This item is the archived peer-reviewed author-version of:

Blocking orthogonal designs with mixed integer linear programming

Reference:

Sartono Bagus, Goos Peter, Schoen Eric D.- Blocking orthogonal designs with mixed integer linear programming
Technometrics : a journal of statistics for the physical, chemical, and engineering sciences - ISSN 1537-2723 - 57:3(2015), p. 428-439

Full text (Publishers DOI): <http://dx.doi.org/doi:10.1080/00401706.2014.938832>

To cite this reference: <http://hdl.handle.net/10067/1276660151162165141>

Blocking Orthogonal Designs with Mixed Integer Linear Programming

Bagus Sartono

Department of Statistics, Bogor Agricultural University, Indonesia

Peter Goos

Faculty of Applied Economics and StatUa Center for Statistics, University of Antwerp, Belgium

Faculty of Bioscience Engineering, University of Leuven, Belgium

Eric Schoen

TNO, Zeist, Netherlands

Faculty of Applied Economics, University of Antwerp, Belgium

June 16, 2014

Abstract

We present a mixed integer linear programming approach to orthogonally block two-level, multi-level and mixed-level orthogonal designs. The approach involves an exact optimization technique which guarantees an optimal solution. It can be applied to many problems where combinatorial methods for blocking orthogonal designs cannot be used. By means of 54-run and 64-run examples, we demonstrate that our approach outperforms two benchmark techniques in terms of the number of estimable two-factor interaction contrasts and in terms of the D-efficiency for models with main effects and some two-factor interaction contrasts. We demonstrate the generic nature of our approach by applying it to the most challenging instances in a catalog of all orthogonal designs of strength 3 with up to 81 runs as well as a small catalog of strength-4 designs. The approach can also be applied to strength-2 designs, but, for these cases, alternative methods described in the literature may perform equally well. Supplementary materials for this article are available online.

Keywords: Aliasing, Confounding, Mixed Integer Linear Programming, Mixed-Level Orthogonal Array, Multi-Level Orthogonal Array, Orthogonal Blocking, Pure-Level Orthogonal Array, Two-Level Orthogonal Array

1 Introduction

Recent research in design of experiments has made many nonregular orthogonal designs available to define the treatments of a factorial experiment. References include Deng and Tang (2002), Ingram and Tang (2005), Sun et al. (2008), Brouwer et al. (2006), Bulutoglu and Margot (2008) and Schoen et al. (2010). Motivated by a calcium fortification study at FrieslandCampina, a dairy company in the Netherlands, we developed a novel approach to arrange these designs in blocks.

The focus in this paper is on designs based on orthogonal arrays (OAs) of strength 3 or 4. If an OA of either strength is used as an experimental design, the main-effect estimators are independent of each other and the estimators of the two-factor interactions are independent of those of the main effects. For strength-3 arrays, the two-factor interaction effects are aliased with each other, while they are independent in strength-4 arrays (Hedayat et al., 1999).

It is often possible to estimate several interaction effects from a design based on a strength-3 array, in spite of the dependence of their estimators. Therefore, an OA of strength 3 is a good design option in the presence of a limited number of two-factor interaction effects, while a strength-4 OA is recommended in case a substantial number of interactions is expected to be active.

Strength-3 arrays often require a substantial number of runs, which makes it harder to conduct all the runs under homogeneous circumstances. Designs of strength 4 require even more runs than designs of strength 3. This necessitates good methods for blocking strength-3 and strength-4 arrays.

Ideally, the blocking is such that the treatment comparisons are not affected by the differences between the blocks. The goal of this paper is to propose a practical strategy for finding blocking arrangements for OAs of strength 3 or 4, so that all the main effects are orthogonal to the block effects and that the confounding of two-factor interactions with the blocks is minimized. To this end, we applied mixed integer linear programming (MILP) techniques to an extensive catalog of strength-3 designs and a smaller one of strength-4 designs. We compare the results from the MILP approach with the results of two benchmark strategies. As our first benchmark strategy, we employ the algorithm of Cook and Nachtsheim (1989) to obtain D-optimal blocking arrangements based on a main-effects-only model. We call this approach the CN benchmark approach. The second benchmark approach is an extension of the modular arithmetic commonly used to block regular designs (see, e.g., Wu and Hamada, 2009) to handle nonregular OAs of strength 3 or higher. We call this approach the MA benchmark approach.

The rest of this paper is organized as follows. In Section 2, we discuss the calcium fortification study (requiring a strength-3 64-run design in eight blocks) and an artificial example (requiring a strength-3 54-run design in nine blocks), and the existing techniques that can be used for

blocking OAs of strength 3 or strength 4. Next, in Section 3, we describe our MILP approach in detail using the OAs from Section 2 as illustrations. In Section 4, we demonstrate the usefulness of the MILP approach by applying it to a rich collection of strength-3 OAs with up to 81 runs. In Section 5, we investigate the potential of our approach to block designs of strength 4. Finally, in Section 6, we summarize the advantages and limitations of our approach.

Throughout the paper, we use the following terms and notation. An OA with m factors, N runs and strength t is an $N \times m$ matrix of symbols. Each column corresponds to a categorical treatment factor and involves a certain number of symbols. These symbols are arranged so that, for every set of t columns, each t -tuple of symbols occurs equally often (Rao, 1947). Every symbol in a given column represents a factor level. In this paper, we denote the levels of an s_i -level factor by $0, 1, \dots, s_i - 1$. We assume that the set of factors can be partitioned in γ subsets, where all m_i factors in a given subset i have the same number of levels, s_i . We denote by $\text{OA}(N; s_1^{m_1} \times s_2^{m_2} \times \dots \times s_\gamma^{m_\gamma}, t)$ the set of non-isomorphic strength- t OAs involving N runs, m_1 factors acting at s_1 levels, m_2 factors acting at s_2 levels, etc. In this paper, we mainly study orthogonal arrays of strength 3, and for these OAs, we drop the explicit reference to the strength t . We drop the exponent m_i when it equals one, i.e. when there is exactly one factor with s_i levels. OAs with $\gamma = 1$ are pure-level arrays, while those with $\gamma > 1$ are mixed-level arrays. Finally, we denote by $\mathbf{0}_{r \times c}$ a zero matrix with r rows and c columns and by $\mathbf{1}_r$ an r -dimensional vector of ones.

2 Examples and benchmarks

In this section, we study two scenarios where a blocked experiment is required. The first scenario, based on a real-life experiment, involves a 64-run multi-level OA. We use that scenario to introduce our first benchmark approach. The second scenario involves an artificial example using a 54-run three-level OA. We use that scenario to introduce our second benchmark approach.

2.1 A calcium fortification experiment involving a mixed-level array

In the summer of 2007, researchers of FrieslandCampina, a dairy company in the Netherlands, conducted an experiment to compare different ways of calcium fortification in liquid food products. They were particularly interested in the presence of calcium after the processing of the food by the stomach. The study involved a laboratory simulation of the intestinal tract in a reaction vessel. The four experimental factors and their numbers of levels are given in Table 1.

The experimenters used eight different calcium sources in the study, as well as four different food products in which to dissolve the calcium. For each of the calcium sources, a high and a low concentration was used. The calcium solution was put into a reaction vessel, together with fluids mimicking those in the human stomach. Two different stomach conditions were studied: a

Table 1: Experimental factors in the calcium fortification study.

Factor	Name	Number of levels
A	Calcium source	8
B	Food product	4
C	Calcium concentration	2
D	Stomach condition	2

Table 2: Levels of the second two-level factor in the four strength-3 arrays of the type $OA(64; 8 \times 4 \times 2^2)$ considered for the calcium fortification experiment.

Array	Levels of the fourth factor
I	01011010010110100101101001011010100101101001011010010110100101
II	0101101001011010010110100110011010011001101001011010010110100101
III	0101101001011010011001100110011010011001100110011010010110100101
IV	0101101001011010011001100110100110010110100110011010010110100101

fasted condition and a fed condition. After a fixed time period, new fluids were added to simulate conditions of the small but important part of the intestinal tract right after the stomach, called the duodenum. After another fixed time period, a first sample was taken to measure the calcium concentration, and new fluids were added to make the conditions similar to those in the next part of the intestinal tract, the ileum. At the end of each experimental run, a second sample was taken to measure the calcium concentration again.

The primary goal of the calcium fortification experiment was to estimate the main effects of the factors, but the experimenters expressed genuine interest in quantifying two-factor interactions as well. For three reasons, it was natural to use one of the four 64-run strength-3 designs of the type $OA(64; 8 \times 4 \times 2^2)$, enumerated by Schoen et al. (2010), as the treatment design for that experiment. First, the run size of 64 is divisible by 8, which was the number of vessels that was available at any time. The OAs therefore might be arranged into eight blocks of size eight. The blocks correspond to eight time points and the block size matches the available number of vessels. Second, as the arrays have strength 3, they allow an independent estimation of the main effects, which were of primary interest. Finally, the designs have a large number of runs, and therefore offered the prospect of being able to estimate many two-factor interaction contrasts.

Table 2 shows the levels of the second two-level factor of each of the four non-isomorphic 64-run strength-3 OAs of the type $OA(64; 8 \times 4 \times 2^2)$. In the representation in the table, we assume that the factor level combinations are sequentially sorted according to the levels of the eight-level factor, the levels of the four-level factor, and the levels of the first two-level factor. All 64 combinations of levels of these three factors appear in the design exactly once. The four arrays only differ with respect to the second two-level factor.

Even when used in a completely randomized design, none of the four arrays of the type

$OA(64; 8 \times 4 \times 2^2)$ permits estimation of all 42 two-factor interaction contrasts. Array I allows 39 two-factor interaction contrasts to be estimated, while the other three arrays offer 41 estimable two-factor interaction contrasts. As a result, before considering the OAs' suitability for use in a blocked experiment, there was an a priori preference for the arrays II, III and IV.

To study the four designs' suitability for use in a blocked experiment, we arranged their runs in eight blocks of size eight using the D-optimal blocking algorithm of Cook and Nachtsheim (1989), as implemented in the OPTTEX procedure of SAS[®]. For strength-3 designs, this approach, to which we refer as the CN benchmark approach, involves a main-effects model for the treatment factors. Whenever the CN benchmark approach returns a blocking arrangement with a D-efficiency of 100%, the resulting design is an orthogonally blocked design for the main effects, implying that the main effects can be estimated independently.

When using 50 random starts of the blocking algorithm, the OPTTEX procedure was able to produce blocking arrangements with a D-efficiency of 100% for all four arrays.

Of course, blocking based on a main-effects model, as in the CN benchmark approach, is not satisfactory, because it ignores the confounding of two-factor interactions with blocks. As the researchers in the calcium fortification study were interested in quantifying two-factor interactions, the confounding of two-factor interactions with the blocks should be minimized. In certain scenarios which, unlike the calcium fortification study, involve pure-level OAs, it is possible to address the confounding of two-factor interactions with blocks, by generalizing the blocking approach for regular OAs using modular arithmetic so that it can also handle nonregular OAs.

2.2 An artificial experiment involving a pure-level OA

As a second benchmark, we study a generalized version of the blocking methodology for regular designs involving modular arithmetic (Wu and Hamada, 2009). Note that this modular arithmetic (MA) benchmark approach is less general than the CN benchmark approach, because it is applicable only for pure-level designs in cases where the number of blocks is a power of the number of levels of each factor. We illustrate the usefulness of the second benchmark approach using a set of four arrays of the type $OA(54; 3^5)$, the runs of which have to be arranged in nine blocks of six. The four OAs were introduced in the literature by Hedayat et al. (1997) and differ in the number of estimable two-factor interaction contrasts. The numbers of estimable two-factor interaction contrasts equal 31, 35, 36, and 39, and, therefore, we refer to the OAs as the 31-2FI array, the 35-2FI array, the 36-2FI array and the 39-2FI array. It is clear that none of the OAs permits estimation of all 40 two-factor interaction contrasts and that, if the design of the experiment were not blocked, the 39-2FI array OA would be preferred.

To arrange the four arrays of the type $OA(54; 3^5)$ in nine blocks of size six, we need to produce a nine-level blocking factor. We do so by combining two three-level factors obtained by decomposing the four degrees of freedom for each two-factor interaction and the eight degrees of

freedom for each three-factor interaction into orthogonal components of two degrees of freedom each. We refer to these three-level factors as blocking basic factors (BBFs). For example, we calculated $F = X_1 + 2X_2 + X_3 \pmod{3}$ as one of the BBFs, where X_i represents the level of the i th factor in the design.

After selecting two BBFs, the blocking factor is defined by the nine level combinations of the BBFs. Subsequently, we need to check whether that blocking factor is indeed orthogonal to the main effects. An exhaustive search revealed that orthogonal nine-block arrangements cannot be obtained for the 31-2FI and 36-2FI arrays. The best arrangement for the 35-2FI array sacrifices the estimability of six two-factor interaction contrasts, leaving 29 estimable ones. The best arrangement for the 39-2FI array sacrifices eight estimable two-factor interaction contrasts, leaving 31 estimable ones. Therefore, in case nine blocks of size six are desired, our recommendation based on the MA benchmark approach would be to use the 39-2FI array.

3 Using mixed integer linear programming to find blocking arrangements

To obtain orthogonal blocking arrangements for given strength-3 or strength-4 OAs, we propose a mixed integer linear programming approach. Linear programming (LP) is a common method to determine the values of a set of decision variables so as to maximize or minimize a particular linear objective function, while satisfying a set of linear constraints. When all the variables in the solution are required to be integer, the method is named integer linear programming (ILP). It is called mixed integer linear programming (MILP) if only some of the variables are required to be integer. In design of experiments, ILP was used earlier by Bulutoglu and Margot (2008) to classify OAs and by Capehart et al. (2011) to construct regular two-level split-plot designs. In this section, we show how MILP can be used to find good blocking arrangements for regular and nonregular strength-3 OAs.

3.1 Description of the methodology

Suppose that we wish to find an optimal orthogonal blocking arrangement of a strength-3 OA $\mathbf{Z}$ involving N runs and m factors, where the i th factor has s_i levels, in b blocks of size N/b . Any arrangement in b blocks can be represented by means of an $N \times b$ indicator matrix $\mathbf{B} = [b_{ij}]$, where b_{ij} equals 1 if the i th run is assigned to the j th block and 0 otherwise. The elements of $\mathbf{B} = [b_{ij}]$ are binary decision variables in our linear programming approach.

We require the blocking factor to be orthogonal to the main effects. Therefore, each level of any treatment factor must appear equally often in each block. This requirement can be cast into a linear constraint. To this end, replace the i th column of $\mathbf{Z}$ by $(s_i - 1)$ orthogonal contrast columns,

normalize these columns such that they have the same length, and denote the normalized columns by $\mathbf{z}_{i(1)}, \mathbf{z}_{i(2)}, \dots, \mathbf{z}_{i(s_i-1)}$. The i th main effect is orthogonal to the blocking factor if and only if $\mathbf{z}_{i(q)}^T \mathbf{B} = \mathbf{0}_{1 \times b}$ for all $q \in \{1, 2, \dots, s_i - 1\}$. If we collect all $p_1 = \sum_{i=1}^m (s_i - 1)$ main-effect contrast columns in a single $N \times p_1$ matrix $\mathbf{X}$, then the orthogonal blocking requirement can be written as $\mathbf{X}^T \mathbf{B} = \mathbf{0}_{p_1 \times b}$.

There may be several ways to achieve orthogonality between the main effects and the blocks. In that case, we prefer a blocking arrangement that enables us to estimate as many two-factor interaction effects as possible. One way in which such a blocking arrangement can be found is by minimizing the confounding between the two-factor interactions and the blocking factor. To quantify this type of confounding, we construct the $N \times p_2$ matrix $\mathbf{W}$ containing the $p_2 = \sum_{i=1}^{m-1} \sum_{j=i+1}^m (s_i - 1)(s_j - 1)$ two-factor interaction contrast vectors. This is done by element-wise multiplication of all pairs of main-effect contrast vectors corresponding to different treatment factors. The elements of the matrix product $\mathbf{D} = \mathbf{W}^T \mathbf{B}$ then measure the extent to which the two-factor interaction contrasts are confounded with the blocks. The LP model to search for an optimal blocking arrangement can then be formulated as “Find $\mathbf{B}$ such that $\mathbf{X}^T \mathbf{B} = \mathbf{0}_{p_1 \times b}$ and the confounding of two-factor interactions with blocks, as measured by $\mathbf{D} = \mathbf{W}^T \mathbf{B}$, is minimal.”

Ideally, $\mathbf{D} = \mathbf{W}^T \mathbf{B} = \mathbf{0}_{p_2 \times b}$, which would mean that the blocks are also orthogonal to all two-factor interaction effects. In many practical situations, this will be impossible to achieve due to the large number of two-factor interaction contrasts of interest. Therefore, in our MILP approach, we sequentially minimize the elements of $\mathbf{D}$ that have the largest absolute value and the sum of the absolute values of all elements of $\mathbf{D}$. If we denote by d_{ij} the elements of $\mathbf{D}$, by $\mathbf{w}_i$ the i th column of $\mathbf{W}$ and by $\mathbf{b}_j$ the j th column of $\mathbf{B}$, so that $d_{ij} = \mathbf{w}_i^T \mathbf{b}_j$, then we first minimize

$$d = \max_{i,j} \{|d_{ij}|\},$$

followed by

$$S = \sum_{i=1}^{p_2} \sum_{j=1}^b |d_{ij}|.$$

By prioritizing the minimization of d , we ensure that none of the two-factor interaction contrasts is severely confounded with any of the blocks. Given that d is minimal, we can then focus on minimizing the remaining, less severe, confounding.

The sequential minimization can be cast into a single objective function, using the big- M method (Winston, 2003). The objective function of our MILP approach then becomes

$$f(\mathbf{B}) = Md + S,$$

where $M \gg 1$.

In brief, our MILP approach involves searching a binary matrix $\mathbf{B}$, which minimizes $f(\mathbf{B})$,

subject to $\mathbf{X}^T \mathbf{B} = \mathbf{0}_{p_1 \times b}$. Its exact implementation, however, involves several additional linear constraints to determine the absolute values $|d_{ij}|$ as well as their maximum d , and to specify the number of runs in each of the b blocks. The detailed specification of our MILP model is as follows:

$$\text{Minimize } f(\mathbf{B}) = Md + \sum_{i=1}^{p_2} \sum_{j=1}^b d_{ij}^+ + \sum_{i=1}^{p_2} \sum_{j=1}^b d_{ij}^- \quad \text{where } M \gg 1 \quad (2)$$

Subject to

- (1) $\mathbf{w}_i^T \mathbf{b}_j - d_{ij}^+ + d_{ij}^- = 0, \quad i = 1, \dots, p_2, j = 1, \dots, b$
- (2) $0 \leq d_{ij}^+ \leq d, \quad i = 1, \dots, p_2, j = 1, \dots, b$
- (3) $0 \leq d_{ij}^- \leq d, \quad i = 1, \dots, p_2, j = 1, \dots, b$
- (4) $\mathbf{X}^T \mathbf{B} = \mathbf{0}_{p_1 \times b},$
- (5) $\mathbf{B}^T \mathbf{1}_N = (N/b) \mathbf{1}_b,$
- (6) $\mathbf{B} \mathbf{1}_b = \mathbf{1}_N,$
- (7) $b_{ij} \in \{0, 1\}.$

In the constraints as well as in the objective function, the absolute values $|d_{ij}|$ are all replaced with two non-negative linear variables d_{ij}^+ and d_{ij}^- . The variable d_{ij}^+ equals d_{ij} if d_{ij} is positive and zero otherwise. The variable d_{ij}^- equals $-d_{ij}$ if d_{ij} is negative and zero otherwise. This is enforced by the constraints (1), (2) and (3), and by the objective function which minimizes the sum of all d_{ij}^+ and d_{ij}^- values. Constraint (1) links the variables d_{ij}^+ and d_{ij}^- to the elements of $\mathbf{D}$, while the constraints (2) and (3) ensure that all d_{ij}^+ and d_{ij}^- values are positive. The objective function makes sure that d_{ij}^+ and d_{ij}^- are never both strictly positive. This is the standard way of linearizing the nonlinear absolute value operator in linear programming. The constraints (2) and (3) define d as an upper bound of all d_{ij}^+ and d_{ij}^- values. The objective function ensures d is the smallest possible upper bound, i.e. the maximum of all d_{ij}^+ and d_{ij}^- values.

Constraint (4) specifies that the blocking factor must be orthogonal to the main effects. The constraints specify the structure of the binary matrix $\mathbf{B}$: every column should contain exactly N/b ones (meaning that every block should have N/b runs), and every row should contain exactly one entry that is equal to 1 (meaning that every run has to be assigned to one block). Finally, constraint (7) imposes that each element b_{ij} of $\mathbf{B}$ is a binary decision variable. The model is a MILP model, because it contains integer decision variables b_{ij} as well as real decision variables (d_{ij}^+ , d_{ij}^- , and d).

When $b = N/s_1$, we add additional constraints to the model to speed up the search for an optimal solution. Consider, for example, the problem of blocking an array of the type OA(64, $8 \times 4 \times 2^2$) in eight blocks of size eight. If such an array is sorted in increasing order of the level of the eight-level factor, the resulting array has equal settings for this factor in the first eight runs. In any orthogonal blocking arrangement, these eight runs have to be assigned to different blocks. Therefore, we sort the runs of the array and add the constraints $b_{ii} = 1$ to the MILP

Table 3: Optimal blocking arrangements for the arrays II-IV of the type $OA(64, 8 \times 4 \times 2^2)$ in Table 2.

Array	Eight-Level Blocking Factor
II	1234785657412368256841373475861213867524872165436452137886374251
III	1234567823761485478365218154236765874132381672544725386156127843
IV	1234567827538461741652834875312656813742638247151527684338642175

formulation, for $i = 1, 2, \dots, 8$. This reduces the number of binary variables in $\mathbf{B}$ by one eighth, and leads to a faster solution of the MILP. Similar constraints speed up the blocking of arrays of the type $OA(64, 4^a \times 2^b)$ in 16 blocks, and of arrays of the type $OA(54, 6 \times 3^a)$ in nine blocks.

We implemented our MILP approach using the OPTMODEL procedure embedded in the OR module of the SAS software and using Matlab. The SAS and Matlab programs are contained within the supplementary materials.

3.2 Examples

3.2.1 Calcium fortification study

We first use the MILP approach to arrange the designs of the type $OA(64; 8 \times 4 \times 2^2)$ given in Table 2 in blocks of size eight. In Section 2.1, we explained that there was an a priori preference for the arrays II, III and IV because they permit the estimation of 41 two-factor interaction contrasts. The MILP approach succeeded in finding orthogonal arrangements in eight blocks for each of these three arrays, so that no estimable two-factor interaction contrasts are lost due to the blocking. Table 3 shows the optimal blocking arrangements produced by the MILP approach. The table was created assuming the runs appear in the same order as in Table 2.

Application of the CN benchmark approach to the three arrays also resulted in orthogonal blocking arrangements. All arrangements permit estimation of 41 two-factor interaction contrasts, just like the blocking arrangements produced by the MILP approach. The MILP approach and the benchmark approach thus perform equally well in terms of the number of estimable interaction contrasts. Nevertheless, differences exist between the blocking arrangements produced by the two methods. To demonstrate this, we computed average D-efficiencies for 10,000 models including all main effects and randomly chosen sets of eight two-factor interaction contrasts for the blocking arrangements produced by the MILP approach and the CN benchmark approach. For each blocked design and randomly chosen model, the D-efficiency was computed as $D = 100 \times |\mathbf{U}^T(\mathbf{I}_N - \mathbf{C}\mathbf{C}^T)\mathbf{U}|^{1/p}/N$, where $\mathbf{U}$ is an $N \times p$ matrix containing all main-effect contrasts and eight two-factor interaction contrasts normalized to length $\sqrt{N}$, $\mathbf{C}$ is an $N \times b$ matrix of normalized and orthogonal blocking contrasts, and $\mathbf{I}_N$ is the $N \times N$ identity matrix. For each OA, we also calculated the average D-efficiencies in the absence of blocking, as $D = 100 \times |\mathbf{U}^T\mathbf{U}|^{1/p}/N$. We present the results in Table 4.

Table 4: Average D-efficiencies (in %) for 10,000 models including eight two-factor interaction contrasts for the arrays II-IV in Table 2, in the absence of blocking and in case the MILP and CN approaches are used for blocking.

Array	Unblocked	MILP	CN
II	99.5	93.1	92.9
III	99.5	93.6	93.1
IV	99.5	93.0	92.8

The table shows that, for each array, the best blocking arrangement in terms of average D-efficiency was found by the MILP approach. The MILP blocking arrangement of array III outperforms all other blocking arrangements. In Section 4, we show that efficiency improvements of the kind observed here can be found when blocking other OAs with the MILP approach as well.

3.2.2 Artificial example

In the supplementary Section A, we study the blocking arrangements of the four arrays of the type $OA(54; 3^5)$ involving nine blocks of six runs, obtained using the MILP approach, the CN benchmark approach, and the MA benchmark procedure. The study includes a calculation of D-efficiencies for the blocked designs for 10,000 models including all main effects and randomly chosen sets of eight two-factor interaction contrasts. The efficiencies for the CN benchmark approach are 0.5%-1.2% smaller than for the MILP approach, while the efficiencies of the MA benchmark approach were 6.0%-7.8% smaller than those obtained using the CN benchmark approach.

We conclude that the MILP approach results in blocking arrangements of arrays of the type $OA(54; 3^5)$ that are generally superior to those produced by the benchmark procedures. Another attractive feature of the MILP approach is its capability to determine whether or not orthogonal blocking is possible for a given OA and blocking structure. The inferior efficiencies, the limited applicability, and the elaborate calculation of the BBFs make the MA benchmark approach unattractive. For this reason, we disregard this approach in the rest of the paper.

4 Application to a catalogue of strength-3 designs

The MILP algorithm is successful in blocking all known series of strength-3 OAs with run sizes ranging from 24 to 81. To substantiate this assertion, we report in this section the results of applying the MILP approach to obtain 136 blocking arrangements of pure-level and mixed-level OAs of 54 different types, and we compare them with those of the CN benchmark approach. We first discuss the series of OAs we have used to test the MILP approach and CN benchmark approach and then discuss the results.

4.1 The catalogue

In total, we studied the blocking of 54 OA types. Each type is defined by its run size, the number of factors and the number of levels of each of the factors. These parameters are given in the first two columns of Table 5. In the column labeled ‘#NI’, the table lists the number of non-isomorphic arrays of each type. The distribution of the number of estimable two-factor interaction contrasts in the absence of blocking is given in the column labeled ‘# Estimable 2FI’. The frequency of each number of estimable contrasts is given by means of a superscript. For example, there are 11 non-isomorphic arrays of the type $OA(32, 4 \times 2^6)$. One of these offers 21 estimable two-factor interaction contrasts, while eight other arrays offer only 20 such contrasts, and the remaining two arrays permit estimation of 19 two-factor interaction contrasts. This is denoted by $21^1, 20^8, 19^2$.

Most of the series studied here contain several arrays with the same number of estimable two-factor interaction contrasts. These arrays may differ in the efficiency with which they estimate the contrast coefficients in the presence as well as in the absence of blocking. We explored this issue by blocking more than one array for some of the OA types. The column ‘nd’ in Table 5 shows the number of arrays we studied of each type.

All the non-isomorphic designs referred to in Table 5 have been generated by the algorithm of Schoen et al. (2010). The numbers of non-isomorphic arrays for the 72-run and 80-run cases are new to the literature.

Table 5: Series of OAs used to compare the MILP approach and the CN benchmark approach. N : run size; #NI: number of isomorphism classes; nd: number of designs used for testing; b : number of blocks; r and r_b : numbers of estimable two-factor interaction contrasts in absence or presence of blocking; *: number of estimable two-factor interaction contrasts after blocking is maximal; s : SOS treatment designs; t : cases where OAs with an r value smaller than the maximum were used; o : MILP problem solved to optimality.

N	Factors	#NI	# Estimable 2FI	nd	b	r	r_b
24	3×2^4	3	$14^2, 11^1$	2	4	14	14^*
	2^{12}	1	11^1	1	12	11^s	0^{*o}
	2^{11}	1	11^1	1	12	11	0^{*o}
27	3^4	1	18^1	1	9	18^s	10^{*o}
32	$4^2 \times 2^4$	2	21^2	2	8	21^s	14^{*o}
	$4^2 \times 2^3$	2	21^2	1	8	21	15^{*o}
	4×2^7	8	21^8	8	8	21^s	14^{*o}
	4×2^6	11	$21^1, 20^8, 19^2$	1	8	20^t	15^{*o}
	2^{16}	5	15^5	5	16	15	0^{*o}
	2^{15}	5	15^5	5	16	15	0^{*o}
36	$3^2 \times 2^2$	3	13^3	3	6	13	13^*
40	5×2^6	1	29^1	1	4	29^s	26^*

continued on next page

Table 5 (continued)

N	Factors	#NI	# Estimable 2FI	nd	b	r	r_b
48	5×2^5	1	29^1	1	4	29	27^*
	2^{20}	3	19^3	3	20	19^s	0^{*o}
	2^{19}	3	19^3	3	20	19	0^{*o}
	$6 \times 4 \times 2^2$	3	$31^2, 29^1$	2	4	31	31^*
	6×2^7	45	35^{45}	4	8	35^s	28^*
	6×2^5	30	$34^1, 33^4, \dots, 27^1$	1	8	34	30^*
	$4 \times 3 \times 2^4$	19	$32^{14}, 31^5$	5	4	32	32^*
	4×2^{11}	560	33^{560}	10	12	33^s	22^{*o}
	4×2^{10}	2217	$33^3, 32^{2212}, 31^2$	3	12	33	22^o
	3×2^9	3	36^3	1	8	36^s	29^*
	3×2^7	3056	$35^{209}, 34^{299}, \dots, 21^8$	1	8	35	31^*
	2^{24}	60	23^{60}	1	24	23	0^{*o}
	2^{23}	130	23^{130}	1	24	23	0^{*o}
54	6×3^3	2	$40^1, 36^1$	1	9	40	34^*
	$3^5 \times 2$	4	41^4	2	9	41	33
	3^5	4	$39^1, 36^1, 35^1, 31^1$	1	18	35^t	20^{*o}
56	7×2^5	7	$40^4, 39^3$	3	4	40	40^*
64	$8 \times 4 \times 2^2$	4	$41^3, 39^1$	3	8	41	41^*
	8×2^7	924	49^{924}	4	8	49^s	42^*
72	8×2^5	192	$45^{13}, 44^{35}, \dots, 37^1$	3	8	45	44^*
	4^6	1	45^1	1	16	45^s	30^{*o}
	4^5	1	45^1	1	16	45	30^o
	$4^5 \times 2^2$	1	46^1	1	16	46^s	31^{*o}
	$4^5 \times 2$	1	45^1	1	16	45	30^o
	$4^4 \times 2^6$	1	45^1	1	16	45^s	30^{*o}
	$4^4 \times 2^5$	1	45^1	1	16	45	30^o
	$4^3 \times 2^8$	2	46^2	1	16	46^s	31^{*o}
	$4^3 \times 2^2$	107	$46^{17}, 45^{11}, \dots, 37^3$	1	16	46	37^{*o}
	$4^2 \times 2^{12}$	2159	45^{2159}	5	16	45^s	30^*
	$4^2 \times 2^5$	104949	$49^{3177}, 48^{16490}, \dots, 31^2$	1	16	49	36^o
	9×2^6	498	$57^{106}, 56^{186}, \dots, 53^6$	2	4	57^s	54^*
	9×2^5	96	$50^{77}, 49^{19}$	4	4	50	50^*
	$6^2 \times 2^2$	6	$45^5, 43^1$	4	12	45	45^*
	$6 \times 3 \times 2^4$	293	$44^{285}, 43^8$	1	12	44	44^*
	$4 \times 3^2 \times 2$	18	23^{18}	4	6	23	23^*
	$3^2 \times 2^{12}$	27	$55^1, 51^8, \dots, 37^3$	1	12	55^s	44^*
				8	12	51^t	$42^3, 43^3, 44^{2*}$
80	$5 \times 4 \times 2^6$	4814	$66^{93}, 65^{59}, \dots, 57^1$	2	4	66^s	63^*
	$5 \times 4 \times 2^5$	217319	$57^{210268}, 56^{5967}, \dots, 47^1$	6	4	57	57^*
81	9×3^4	2	64^2	2	9	64^s	56^*
	9×3^3	3	$60^1, 58^1, 54^1$	1	9	58^t	58^*
	3^{10}	1	60^1	1	27	60^s	inf
				1	9	60^s	52^*
	3^9	1	60^1	1	27	60	36^{*o}

Table 5 includes the most challenging types of strength-3 OAs, namely the OAs involving the largest numbers of factors possible for a given set of factor levels. For example, the arrays of the type $OA(40; 5 \times 2^6)$ are included because six is the maximum number of two-level factors in an array of the type $OA(40; 5 \times 2^a)$. Many of the arrays in the table belong to the class of second

order saturated (SOS) designs, meaning that the rank of their model matrix for a model with an intercept, all main effects and all two-factor interactions equals the number of runs (Cheng et al., 2008). Arranging such arrays in b blocks which are orthogonal to the main effects automatically reduces the number of estimable two-factor interaction contrasts by $b - 1$, independent of what orthogonal blocking approach is used. SOS designs are therefore not suitable for evaluating the ability of the MILP approach to maximize the number of estimable two-factor interaction contrasts. Obviously, the MILP approach remains useful as a technique to search for a blocking arrangement of SOS designs that allows independent main-effect estimates, which is a challenge in itself. Also, the MILP approach prioritizes minimization of the most severe confounding such that no single interaction is completely confounded with blocks, if possible.

In Table 5, we indicate OAs that belong to the class of SOS designs by means of the superscript s . Whenever we encountered an SOS design, we also studied similar OAs which do not belong to the class of SOS designs. Table 5 also includes results for these designs. For example, all arrays of the type $\text{OA}(48, 3 \times 2^9)$ are SOS designs. We therefore studied arrays with fewer two-level factors until we obtained designs which do not possess the SOS property. It turns out that the maximum number of two-level factors for which non-SOS designs exist is seven, which is why we have also studied arrays of the type $\text{OA}(48, 3 \times 2^7)$. As a matter of fact, 209 arrays of that type offer 35 estimable two-factor interaction contrasts, while 38 degrees of freedom are available after estimating the main effects and the block effects. In that case, arranging the OAs in b blocks does not necessarily lead to a decrease of $b - 1$ in the number of estimable two-factor interaction contrasts.

4.2 Blocking results

In this section, we discuss the results of applying the MILP approach to the OAs listed in Table 5. We arranged the arrays in the maximum number of blocks for which an orthogonal arrangement might exist. That number of blocks, b , is shown in the sixth column of Table 5. The corresponding blocking arrangements are given in the supplementary materials.

We studied the largest numbers of blocks possible because it is generally harder to find optimal orthogonal blocking arrangements for larger numbers of blocks. Moreover, as soon as a blocking arrangement with a large number of small blocks has been found, it is possible (but not necessarily optimal) to construct orthogonal blocking arrangements with fewer blocks simply by merging blocks. Merging blocks generally leads to a larger number of estimable two-factor

interaction contrasts.

For all but 32 designs that were orthogonally blocked by the MILP approach, an orthogonal blocking arrangement was also produced by the CN benchmark approach. In the remainder of this section, we discuss the computing times for both approaches, the sizes of the MILP problems in terms of the numbers of variables and constraints, the numbers of estimable two-factor interaction contrasts, and the D-efficiencies for several models.

4.2.1 Computing times and size of the MILP problem

The MILP approach is computationally demanding due to the branch-and-bound procedure it uses to make all elements b_{ij} of the blocking matrix $\mathbf{B}$ binary. Roughly speaking, the branch-and-bound tree involves many nodes (at each of which a binary decision variable is either set to zero or one), and all these nodes need to be evaluated to ensure the optimality of the solution found. To limit the computing time, we provided the software with a maximum number of nodes to evaluate, hereby accepting the possibility that the program may stop before it obtains an optimal solution. It turned out that evaluating 10,000 nodes sufficed to obtain a good solution in all but three cases that we studied. This is because the branch-and-bound procedure seems capable of finding a good solution relatively fast, and the bulk of the computing time serves to confirm that that solution is a high-quality one.

Table 5 lists the 54 OA types we studied. Two OA types ($\text{OA}(72, 3^2 \times 2^{12})$ and $\text{OA}(81, 3^{10})$) were considered twice, resulting in 56 cases in total. The table contains three kinds of results: (i) in 26 cases (indicated by the superscript o in the table’s last column), the MILP problem was solved to optimality before the maximum number of nodes was reached; (ii) in 29 cases, a solution was found for the MILP problem, but the maximum number of nodes did not suffice to establish the optimality of the solution or to find a better solution; and (iii) for the array of the type $\text{OA}(81; 3^{10})$ in 27 blocks, the MILP problem turned out to be infeasible. This is indicated in the table with the abbreviation ‘inf’. The time required by the MILP approach to establish the infeasibility of the problem was only a few seconds. Note that it was possible to obtain an orthogonal arrangement in nine blocks for the array of the type $\text{OA}(81; 3^{10})$, which demonstrates that it is easier to find an orthogonal blocking arrangement for large blocks than for small ones.

Detailed computing times for the MILP approach are given in the supplementary tables. Here, we compare the MILP computing times with those of the CN benchmark approach. To do so, we first used as few as 50 tries of the CN benchmark algorithm. For arrays that could

Table 7: Number of designs (ND) blocked orthogonally after T tries of the CN benchmark approach, the approximate average computing time (in minutes) for the CN benchmark approach and the average computing time (in minutes) for the MILP approach.

T	ND	CN time	MILP time
$T \leq 50$	59	0.02	8
$50 < T \leq 500$	25	0.2	10
$500 < T \leq 5000$	14	2	15
$5000 < T \leq 50,000$	5	20	11
$T > 50,000$	32	> 20	22

not be blocked orthogonally using 50 tries, we conducted 500 tries. For arrays that could not be blocked orthogonally using 500 tries, we conducted 5,000 tries, and, finally, for all cases where 5,000 tries did not suffice, we performed 50,000 tries. The numbers of cases for which 50, 500, 5,000 and 50,000 tries sufficed are shown in Table 7, along with the number of cases where 50,000 tries was not enough to find an orthogonal blocking arrangement. Computing times for the CN benchmark approach increase roughly linearly with the number of tries, from 1.2 seconds for 50 tries to 20 minutes for 50,000 tries.

In Table 7, we compare the computing times for the MILP approach with those of the CN benchmark approach. The 27-block arrangement of $OA(81; 3^{10})$ was excluded from the comparison because it was infeasible. The table thus reports the results for 135 blocked designs.

The CN benchmark was able to arrange more than 70% of the OAs in blocks in 5,000 tries or less. So, in general, the CN benchmark requires less computation time than our MILP approach. However, our main reason to develop a MILP approach was to minimize the confounding of the two-factor interaction contrasts with the blocks. This type of confounding is ignored in the CN benchmark approach.

For 32 of the 135 designs, the CN benchmark approach did not manage to produce an orthogonal blocking arrangement using 50,000 tries. For these cases, the MILP approach did produce an orthogonal blocking arrangement in 22 minutes, on average. The supplementary tables give a more detailed account of the computing times of the CN benchmark approach.

We conclude that the MILP approach produces an orthogonally blocked design within an acceptable computing time, and that it is capable of producing orthogonal blocking arrangements even for challenging cases where the CN benchmark approach fails. A strength of the MILP approach is its ability to identify cases where an orthogonal blocking arrangement does not exist, unlike the CN benchmark approach.

4.2.2 Estimable two-factor interaction contrasts

The last two columns of Table 5, labeled r and r_b , show the number of estimable two-factor interaction contrasts in the absence of blocking and in the presence of blocking, respectively. Whenever $r - r_b < b - 1$, fewer than $b - 1$ estimable two-factor interaction contrasts have to be sacrificed to obtain a good blocking arrangement. In particular, for all cases where $r = r_b$, no estimable two-factor interaction contrasts are lost due to the blocking. This is the best possible result.

In Table 5, we have indicated all the cases where the number of estimable two-factor interaction contrasts, r_b , is maximal with an asterisk. In most cases, this is because r_b reaches the upper bound

$$UB = \min \left\{ r, N - \left[b + \sum_{i=1}^m (s_i - 1) \right] \right\}, \quad (1)$$

where N is the number of runs, m is the number of factors, b is the number of blocks, and s_i is the number of levels of the i th factor. The upper bound UB is the minimum of the number of estimable two-factor interaction contrasts in the absence of blocking and the degrees of freedom remaining after fitting a model including the treatment factors' main effects and the blocks.

To illustrate the calculation of UB , consider arranging the array of the type $OA(27; 3^4)$ in nine blocks, so that $b = 9$, $m = 4$ and $s_1 = \dots = s_4 = 3$. In the absence of blocking, this array allows the estimation of 18 two-factor interaction contrasts, so that $r = 18$. Fitting a model involving a nine-level blocking factor and the main effects of all four three-level factors requires $9 + 4 \times 2 = 17$ degrees of freedom. Therefore, at most $27 - 17 = 10$ degrees of freedom are available to estimate two-factor interaction contrasts. Hence, the upper bound is $UB = \min(18, 10) = 10$. The blocking arrangement we obtained using the MILP approach indeed permits estimation of 10 two-factor interaction contrasts. For this reason, we put an asterisk in the final column for the array of the type $OA(27; 3^4)$ in Table 5.

The fact that the asterisks in the table are numerous shows that the MILP approach is very powerful. We are not sure whether the reported r_b value is the maximum one possible in only 15 of the 136 blocked designs. Six of these involve arrays of the type $OA(72; 3^2 \times 2^{12})$ with an r value of 51. However, two other arrays of that type do have blocking arrangements with the maximal r_b value of 44. Four further blocking arrangements for which we are not sure whether the reported r_b value is maximal involve 64-run arrays, namely those of the types $OA(64; 4^5)$, $OA(64; 4^5 \times 2)$, $OA(64; 4^4 \times 2^5)$ and $OA(64; 4^2 \times 2^5)$. The remaining five cases involve arrays of

the types $\text{OA}(48; 4 \times 2^{10})$ (three cases) and $\text{OA}(54; 3^5 \times 2)$ (two cases).

In four cases, indicated with the superscript t in Table 5, we blocked an array with a suboptimal value of r to maximize the r_b value. First, the 39-2FI array of the type $\text{OA}(54, 3^5)$ cannot be orthogonally blocked when 18 blocks of size three are desired. Therefore, we tried to arrange the 36-2FI array in 18 blocks. As this also failed, we continued with the 35-2FI array. For this array, the MILP approach did yield an orthogonal blocking arrangement in 18 blocks of size three. The arrangement has 20 estimable two-factor interaction contrasts. While the value of 20 does not equal the upper bound UB in this case, we confirmed the optimality of the r_b value using the MCS algorithm of Schoen et al. (2010): there is only one way to extend the original array by adding an 18-level factor, and this extension allows estimation of 20 two-factor interaction contrasts. Because of the optimality of r_b , we also indicated that case with an asterisk.

Two other cases where we obtained an optimal r_b value by blocking an array with a suboptimal value of r involve the arrays of the types $\text{OA}(32, 4 \times 2^6)$ and $\text{OA}(81, 9 \times 3^3)$. In these cases, the r_b value equals the upper bound UB . Finally, the array of the type $\text{OA}(72, 3^2 \times 2^{12})$ with an r value of 55 can be orthogonally blocked in 12 blocks of size six. However, the design is an SOS design. For this reason, we also studied the eight arrays with an r value of 51. Two of the eight arrangements attain the upper bound UB for r_b , while the other six did not.

Eight cases in Table 5 have an r_b value of zero. These cases involve two-level arrays with $N/2$ or $N/2 - 1$ factors. As noted by Schoen and Mee (2012), these OAs are all fold-over designs. For these arrays, the only blocks of size two that result in independent main-effect estimates consist of mirror image pairs. All two-factor interaction contrasts are then completely confounded with the blocks, so that r_b always equals zero. Given this result, it is natural to study blocks of size four. In the supplementary Section B, we show that, for any two-level fold-over design with a number of factors greater than or equal to $3N/8 + 1$ and greater than or equal to $N/2 - 4$, the only way to obtain arrangements involving four runs per block is to merge the blocks of size two. In any case, the r_b value of zero we obtained for the two-level designs in blocks of size two is optimal, which is why we have given these cases an asterisk as well.

In the supplementary tables, we compare the number of estimable two-factor interaction contrasts for the blocking arrangements produced by the MILP approach and the CN benchmark approach. However, most often, only a limited number of interactions turn out to be active when analyzing data from a designed experiment. Therefore, we also study the performance of the blocking arrangements obtained using the two approaches in terms of estimation efficiency for

Figure 1: Average D-efficiencies over 10,000 models with 4, 8 or 16 interaction contrasts based unblocked (white bullets) and MILP blocked (black bullets) versions of the designs.

models including a few interactions.

4.2.3 Efficiencies

Figure 1 shows the designs' average D-efficiencies based on 10,000 models with 4, 8 or 16 randomly chosen interaction contrasts and all the main effects. The efficiencies for the unblocked arrangements are shown as white bullets and those for the blocked arrangements are shown as black bullets; see Section 3.2.1 for the formulas used for the calculation.

We excluded the 20 pure two-level designs in blocks of size two from our efficiency study, because all their interaction contrasts are completely confounded with blocks and their D-efficiencies are zero for any model including two-factor interactions. Detailed efficiency results as well as Matlab programs used for the calculations are available in the supplementary tables and pro-

grams, respectively. The efficiencies for 16 interaction contrasts were only calculated for the 32 designs with 40 or more runs, as the smaller designs do not permit estimation of models with this large a number of interaction contrasts. In the figure, the white bullets represent the efficiencies of the designs in the absence of blocking, while the black bullets represent their efficiencies in the presence of blocking.

The figure shows that there are cases where an orthogonal blocking arrangement comes at a high price in terms of efficiency. One such case involves an array of the type $OA(54; 3^5)$ in 18 blocks. The average efficiency for models with four interaction contrasts drops from 99% to 69% in that case. For models involving eight contrasts, the average efficiency drops from 98% to 30%, while, for 16 contrasts, it drops to 0%. Since there is only one isomorphism class for this blocked design, this is the best possible efficiency. Of course, the efficiency drops shown in Figure 1 were computed assuming fixed block effects. If the block effects were considered random, the drop in efficiency would be smaller, especially if the variance of the random block effects were small.

We compared the performance of the MILP approach and the CN benchmark approach in terms of estimation efficiency for 49 of the 59 designs where the benchmark approach managed to find an orthogonal blocking arrangement in 50 tries. We exclude 10 two-level cases due to their zero efficiencies. For models with four interaction contrasts, 82% of the OAs had a better average efficiency when blocked using the MILP approach than when blocked using the CN benchmark approach. The average difference in efficiency was 1.8%. For models with 8 and 16 interaction contrasts, the MILP approach worked better in 73% and 94% of the cases, with average differences in efficiency of 3.8% and 14%, respectively. In those cases where the blocking arrangement produced by the CN benchmark approach was best, the difference in average efficiency ranged from 3.3% to 25.9%. The large difference in average efficiencies is due to the results for the eight arrays of the type $OA(32; 4 \times 2^7)$ we studied. For three of these OAs, the MILP approach leads to much smaller efficiencies than the CN benchmark approach. For the other arrays of that type, the performance of the MILP approach is similar to or even better than that of the CN benchmark algorithm. Both for the MILP approach and for the CN benchmark approach, the efficiencies strongly depend on the specific array of the type $OA(32; 4 \times 2^7)$ used, which suggests that it is useful to try out several arrays of the same type when planning a blocked experiment.

From the efficiency comparisons between the MILP approach and the CN benchmark approach, we conclude that the MILP approach generally results in orthogonal blocking arrange-

ments that have a D-efficiency that is roughly 1-4 % better than those produced by the CN benchmark approach.

5 Strength-4 designs

In this section, we compare the MILP approach and the CN benchmark approach when applied to strength-4 designs involving 32–64 runs. The supplementary tables include the blocking arrangements obtained using the MILP approach as well as full details of the comparisons we made. Here, we summarize our findings.

Unlike the strength-3 designs, strength-4 designs permit estimation of a model involving all main effects and all two-factor interaction effects when run in a completely randomized fashion. For this reason, the CN benchmark approach can be applied in two different ways. First, the model specified for the treatment factors can be a simple main-effects model, as was done in the previous sections for the strength-3 designs. Second, the model specified for the treatment factors can be one involving main effects and two-factor interactions. We studied both versions of the CN benchmark approach for the strength-4 designs.

Table 8 shows the strength-4 designs studied, along with the numbers of blocks b , the computing time for the MILP procedure, and the average efficiencies for 10,000 models containing all main effects and eight two-factor interaction contrasts. For every case, we used 5,000 tries of the two versions of the CN benchmark approach.

For the 64-run designs, we considered arrangements in eight as well as 16 blocks. There is a substantial increase in computing time when 16 blocks are desired, showing that it is harder to arrange a given design in many small blocks than in few large blocks. In all but one case, the arrangements in eight blocks were found within half an hour. Finding arrangements in 16 blocks took as long as nine hours in one case. For this case, both versions of the CN benchmark were unable to find an orthogonal blocking arrangement in 5,000 tries.

As shown in the columns labeled ‘ r ’ and ‘ r_b ’, the eight-block arrangements as well as three of the five 16-blocks arrangements obtained by the MILP approach permit estimation of all two-factor interaction contrasts. The two remaining 16-block arrangements had one and three fewer estimable interaction contrasts in the presence of blocking.

The columns labeled ‘ME’ and ‘ME+INT’ in Table 8 report the D-efficiencies for the two versions of the CN benchmark approach, in case an orthogonal blocking arrangement was found.

Table 8: Strength-4 designs arranged in blocks using the MILP approach and the CN benchmark approaches, as well as the MILP computing time (in minutes) and the D-efficiencies for the MILP approach and the benchmark approaches (ME: main-effects model; ME+INT: model involving main effects and interactions).

b	N	Factors	Design	Time	r	r_b	Average D-efficiency		
							ME	ME+INT	MILP
8	32	2^6	1	3	15	15	78.46		79.56
			2	3	15	15	38.10	86.16	86.45
	48	3×2^5	1	17	20	20	87.80		86.54
			2	22	20	inf		-	
			3	0	20	inf		-	
	64	4×2^6	1	20	33	33	91.83	96.53	94.23
			2	134	33	33	92.20	96.68	96.60
	64	2^8	1	8	28	28	91.00		94.01
			2	10	28	28	92.56		94.57
			3	7	28	28	92.56		94.75
16	64	4×2^6	1	65	33	32	80.74		81.61
			2	286	33	33	80.30		81.13
	64	2^8	1	67	28	28	80.76		83.77
			2	522	28	28			82.20
			3	111	28	25	77.94		76.96

The first approach, where a main-effects model was specified, was much more successful in finding orthogonal blocking arrangements, than the second, where a model involving main effects and interactions was specified. The latter approach found orthogonal blocking arrangements for three cases only, compared to 12 for the former approach. The MILP approach produced an orthogonal blocking arrangement in 13 cases, and indicated that the two remaining cases are infeasible.

Comparing the D-efficiencies of the MILP approach and the first version of the CN benchmark approach (involving the main-effects model), we see that the MILP approach leads to efficiencies that are 48% better in one case, a few percent better in 11 cases, and 1.0-1.3% worse in two cases. The second version of the CN benchmark approach (involving the model with main effects and interactions) usually fails to produce an orthogonal blocking arrangement. However, when it does succeed in producing one, it is a high-quality blocking arrangement in terms of efficiency.

6 Discussion

This paper presents a mixed integer linear programming (MILP) approach for finding orthogonal blocking arrangements of orthogonal designs with a strength of 3 or 4, which are typically used

when two-factor interactions are of interest. The use of the MILP approach is preferable to using conventional methods for at least two reasons. First, the method explicitly requires orthogonal blocking. This feature is not shared by the benchmark approaches we discussed, because both of these approaches require checking the orthogonality between blocks and main effects.

The second reason to prefer the MILP approach is the fact that it directly addresses the confounding between the two-factor interactions and the blocks. This is accomplished by first minimizing the worst kind of confounding between the two-factor interaction contrasts and the blocks, and subsequently minimizing the total confounding. The estimation efficiencies for models involving a few interaction contrasts are usually 1-2% higher for blocking arrangements produced by the MILP approach than for those produced by the CN benchmark approach.

A very large body of literature on blocking is restricted to regular two-level designs. One of the most recent papers on this issue presents minimum aberration blocking arrangements for 128-run designs involving up to 64 factors (Xu and Mee, 2010). Following a suggestion from one of our referees, we compared the performance of the MILP approach to that of these blocking arrangements. Therefore, we studied the blocking of 128-run minimum aberration designs in 10, 20, 30 and 40 factors in 16 or 32 blocks. For the 40-factor design involving 16 and 32 blocks and for the 30-factor design involving 32 blocks, the MILP formulation was too big to solve on a computer with 16 GB of RAM. For the remaining five cases, we obtained a solution in 8–74 hours.

For the minimum aberration blocking arrangements as well as for those obtained with the MILP approach, we determined $A_{2.1}$ values, which measure the sum of squared correlations between block contrasts and two-factor interaction contrasts, the r_b values as discussed in Section 4.2.2, and the average D-efficiencies over 10,000 models with all the main effects and randomly chosen sets of eight two-factor interactions, corrected for blocking. For the 30-factor design in 16 blocks, the $A_{2.1}$ value of the MA blocked design was 50, while the MILP design has a value of 58, which is worse. However, the r_b values of the MA and MILP blocked designs were 123 and 126, respectively, while the respective efficiencies were 0.26 and 0.70. A comparison of the other four pairs of blocking arrangements shows similar features. All details as well as the blocked designs are available in the supplementary materials.

Our findings for the nonregular strength-3 and strength-4 designs suggest that designs with up to 81 runs can be blocked in an acceptable computing time using the MILP approach. Our results for the 128-run case suggest two interesting research topics. First, it is useful to seek MILP

methods dedicated to two-level designs which results in faster solutions. Second, nonregular blocking of regular designs might result in a more efficient detection of two-factor interactions than regular blocking techniques.

The objective function in our MILP method prioritizes the minimization of the largest absolute value in the matrix $\mathbf{W}^T\mathbf{B}$ over the minimization of the sum of all absolute values of the elements of $\mathbf{W}^T\mathbf{B}$. One of the referees suggested that some researchers might prefer prioritizing the sum of the absolute values instead of the maximum absolute value. To demonstrate the effect of this kind of prioritization, we arranged an array of the type $\text{OA}(48; 2^{10})$ in six blocks of size eight following our original approach and following the suggestion of the referee. Using our original approach, the maximum absolute value inside $\mathbf{W}^T\mathbf{B}$, which is a 45×6 matrix, equals 4, while the sum of the absolute elements equals 480. Using the referee’s suggestion, the maximum absolute element of $\mathbf{W}^T\mathbf{B}$ equals 8, while the sum of absolute elements drops to 248. The average D-efficiency over 10,000 models with all the main effects and eight two-factor interactions is 90.99% for the original approach and 90.39% for the alternative approach. In our Matlab program in the supplementary materials, we allow the user to choose which prioritization is used.

Finally, the MILP approach can also be applied to strength-2 orthogonal designs. However, strength-2 designs are used when there is less emphasis on estimating the two-factor interactions. The main concern when arranging strength-2 designs in blocks is therefore to ensure that the main effects are orthogonal to the blocks. Moreover, as strength-2 designs are generally smaller than strength-3 or strength-4 designs, complete catalogs exists of all possible orthogonal blocking arrangements. Schoen et al. (2013) show how to use such catalogs to identify the best possible blocking arrangements in terms of several criteria. Therefore, while our MILP approach is fully applicable to strength-2 designs as well, its added value compared to existing methods seems smaller for that type of designs than for designs with a higher strength.

SUPPLEMENTARY MATERIAL

Supplementary sections.pdf: Section A provides a detailed explanation of the MA benchmark approach based on interaction components that are calculated by modular arithmetic. Section B is a study of two-level fold-over designs with blocks of size four, discussing numbers of factors for which these blocks can only be formed by combining two fold-over pairs.

Supplementary tables.zip: Excel files with details on the computing times for the MILP and

CN blocking approaches, as well as efficiencies of the resulting blocking arrangements.

Programs.zip: Matlab and SAS programs for the MILP blocking approach, as well as Matlab programs to compute efficiencies of the blocked designs under various models.

Designs.zip: Text files with all the blocking arrangements discussed in the paper.

Acknowledgments

The research of the first and third author was supported by the Flemish Fund for Scientific Research FWO. The authors are grateful to Nha Vo-Thanh for the MATLAB code he developed and for his help when preparing a revised version of this paper. The authors are also grateful to two referees and an associate editor; their comments led to a more in-depth treatment of the proposed blocking approach. Finally, thanks are due to Ellen van den Heuvel and the management of FrieslandCampina for their help and for allowing us to reveal some details of the calcium fortification study.

References

- Brouwer, A. E., Cohen, A. M., and Nguyen, M. V. M. (2006). Orthogonal arrays of strength 3 and small run sizes. *Journal of Statistical Planning and Inference*, 136:3268–3280.
- Bulutoglu, D. A. and Margot, F. (2008). Classification of orthogonal arrays by integer programming. *Journal of Statistical Planning and Inference*, 138:654–666.
- Capehart, S. R., Keha, A., Kulahci, M., and Montgomery, D. C. (2011). Designing fractional factorial split-plot experiments using integer programming. *International Journal of Experimental Design and Process Optimisation*, 2:34–57.
- Cheng, C. S., Mee, R. W., and Yee, O. (2008). Second order saturated orthogonal arrays of strength three. *Statistica Sinica*, 18:105–119.
- Cook, R. D. and Nachtsheim, C. J. (1989). Computer-aided blocking for factorial and response-surface designs. *Technometrics*, 31:339–346.
- Deng, L. Y. and Tang, B. (2002). Design selection and classification for Hadamard matrices using generalized minimum aberration criteria. *Technometrics*, 44:173–184.

- Hedayat, A. S., Seiden, E., and Stufken, J. (1997). On the maximal number of factors and the enumeration of 3-symbol orthogonal arrays of strength 3 and index 2. *Journal of Statistical Planning and Inference*, 58:43–63.
- Hedayat, A. S., Sloane, N. J. A., and Stufken, J. (1999). *Orthogonal Arrays: Theory and Applications*. Springer, New York.
- Ingram, D. and Tang, B. (2005). Minimum G aberration design construction and design tables for 24 runs. *Journal of Quality Technology*, 37:101–114.
- Rao, C. R. (1947). Factorial experiments derivable from combinatorial arrangements of arrays. *Journal of the Royal Statistical Society Supplement*, 9:128–139.
- Schoen, E. D., Eendebak, P. T., and Nguyen, M. V. M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs*, 18:123–140.
- Schoen, E. D. and Mee, R. W. (2012). Two-level designs of strength 3 and up to 48 runs. *Journal of the Royal Statistical Society Series C*, 61:163–174.
- Schoen, E. D., Sartono, B., and Goos, P. (2013). Optimal blocking arrangements for general resolution-3 designs. *Journal of Quality Technology*, 45:166–187.
- Sun, D. X., Li, W., and Ye, K. Q. (2008). An algorithm for sequentially constructing nonisomorphic orthogonal designs and its applications. *Statistics and Application*, 6:144–158.
- Winston, W. L. (2003). *Operations Research: Applications and Algorithms*. 4th edition; Boston: Duxbury Press.
- Wu, C. F. J. and Hamada, M. (2009). *Experiments: Planning, Analysis and Parameter Design Optimization*. Wiley, New York, 2nd edition.
- Xu, H. and Mee, R. W. (2010). Minimum aberration blocking schemes for 128-run designs. *Journal of Statistical Planning and Inference*, 140:3213–3229.