf{Z}_k} + 2^{k+1}).
\end{aligned}
$$

Together with the start tuples $Y_{\mathbf{X}_2}, Y_{\mathbf{Y}_2}$ *or* $Y_{\mathbf{X}_3}, Y_{\mathbf{Y}_3}$ *from Lemma 4.3, a recursive construction for all* $k \geq 2$ *is completely specified.*

The saturated $\mathbf{X}$ is always in original Yates order, i.e. $Y_{\mathbf{X}_k} = (1, \ldots, 2^k - 1)$. Algorithmically, it suffices to construct the reshuffled $\mathbf{Y}$, since $\mathbf{Z}$ is a direct consequence and is also not needed for the construction (see below).

Example. Applying Proposition 4.3 for obtaining the 16×15 matrices (i.e. $k + 2 = 4$) yields

- columns 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15 for $\mathbf{X}$
- columns 2,3,1,8,10,11,9,12,14,15,13,4,6,7,5 for $\mathbf{Y}$ and consequently
- columns 3,1,2,12,15,13,14,4,7,5,6,8,11,9,10 for $\mathbf{Z}$.

The following lemma states the construction of the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ for Equation (3) from the matrices $\mathbf{X}$ and $\mathbf{Y}$.

Lemma 4.4 (restated from Shi and Tang 2020)**.** *Let* $\mathbf{X}$ *and* $\mathbf{Y}$ *be* $2^{k-2} \times (2^{k-2} - 1)$ *matrices according to Proposition 4.3.*

(i) $2^k \times 2^{k-2}$ *matrices for constructing Shi and Tang's Family 2 from Equation (3) are given as*

$$
\mathbf{A} = \begin{pmatrix} \mathbf{0}_{n/4} & \mathbf{X} \\ \mathbf{0}_{n/4} & \mathbf{X} \\ \mathbf{1}_{n/4} & 1 +_2 \mathbf{X} \\ \mathbf{1}_{n/4} & 1 +_2 \mathbf{X} \end{pmatrix}, \quad \textit{and} \quad \mathbf{B} = \begin{pmatrix} \mathbf{0}_{n/4} & \mathbf{Y} \\ \mathbf{1}_{n/4} & 1 +_2 \mathbf{Y} \\ \mathbf{0}_{n/4} & \mathbf{Y} \\ \mathbf{1}_{n/4} & 1 +_2 \mathbf{Y} \end{pmatrix}. \tag{13}
$$

(ii) The construction (i) yields an SOA$(n, n/4, 8, 3)$ *with properties* α *and* β*, if* $\mathbf{c}_\ell$ *is chosen as a column from the saturated OA*$(2^k, 2^k - 1, 2, 2)$ *that is unequal to all three of* $\mathbf{a}_\ell$*,* $\mathbf{b}_\ell$ *and* $\mathbf{a}_\ell +_2 \mathbf{b}_\ell$*.*

(iii) If each first column is omitted, and one of them is used instead for each column of $\mathbf{C}$*, the result is an SOA*$(n, n/4 - 1, 8, 3+)$ *corresponding to Shi and Tang's Family 3.*

The design construction of Equation (13) for a design with $n = 2^k$ rows based on $n/4 \times n/4 - 1$ matrices $\mathbf{X}$ and $\mathbf{Y}$ and (denoting the Yates matrix column tuple for $\mathbf{Y}$ as $Y_{\mathbf{Y}}$) can be stated as follows in Yates matrix notation:

Corollary 4.1. *The construction of Lemma 4.4 is equivalent to using the following tuples of Yates matrix columns for constructing the matrices (omitting the respective first column for the strength* 3+ *construction):*

- $Y_{\mathbf{A}} = (n/2, n/2 + Y_{\mathbf{X}}) = (n/2, n/2 + 1, \ldots, 3n/4 - 1)$,
- $Y_{\mathbf{B}} = (n/4, n/4 + Y_{\mathbf{Y}})$,
- $Y_{\mathbf{A}+_2\mathbf{B}} = (n/2 + n/4, n/2 + n/4 + Y_{\mathbf{Z}})$.

Proof. The Yates column numbers follow from careful considerations regarding the structure of Yates matrices. □

The following lemma proposes a choice of columns for $\mathbf{C}$ such that the strength 3+ SOA of Shi and Tang's Family 3 becomes an OSOA, and the strength 3 SOA of Shi and Tang's Family 2 has a single pair of correlated columns only.

Lemma 4.5. *The following column choices for the constructions of Shi and Tang's families 2 and 3 are beneficial for obtaining orthogonal columns:*

(i) For Family 3, choosing **C** *as the matrix of the first* $n/4-1$ *Yates columns guarantees that an OSOA*$(n, n/4-1, 8, 3+)$ *is obtained.*

(ii) For Family 2, choosing **C** *as a matrix of the first* $n/4-1$ *Yates columns with exactly one column duplicated guarantees that there is only a single column pair with non-zero correlation.*

Proof. The proposed column choices fulfill the assumptions of Lemma 3.4. (i) follows from Lemma 3.5. (ii) also follows from that lemma, if one realizes that any $n/4-1$ column sub matrix that does not contain the pair with the same **C** column has orthogonal columns according to the lemma. □

Appendix D provides two small applications of the construction.

5 SOA constructions with m columns in s^2 levels

This section uses the construction from Lemma 3.2 with Equation (2). Lemmata 3.7, 3.8 and 3.9 gave conditions for a strength 2+ SOA, an OSOA or a strength 3− SOA. The constructions of this section are based on these lemmata.

5.1 Constructions by Zhou and Tang (2019)

The constructions by Zhou and Tang (2019) are similar to the constructions by Li et al. (2021), which were of course developed later, but were already presented in the previous section. Basically, the matrices **A** and **B** from the Li et al. constructions are used, and the unnecessary matrix **C** is omitted. Generally, the constructions by Zhou and Tang have the same number of columns or one more column, and yield s^2 levels instead of s^3.

Where Li et al. obtained an OSOA$(n, n/2-2, 8, 3)$ from a doubled Hadamard matrix with $n/2$ rows, Zhou and Tang obtained an OSOA$(n, n/2-1, 4, 3-)$.

Where Li et al. obtained an OSOA$(s^k, 2\lfloor(s^{k-1}-1)/(2(s-1))\rfloor, s^3, 2^*)$ from a regular saturated OA in s^{k-1} runs, Zhou and Tang obtained an OSOA$(s^k, (s^{k-1}-1)/(s-1), s^2, 2+)$.

Where Li et al. obtained an OSOA$(n, 2\lfloor m/2\rfloor, s^3, 3)$ from an OA$(n/s, m, s, 3)$ or an OSOA$(n, 2\lfloor m/2\rfloor, s^3, 2^*)$ from an OA$(n/s, m, s, 2)$, Zhou and Tang obtained an OSOA$(n, m, s^2, 3-)$ (although they did not claim that strength for their general construction) or an OSOA$(n, m, s^2, 2+)$. For this construction, they used **A** and **B** in switched roles (their Theorem 4), which can be improved upon: Using **A** and **B** from Equation (11) in Equation (2), combined with the construction of **A** with the modification presented in Equation (12), it is sometimes possible to achieve a strength 3− OSOA in spite of using a matrix **V** with OA strength 2, e.g. when using an OA$(9, 3, 3, 2)$ in the role of **V**.

5.2 He et al.'s construction of strength 2+ SOAs for regular 2-level fractions

He, Cheng and Tang (2018) provided a construction based on regular 2-level fractions. The columns for both **A** and **B** are chosen from a saturated regular 2-level array **S**. He et al. (2018) proved that it is necessary and sufficient for strength 2+ that the columns of **S** that are not used in **A** make up an SOS design **X**, where an SOS design is characterized as follows: all columns of **S** that do not belong to a main effect of **X** contain a two-factor interaction of a pair of columns in **X**. $\mathbf{A} = \overline{\mathbf{X}}$ (where the overline denotes the complement in **S**) is the largest possible choice of **A** for a given SOS design **X**, and suitable columns for **B** must be picked from **X**. According to Lemma 3.8, populating **B** with distinct columns from **X** (if possible) makes **D** an OSOA; a pair-matching algorithm for bipartite graphs can help to find distinct columns for **B**. If **A** has OA strength 3, the resulting (O)SOA **D** has strength 3−. However, it is not obvious how to ensure OA strength 3 for **A** in a systematic way.

5.2.1 Constructions for SOS designs

He et al. gave four constructions for an SOS design as follows: For a total of $k \geq 4$ independent columns, let $P = P(\{\mathbf{a}_1, \ldots, \mathbf{a}_{k_1}\})$ denote the set of all effects pertaining to $k_1 \geq 2$ columns (i.e., $2^{k_1}-1$ elements), $Q = Q(\{\mathbf{b}_1, \ldots, \mathbf{b}_{k_2}\})$ the set of all effects pertaining to the remaining $k_2 = k - k_1 \geq 2$ columns (i.e., $2^{k-k_1}-1$ elements). Then, the following column choices yield SOS designs (where $+_2$ between a column and a set denotes the set of separate sums):

(i) $C_1 = P \cup Q$ ($2^{k_1} + 2^{k-k_1} - 2$ elements),
(ii) $C_2 = (P - \{\mathbf{a}_1\}) \cup (Q - \{\mathbf{b}_1\}) \cup \{\mathbf{a}_1 +_2 \mathbf{b}_1\}$ ($2^{k_1} + 2^{k-k_1} - 3$ elements),
(iii) $C_3 = (P - \{\mathbf{a}_1\}) \cup (\mathbf{a}_1 +_2 Q)$ ($2^{k_1} + 2^{k-k_1} - 3$ elements),
(iv) $C_4 = (\mathbf{b}_1 +_2 P) \cup (\mathbf{a}_1 +_2 (Q - \{\mathbf{b}_1\}))$ ($2^{k_1} + 2^{k-k_1} - 3$ elements).

The minimum possible number of columns for an SOS design determines the maximum possible number m_k of columns for the SOA from this construction. He et al. stated that $2^k - 2^{\lfloor k/2 \rfloor} - 2^{k-\lfloor k/2 \rfloor} + 2 \leq m_k \leq 2^k - 1 - M(k)$, with $M(k)$ the maximum number of columns in a strength 4 OA. The upper bound in that inequality can be slightly tightened by realizing that the number of columns m in an SOS design must fulfill the quadratic inequality $m + m(m-1)/2 \geq 2^k - 1$ for m. It is likely that incorporation of structural requirements would lead to further tightening of the upper bound for m_k.

5.2.2 Implementation of the construction

The implementation of the construction has the following steps:

- Allocate P and Q with k_1 and k_2 columns, $k_1 + k_2 = k$, $|k_1 - k_2| \leq 1$; this choice of k_1 and k_2 minimizes the number of columns of the SOS design.
- Define R_{SOS} as one of C_2, C_3 or C_4.
- Obtain $\mathbf{A}$ as the matrix of the columns from the saturated regular $2^k \times (2^k - 1)$ array $\mathbf{S}$ that are not in R_{SOS}, and define A as the set of the columns of $\mathbf{A}$.
- For each column $\mathbf{a}_j \in A$, define the set S_j as those columns $\mathbf{c} \in R_{SOS}$ that yield a triple of OA strength 3 when added to any pair $\{\mathbf{a}_\ell, \mathbf{a}_j\} \subset A$, $\ell \neq j$ that involves column $\mathbf{a}_j$.
- Define a bipartite graph G with the vertices from A (type 1) and R_{SOS} (type 2) and edges between $\mathbf{a}_j$ and all elements of S_j. (There is at least one edge for each element of A.)
- Create $\mathbf{B}$ from the columns $\mathbf{b}_j \in R_{SOS}$ according to the following matching approach:
  - Match a $\mathbf{b}_j$ to each $\mathbf{a}_j \in A$, using a maximum bipartite matching algorithm on the graph G. *If this step matches all columns of A, $\mathbf{B}$ will have strength 2 and column orthogonality will be achieved.*
  - For any unmatched $\mathbf{a}_j \in A$, assign an arbitrary element of S_j as $\mathbf{b}_j$.
- Apply the algorithm by Weng (2014) for improving ϕ_p, permuting levels in the columns of $\mathbf{A}$ and $\mathbf{B}$.
- Return $\mathbf{D} = s\mathbf{A} + \mathbf{B}$ from the optimized permutation pattern.

5.3 He et al.'s construction of strength 2+ SOAs for regular s level fractions ($s \geq 3$)

The construction of this section works for primes or prime powers $s \geq 3$ via a saturated regular fraction $\mathbf{S}$ which is an OA$(s^k, (s^k-1)/(s-1), s, 2)$, $k \geq 3$ (see Section 2.3.1 for the construction of $\mathbf{S}$). An SOA obtained by this construction has s^k runs with up to $m = (s^k-1)/(s-1) - ((s-1)^k - 1)/(s-2)$ columns, e.g. six 9-level columns in 27 runs ($s = 3$, $k = 3$), eight 16-level columns in 64 runs ($s = 4$, $k = 3$), 25 9-level columns in 81 runs ($s = 3$, $k = 4$), or 45 16-level columns in 256 runs ($s = 4$, $k = 4$). The necessary and sufficient conditions for the existence of an SOA of strength 2+ were given in Lemma 3.7.

The following coarse implementation steps can be followed for an implementation that guarantees orthogonality (through a strength 2 matrix $\mathbf{B}$, according to Lemma 3.8), whenever that is compatible with strength 2+ for the requested number of runs and columns. The details are very similar to Section 5.2.2 and are therefore omitted.

- The matrix $\mathbf{A}$ is populated with the $m = (s^k-1)/(s-1) - ((s-1)^k - 1)/(s-2)$ linear combinations of at least two of the k basic columns (contained in $\mathbf{S}$) whose first coefficient is 1 (holds for all columns of $\mathbf{S}$) and whose further coefficients include at least one element $s-1$. The set of the columns of $\mathbf{A}$ is denoted as A.
- The set of the remaining columns of $\mathbf{S}$ is denoted as R.
- The elements of A and R make up the two types of vertices in a bipartite graph G.
- The matrix $\mathbf{B}$ is populated with a selection of the elements of R: $\mathbf{b}_j$ is chosen to yield a triple of OA strength 3 with any pair $(\mathbf{a}_\ell, \mathbf{a}_j), j \neq \ell$ (cf. Lemma 3.7). Like in Section 5.2, a set S_j of permissible columns is identified for each position j, and G has edges between $\mathbf{a}_j$ and all elements of S_j. Use of an algorithm for maximal matching in bipartite graphs ensures that an SOA with many pairs of orthogonal columns will be obtained.
- The SOA$(s^k, m, s^2, 2+)$ is then obtained as $\mathbf{D} = s\mathbf{A} + \mathbf{B}$.

- Optimization of permutations in columns of **A** and **B** improves the space filling behavior.

6 Overview of sizes, strengths and constructions

Of course, the larger the strength, the larger the run size requirements for the same number of levels. Aspects in the choice of a suitable (O)SOA are

- the affordable run size n
- the required number of columns m
- the required number of levels per column s^r
- and the required balance properties, reflected by the strength or column orthogonality.

Table 5 lists the different constructions that are covered in this paper. Tables 6, 7 and 8 give the numbers of columns for the different constructions for some n and $s = 2$ to $s = 4$. For $s = 3$ and $s = 4$, the maximum sizes of the relevant OAs underlying the constructions have been taken from the MinT database (Schmid, Schürer and others). The OAs are available in R package **DoE.base**. The tables show that the strength 2+ SOAs by He, Cheng and Tang have a lot more columns than the strength 3− ($s = 2$) or 2+ ($s > 2$) OSOAs by Zhou and Tang, i.e. column orthogonality comes at a cost for this strength. For the strength 3 (O)SOAs, the numbers of columns obtainable with and without orthogonality are almost identical. Here, Li, Liu and Yang have increased the number of columns obtainable with s^3 levels by dropping the 3D projection properties (strength 2^*). On the other end, the OSOAs by Liu and Liu (2015) are available for smaller numbers of columns only; this is the price for their attractive feature of 3-orthogonality, which is beneficial for the estimation of second order linear models, so that their construction is worth being studied. For 2-level factors, Shi and Tang constructed strength 3+ SOAs that have the 2D and 3D balance properties of strength 4 SOAs. As the number of columns that can be accommodated in a strength 4 SOA is very small (e.g. 7 columns in 729 runs for $s = 3$), strength 3+ is quite attractive. Unfortunately, there are so far no strength 3+ constructions for $s > 2$.

Table 6 is limited to regular fractions or tables based on Hadamard matrices. Where strong non-regular OAs exist, more columns may be possible for the He and Tang construction. Known such cases are strength 4 SOAs with 7 columns in 16 levels from an $OA(128, 15, 2, 4)$, or 9 columns in 16 levels from an $OA(256, 19, 2, 4)$; both these OAs can be found in Mee 2009 and are available in R package DoE.base. If more non-regular arrays are found, more such cases will arise.

Table 5: Overview of the construction methods of this paper. HT=He and Tang (2013), LL=Liu and Liu (2015), HCT=He, Cheng and Tang (2018), LLY=Li, Liu and Yang (2021), ZT=Zhou and Tang (2019), ST=Shi and Tang (2020). p^q indicates a prime or prime power; where this is restricted to $\neq 2$, this indicates that the construction is not recommended for $s = 2$ because there are better possibilities for that case. The LL construction yields 3-orthogonal columns for strengths 3 and 4. The HCT construction achieves orthogonal columns under some circumstances. The entries for the LLY and ZT constructions assume that **A** is constructed according to the modification (12). The ST constructions enable additional balance properties, even where strength 3+ is not achieved; the construction of the matrix **C** is assumed to follow this paper.

Eq.	levels	input	s	rule for s	n	m'	t	OSOA	Source
(2)	s^2	$\text{OA}(n, m, s, 2)$	s	none	n	m	2	no	HT 2013
(3)	s^3	$\text{OA}(n, m, s, 3)$	s	none	n	$m - 1$	3	no	HT 2013
Table 3	s^4	$\text{OA}(n, m, s, 4)$	s	none	n	$\lfloor m/2 \rfloor$	4	no	HT 2013
Table 3	s^5	$\text{OA}(n, m, s, 5)$	s	none	n	$\lfloor (m-1)/2 \rfloor$	5	no	HT 2013
(2)	s^2	$\text{OA}(n, m, s, 2)$	s	none	n	$2\lfloor m/2 \rfloor$	2	yes	LL 2015
(3)	s^3	$\text{OA}(n, m, s, 3)$	s	none	n	$2\lfloor m/4 \rfloor$ or $2\lfloor m/4 \rfloor + 1$	3	yes	LL 2015
Table 3	s^4	$\text{OA}(n, m, s, 4)$	s	none	n	$2\lfloor m/4 \rfloor$	4	yes	LL 2015
(2)	4	2, k, (m)	2	2	2^k	$2^k - 2^{\lfloor k/2 \rfloor} - 2^{k - \lfloor k/2 \rfloor} + 2$	2+	n/y	HCT 2018
(2)	s^2	s, k, (m)	s	$p^q \neq 2$	s^k	$\dfrac{s^k - 1}{s - 1} - \dfrac{(s-1)^k - 1}{s - 2}$	2+	n/y	HCT 2018
(3)	s^3	$\text{OA}(n/s, m, s, 2)$	s	none	n	$2 \cdot \lfloor m/2 \rfloor$	2* or 3	yes	LLY 2021
(3)	s^3	$\text{OA}(n/s, m, s, 3)$	s	none	n	$2 \cdot \lfloor m/2 \rfloor$	3	yes	LLY 2021
(3)	8	m and/or n	2	2	$8 \cdot \left\lceil \dfrac{m+2}{4} \right\rceil$	$n/2 - 2$	3	yes	LLY 2021
(3)	s^3	s, k, (m)	s	$p^q \neq 2$	s^k	$2 \cdot \left\lfloor \dfrac{s^{k-1} - 1}{2(s-1)} \right\rfloor$	2* or 3	yes	LLY 2021
(2)	s^2	$\text{OA}(n/s, m, s, 2)$	s	none	n	m	2+ or 3−	yes	ZT 2019
(2)	4	m and/or n	2	2	$8 \cdot \lceil (m+1)/4 \rceil$	$n/2 - 1$	3−	yes	ZT 2019
(2)	s^2	s, k, (m)	s	$p^q \neq 2$	s^k	$\dfrac{s^{k-1} - 1}{s - 1}$	2+ or 3−	yes	ZT 2019
(3)	8	n, (m)	2	2	2^k, ≥ 16	$5n/16$	3	no	ST 2020
(3)	8	n, (m)	2	2	2^k, ≥ 16	$n/4$	3	no	ST 2020
(3)	8	n, (m)	2	2	2^k, ≥ 16	$n/4 - 1$	3+	yes	ST 2020

Table 6: Achievable column numbers for (O)SOAs from regular 2-level fractional factorials or Hadamard matrices.

	4 levels			8 levels					16 levels
n	HT, 2	HCT, 2+	ZT, 3-	HT, 3	LLY, 3	ST, 3	ST, 3+	LL, 3	HT, 4
16	15	10	8	7	6	5	3	4	2
32	31	22	16	15	14	10	7	8	3
64	63	50	32	31	30	20	15	16	4
128	127	106	64	63	62	40	31	32	5
256	255	226	128	127	126	80	63	64	8
512	511	466	256	255	254	160	127	128	11
1024	1023	962	512	511	510	320	255	256	16

Note:
HT=He and Tang 2013, HCT=He, Cheng and Tang 2018, LL=Liu and Liu 2015, LLY=Li, Liu and Yang 2021, ST=Shi and Tang 2020, ZT=Zhou and Tang 2019.

Table 7: Achievable column numbers for (O)SOAs from 3-level fractional factorials.

	9 levels			27 levels				81 levels
n	HT, 2	HCT, 2+	ZT, 2+	LL, 2*	HT, 3	LLY, 3	LL, 3	HT, 4
81	40	25	13	12	9	10	4	2
243	121	90	40	40	19	20	10	5
729	364	301	121	120	55	56	28	7
2187	1093	966	364	364	111	112	56	13
6561	3280	3025	1093	1092	247	248	124	20

Note:
HT=He and Tang 2013, HCT=He, Cheng and Tang 2018, LL=Liu and Liu 2015, LLY=Li, Liu and Yang 2021, ZT=Zhou and Tang 2019.

Table 8: Achievable column numbers for (O)SOAs from 4-level fractional factorials.

	16 levels			64 levels				256 levels
n	HT, 2	HCT, 2+	ZT, 2+	LL, 2*	HT, 3	LLY, 3	LL, 3	HT, 4
256	85	45	21	20	16	16	8	2
1024	341	220	85	84	40	40	20	5
4096	1365	1001	341	340	125	126	62	10

Note:
HT=He and Tang 2013, HCT=He, Cheng and Tang 2018, LL=Liu and Liu 2015, LLY=Li, Liu and Yang 2021, ZT=Zhou and Tang 2019.

7 Discussion

SOAs and OSOAs of practical importance exist in many varieties: one can obtain LHDs by e.g. constructing an $\text{OSOA}(729, 10, 729, 2^*)$ from an $\text{OA}(81, 10, 9, 2)$, or an $\text{OSOA}(512, 8, 512, 2^*)$ from an $\text{OA}(64, 9, 8, 2)$. A 3-orthogonal $\text{OSOA}(512, 4, 512, 3)$ by the Liu and Liu (2015) construction, obtained from an $\text{OA}(512, 9, 8, 3)$, is a very good choice for inspecting four factors in detail. The classical computer experimentation with relatively few quantitative factors will benefit from such LHD-like (O)SOAs.

(O)SOAs that are not LHDs themselves can be used for creating LHDs by level expansion; it has not been investigated inhowfar this brings an advantage over direct expansion from an OA.

Exploration of responses for many quantitative variables can be supported by (O)SOAs with smaller numbers of levels, e.g. 4, 8, 9, 16 or 27 levels. For up to $n/4 - 1$ 8-level columns, strength 3+ OSOAs may be attractive because of their stratification properties and their column orthogonality; the latter can be guaranteed by a modification proposed in this paper. The strength 3+ property has so far only been implemented for $s = 2$, i.e., for 8-level columns. Strength 3+ (O)SOAs with more columns than strength 3 OAs or strength 4 SOAs for $s > 2$ would be a very attractive invention. Strength 3 or stronger OSOAs by Liu and Liu (2015) are attractive because of their 3-orthogonal columns. Where OAs with OA strength $t \geq 2$ exist for the number of levels and columns under consideration, these may be preferable to SOAs (see also Section 3.3).

If a quantitative variable is easy to realize at different levels, it may be attractive to use designs that offer the chance to learn something about the functional form of the response surface by providing more levels than the usual orthogonal arrays, even if one does not use LHDs. (O)SOAs can ensure that, while at the same time preserving attractive projection properties over coarser grids. Many questions remain open with respect to practical usefulness of the various types of (O)SOAs, relative to other types of arrays like level-expanded OAs or LHDs. Their investigation will become easier once SOAs are implemented in software, which is work in progress. It is planned to investigate the relative merits of different (O)SOAs and OAs in a different piece of work.

He and Tang (2014) pointed out the existence of some further SOAs without providing explicit constructions; these have not been included. Furthermore, this paper did not discuss sliced SOAs (Liu and Liu 2015) or nearly strong OSOAs (Li, Liu and Yang 2021b). Like in most of the SOA literature, the mixed level case was also completely ignored; He, Cheng and Tang (2018) are among the few authors who devoted a small part of their paper to that case.

The SOA construction by He, Cheng and Tang (2018) permits particularly large numbers of columns with s^2 levels to be accommodated with SOA strength 2+. Jiang, Wang and Wang (2021) presented a construction based on Addelman and Kempthorne (1961) OAs whose number of runs is twice an odd prime power. Their construction supplements the strength 2+ SOAs in s^2 levels and s^k runs from the HCT construction with strength 2+ SOAs in s^2 levels and $2s^k$ runs, e.g. 9, 43 or 165 9-level columns in 54 runs, 162 or 486 runs. The construction works with ingredients of the Addelman and Kempthorne construction, and it is not obvious how to use a ready-made Addelman and Kempthorne array for obtaining the construction from that array. It will hopefully be possible to translate the construction into a rule for the columns to pick for **A** and **B** from an Addelman and Kempthorne array.

If one would consider the columns of an SOA for experimentation with qualitative factors, many SOAs would be supersaturated, i.e. would require the estimation of more main effects coefficients than there are runs. This is usually not the intention. It may be possible to collapse selected SOA columns and use them for a qualitative factor with fewer levels, while using most columns for quantitative variables, e.g. in linear models with low order polynomials. It is a benefit of (O)SOAs that they can also be used for obtaining insights at coarser discretized versions of the quantitative experimental variables, due to their stratification properties.

This discussion closes with another appeal to replace the misleading expression “strong orthogonal arrays” with the more adequate “stratum orthogonal arrays”, since the OA strength of an SOA with SOA strength 4 may easily be 1 only.

8 References

Addelman, S. and Kempthorne, O. (1961). Some main-effect plans and orthogonal arrays of strength two. *Annals of Mathematical Statistics* **4**, 1167-1176.

Bingham, D., Sitter, R. and Tang, B. (2009). Orthogonal and nearly orthogonal designs for computer experiments. *Biometrika* **96**, 51-65.

Bose, R.C. (1938). On the application of Galois fields to the problem of construction of Hyper-Graeco-Latin Squares. *Sankhya* **3**, 323-338.

Bush, K.A. (1952). Orthogonal arrays of index unity. *Annals of Mathematical Statistics* **23**, 426-434.

Bose, R.C. and Bush, K.A. (1952). Orthogonal arrays of strength two and three. *Annals of Mathematical Statistics* **23**, 508-524.

Grömping, U. (2021). R package **DoE.base**, version 1.2. (used for OAs and utility functions; it will likely be possible to drop the dependence on this package)

He, Y., Cheng, C.S. and Tang, B. (2018). Strong orthogonal arrays of strength two plus. *The 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.

He, Y. and Tang, B. (2014). A characterization of strong orthogonal arrays of strength three. *The Annals of Statistics* **42**, 1347-1360.

Hedayat, A.S., Sloane, N.J.A. and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications.* Springer-Verlag, New York. https://doi.org/10.1007/978-1-4612-1478-6.

Jiang, B., Wang, Z. and Wang, Y. (2021). Strong orthogonal arrays of strength two-plus based on the Addelman-Kempthorne method. *Statistics and Probability Letters* **175**, 109-114.

Li, W., Liu, M.-Q. and Yang, J.-F. (2021, in press). Construction of column-orthogonal strong orthogonal arrays. *Statistical Papers* https://doi.org/10.1007/s00362-021-01249-w.

Li, W., Liu, M.-Q. and Yang, J.-F. (2021). Construction of column-orthogonal nearly strong orthogonal arrays. *Journal of Statistical Planning and Inference* **215**, 184-192.

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

Mee, R. (2009). *A Comprehensive Guide to Factorial Two-level Experimentation.* Springer-Verlag, Berlin.

Schmid, W.Ch. and Schürer, R. and others (ongoing, accessed August 26 2021). MinT – A web based database for querying optimal (t, m, s)-net parameters. http://mint.sbg.ac.at/index.php.

Shi, L. and Tang, B. (2020). Construction results for strong orthogonal arrays of strength three. *Bernoulli* **26**, 418-431. https://doi.org/10.3150/19-BEJ1130

Tang, B. (1993). Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association* **88**, 1392-1397.

Wang, Y., Sun, F. and Xu, H. (2020). On design orthogonality, maximin distance, and projection uniformity for computer experiments. *Journal of the American Statistical Association.* https://doi.org/10.1080/01621459.2020.1782221.

Weng, J. (2014). Maximin Strong Orthognal Arrays. *Master's thesis* at Simon Fraser University under supervision of Boxin Tang and Jiguo Cao.

Xiao, Q. and Xu, H. (2018). Construction of maximin distance designs via level permutation and expansion. *Statistica Sinica* **28**, 1395-1414.

Xu, H. and Wu, C.F.J. (2001). Generalized minimum aberration for factorial designs. *The Annals of Statistics* **29**, 1066-1077.

Ye, K.-Q. (1998). Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association* **93**, 1430-1439.

Zhou, Y.D. and Tang, B. (2019). Column-orthogonal strong orthogonal arrays of strength two plus and three minus. *Biometrika* **106**, 997-1004. https://doi.org/10.1093/biomet/asz043

Appendix A: Construction by Liu and Liu (2015)

Let $\mathbf{V}$ be an OA(n, m_{oa}, s, t). The algorithm of Liu and Liu (2015) proceeds as follows: Define a block diagonal $m_{\text{oa}} \times 2k$ matrix $\mathbf{R}$ that has

- k diagonal blocks of identical $b \times 2$ matrices, where $b = t$ for even t and $b = t + 1$ for odd t,
- followed by $q = m_{\text{oa}} - bk$ rows of zeroes (none if $q = 0$).

The design $\mathbf{D}$ is obtained as $\mathbf{D} = \mathbf{VR}$; remember that Liu and Liu denoted the levels in $\mathbf{V}$ by $-(s-1), -(s-3), \dots, +(s-1)$. Each column of the $b \times 2$ matrix holds exactly one element of $s^0 = 1, s^1 = s, \dots, s^{t-1}$ (where the list stops at s for $s = 2$), with an additional zero element for odd t; these values carry a positive or negative sign. It is thus straightforward, if a little bit tedious, to define $\mathbf{A}$, $\mathbf{B}$ etc. according to the following rule:

- $\mathbf{a}_\ell$ is obtained from the unique column $\mathbf{v}_j$ of $\mathbf{V}$ for which column $\mathbf{r}_\ell$ holds the entry $\pm s^{t-1}$,
- $\mathbf{b}_\ell$ is obtained from the unique column $\mathbf{v}_j$ of $\mathbf{V}$ for which column $\mathbf{r}_\ell$ holds the entry $\pm s^{t-2}$,
- and so forth.

Where the entry in the matrix $\mathbf{R}$ is positive, $\mathbf{v}_j$ is used directly; where the entry in the matrix $\mathbf{R}$ has a minus sign, $s - 1 - \mathbf{v}_j$ is used (reversal of levels). Equations (5) to (8) gave the results of these allocations for $t = 2$ to $t = 4$. The matrix constructions behind these equations are detailed below:

For a strength 2 OA$(n, m, s, 2)$ called $\mathbf{V}$, the $2\lfloor m/2 \rfloor$ columns of the matrices $\mathbf{A}$ and $\mathbf{B}$ are obtained as follows:

$$\mathbf{a}_\ell = \begin{cases} \mathbf{v}_{\ell+1} & \ell \text{ odd} \\ \mathbf{v}_{\ell-1} & \ell \text{ even} \end{cases}, \qquad \mathbf{b}_\ell = \begin{cases} \mathbf{v}_\ell = \mathbf{a}_{\ell+1} & \ell \text{ odd} \\ s - 1 - \mathbf{v}_\ell = s - 1 - \mathbf{a}_{\ell-1} & \ell \text{ even} \end{cases}, \quad \ell = 1, \dots, 2\lfloor m/2 \rfloor,$$

where $1 \le j \le \lfloor m/2 \rfloor$.

For a strength 3 OA$(n, m, s, 3)$ called $\mathbf{V}$, the $2\lfloor m/4 \rfloor$ columns of the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are obtained as follows:

$$\mathbf{a}_\ell = \begin{cases} \mathbf{v}_{2\ell+1} & \ell \text{ odd} \\ \mathbf{v}_{2\ell-3} & \ell \text{ even} \end{cases}, \quad \mathbf{b}_\ell = \mathbf{v}_{2\ell}, \quad \mathbf{c}_\ell = \begin{cases} \mathbf{v}_{2\ell-1} = \mathbf{a}_{\ell+1} & \ell \text{ odd} \\ s - 1 - \mathbf{v}_{2\ell-1} = s - 1 - \mathbf{a}_{\ell-1} & \ell \text{ even} \end{cases}, \quad \ell = 1, \dots, 2\lfloor m/4 \rfloor.$$

If $m - 4\lfloor m/4 \rfloor = 3$, an additional column can be added as follows:

$$\mathbf{a}_{2\lfloor m/4 \rfloor+1} = \mathbf{v}_m, \quad \mathbf{b}_{2\lfloor m/4 \rfloor+1} = \mathbf{v}_{m-1}, \quad \mathbf{c}_{2\lfloor m/4 \rfloor+1} = \mathbf{v}_{m-2}.$$

For a strength 4 OA$(n, m, s, 4)$ called $\mathbf{V}$, the $2\lfloor m/4 \rfloor$ columns of the matrices $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$ and $\mathbf{A}_4$ are obtained as follows ($\ell = 1, \dots, \lfloor m/4 \rfloor$): For odd ℓ,

$$\mathbf{a}_{1;\ell} = \mathbf{v}_{2\ell+2}, \quad \mathbf{a}_{2;\ell} = \mathbf{v}_{2\ell+1}, \quad \mathbf{a}_{3;\ell} = \mathbf{v}_{2\ell} = \mathbf{b}_{\ell+1}, \quad \mathbf{a}_{4;\ell} = \mathbf{v}_{2\ell-1} = \mathbf{a}_{\ell+1},$$

for even ℓ,

$$\mathbf{a}_{1;\ell} = \mathbf{v}_{2\ell-3}, \quad \mathbf{a}_{2;\ell} = \mathbf{v}_{2\ell-2}, \quad \mathbf{a}_{3;\ell} = s - 1 - \mathbf{v}_{2\ell-1} = s - 1 - \mathbf{b}_{\ell-1}, \quad \mathbf{a}_{4;\ell} = s - 1 - \mathbf{v}_{2\ell} = s - 1 - \mathbf{a}_{\ell-1}.$$

Appendix B: Proof of Proposition 4.1

Let $\mathbf{A}$ and $\mathbf{B}$ be OA$(n, m, s, 2)$, and let $\mathbf{A}^*$ and $\mathbf{B}^*$ denote those matrices after subtracting $(s-1)/2$ (i.e. centered versions of the matrices). Li et al.'s (2021) algorithm proceeds as follows:

a) Obtain an $n \times 2m'$ array $\mathbf{C} = (\mathbf{C}_1, \dots, \mathbf{C}_{m'/2})$ by interleaving the columns of $\mathbf{A}$ and $\mathbf{B}$ as follows:

$$\mathbf{C}_\ell = (\mathbf{a}_{2\ell-1}, \mathbf{b}_{2\ell-1}, \mathbf{a}_{2\ell}, \mathbf{b}_{2\ell}), \quad \ell = 1, \dots, m'/2.$$

b) Obtain the column-centered matrix $\mathbf{C}^*$ by subtracting $(s-1)/2$ from each element of $\mathbf{C}$, so that elements are in the interval $[-(s-1)/2, (s-1)/2]$, i.e. $\mathbf{C}^*$ interleaves $\mathbf{A}^*$ and $\mathbf{B}^*$.

c) Obtain $n \times 2$ matrices $\mathbf{D}^*_\ell = C^*_\ell \mathbf{V}$, with

$$\mathbf{V} = \begin{pmatrix} s^2 & s & 0 & 1 \\ -1 & 0 & s^2 & s \end{pmatrix}^\top.$$

d) Obtain the $n \times m'$ design matrix

$$\mathbf{D} = (\mathbf{D}^*_1, \dots, \mathbf{D}^*_{m'/2}) + (s^3 - 1)/2.$$

The first column of $\mathbf{D}_\ell^*$ is the $2\ell - 1^{th}$ column of $\mathbf{D}^*$,

$$\mathbf{d}_{2\ell-1}^* = s^2\mathbf{a}_{2\ell-1}^* + s\mathbf{b}_{2\ell-1}^* + \mathbf{a}_{2\ell}^*,$$

the second column is

$$\mathbf{d}_{2\ell}^* = s^2\mathbf{a}_{2\ell}^* + s\mathbf{b}_{2\ell}^* - \mathbf{a}_{2\ell-1}^*.$$

Clearly, $\mathbf{D}^* = s^2\mathbf{A}^* + s\mathbf{B}^* + \mathbf{C}^*$ with the columns of $\mathbf{C}^*$ obtained from $\mathbf{A}^*$. Now, observe that the superscript * stands for subtraction of a constant only; the only position in which this matters is the subtraction of $\mathbf{a}_{2\ell-1}^*$, for which the "$-$" after subtraction of the center value corresponds to a reversal of the levels, which can also be written as $s - 1 - \mathbf{a}_{2\ell-1}$ for the original coding $0, \ldots, s-1$.

Appendix C: Start values and example constructions for Shi and Tang Family 1

$n = 2^k$ for even $k \geq 4$

The start values are $Y_\mathbf{A} = (1, 2, 4, 8, 15)$ and $Y_\mathbf{B} = (12, 9, 3, 6, 5)$. This implies $Y_{\mathbf{A}+_2\mathbf{B}} = (13, 11, 7, 14, 10)$. There are many possibilities for $Y_\mathbf{C}$, e.g. 2, 1, 1, 1, 1, since one only has to avoid choosing $\mathbf{c}_\ell$ coincident with $\mathbf{a}_\ell$, $\mathbf{b}_\ell$ or $\mathbf{a}_\ell +_2 \mathbf{b}_\ell$.

$n = 2^k$ for odd $k \geq 7$

The start values are
$Y_\mathbf{A}$=(1, 2, 4, 8, 15, 17, 18, 20, 24, 31, 33, 34, 36, 40, 47, 49, 50, 52, 56, 63, 65, 66, 68, 72, 79, 81, 82, 84, 88, 95, 97, 98, 100, 104, 111, 113, 114, 116, 120, 127),
$Y_\mathbf{B}$=(42, 37, 25, 3, 117, 74, 41, 10, 14, 102, 92, 69, 23, 6, 83, 90, 73, 71, 21, 86, 54, 28, 7, 5, 57, 61, 44, 26, 19, 53, 60, 12, 9, 13, 58, 55, 62, 35, 27, 38).
Note that the 28th entry for $\mathbf{B}$ was corrected from 22 to 26 versus Table 1 of Shi and Tang (or from $\mathbf{e}_2\mathbf{e}_3\mathbf{e}_5$ to $\mathbf{e}_2\mathbf{e}_4\mathbf{e}_5$ in their notation).
$Y_\mathbf{A}$ and $Y_\mathbf{B}$ imply $Y_{\mathbf{A}+_2\mathbf{B}}$=(43, 39, 29, 11, 122, 91, 59, 30, 22, 121, 125, 103, 51, 46, 124, 107, 123, 115, 45, 105, 119, 94, 67, 77, 118, 108, 126, 78, 75, 106, 93, 110, 109, 101, 85, 70, 76, 87, 99, 89).
Columns of $\mathbf{C}$ can be most conveniently chosen from the multiples of 16 that do not occur in any of the matrices.

Special case $k = 5$ $(n = 32)$

The maximum number of columns in an SOA$(2^5, m, 8, 3)$ with property α is $m = 9 < 10 = 5 \cdot 2^{5-4}$. This maximal SOA is e.g. obtained with $\mathbf{A}$ chosen as the GMA design 9-4.1 (Yates columns 1, 2, 4, 8, 16, 7, 11, 19, 29) and $\mathbf{B}$ consisting of Yates columns 24, 20, 9, 6, 5, 27, 17, 12, 3. Then $\mathbf{A} +_2 \mathbf{B}$ consists of Yates columns 25, 22, 13, 14, 21, 28, 26, 31, 30. For $\mathbf{C}$, one can use e.g. Yates column 10 or 15 for all columns.

Appendix D: Example constructions for Shi and Tang Families 2 and 3

Section 4.2 provided the recursive construction for Families 2 and 3. It will be applied to two examples in this appendix.

Example: Constructing an SOA$(64, 16, 8, 3)$ or an OSOA$(64, 15, 8, 3+)$

$64 = 2^6$, i.e. $k = 6$ is even. We need a matrix $\mathbf{Y}$ with $2^{k-2} = 16$ rows in order to obtain matrices $\mathbf{A}$ and $\mathbf{B}$ with 2^k rows. We already saw in Section 4.2 that $Y_\mathbf{Y} = c(2, 3, 1, 8, 10, 11, 9, 12, 14, 15, 13, 4, 6, 7, 5)$, which arises from applying Proposition 4.3 to the start vector $Y_{\mathbf{Y}_2} = 231$ (with $k = 2$ in the proposition). Corollary 4.1 tells us that $\mathbf{A}$ holds Yates matrix columns 33 to 47 (in that order) and $\mathbf{B}$ holds Yates matrix columns $16 + Y_\mathbf{Y}$, and Lemma 4.5 tells us to use Yates matrix columns 1 to 15 for obtaining the OSOA with $n/4 - 1 = 15$ columns. For obtaining the SOA with 16 columns, one can add Yates column 32 to $\mathbf{A}$, Yates column 16 to $\mathbf{B}$, and an arbitrary column from Yates columns 1 to 15 to matrix $\mathbf{C}$, so that matrix $\mathbf{C}$ has one duplicate column pair. According to Lemma 4.5, this implies orthogonality for most column pairs, with the exception of a non-zero correlation for the pair that have the same **C** matrix column.

Example: Constructing Family 2 and Family 3 designs in 32 and 128 runs (k odd)

The start values for a design in 2^5 runs ($k = 5$) have $2^{5-2} - 1$ elements and were given in Lemma 4.3 as $Y_{\mathbf{X}} = 1234567$, $Y_{\mathbf{Y}} = 7521643$, and $Y_{\mathbf{Z}} = 6715324$.
The resulting start columns for matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{A} +_2 \mathbf{B}$ are given as
$Y_{\mathbf{A}} = (16, 17, 18, 19, 20, 21, 22, 23)$,
$Y_{\mathbf{B}} = (8, 15, 13, 10, 9, 14, 12, 11)$,
$Y_{\mathbf{A}+_2\mathbf{B}} = (24, 30, 31, 25, 29, 27, 26, 28)$,
for an SOA$(32, 8, 8, 3)$ with properties α and β.

According to Lemma 4.5, eight corresponding columns for **C** should be obtained from Yates columns 1 to 7, with one duplicate, and the resulting array has a non-zero correlation for the pair of columns that share the same **C** column. If one only needs seven columns, omitting the first columns from **A** and **B** and using Yates columns 1 to 7 for **C** yields an OSOA$(32, 7, 8, 3+)$, since the array is from Shi and Tang's Family 3 and additionally fulfills all requirements of Lemma 3.5.

One step of the recursion yields

$Y_{\mathbf{X}} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, ..., 31)$,
$Y_{\mathbf{Y}} = (7, 5, 2, 1, 6, 4, 3, 16, 23, 21, 18, 17, 22, 20, 19, 24, 31, 29, 26, 25, 30, 28, 27, 8, 15, 13, 10, 9, 14, 12, 11)$,
$Y_{\mathbf{Z}} = (6, 7, 1, 5, 3, 2, 4, 24, 30, 31, 25, 29, 27, 26, 28, 8, 14, 15, 9, 13, 11, 10, 12, 16, 22, 23, 17, 21, 19, 18, 20)$
for the construction of an OSOA$(128, 31, 8, 3+)$, whose Yates matrix columns are
$Y_{\mathbf{A}} = (65, \ldots, 95)$,
$Y_{\mathbf{B}} = (39, 37, 34, 33, 38, 36, 35, 48, 55, 53, 50, 49, 54, 52, 51, 56, 63, 61, 58, 57, 62, 60, 59, 40, 47, 45, 42, 41, 46, 44, 43)$,
$Y_{\mathbf{A}+_2\mathbf{B}} = (102, 103, 97, 101, 99, 98, 100, 120, 126, 127, 121, 125, 123, 122, 124, 104, 110, 111, 105, 109, 107, 106, 108, 112, 118, 119, 113, 117, 115, 114, 116)$.

Orthogonal columns are guaranteed by choosing Yates columns 1 to 31 for matrix **C**. The analogous construction of the Family 2 SOA$(128, 32, 8, 3)$ with properties α and β additionally uses Yates columns 64 and 32 as the first columns of matrices **A** and **B**, and adds another column from 1 to 31 to matrix **C**.