

Row-column arrangements of regular and nonregular fractional factorial two-level designs

Nha Vo-Thanh

Biostatistics Unit, Institute of Crop Science, University of Hohenheim, Germany

Eric D. Schoen

Faculty of Bioscience Engineering, KU Leuven, Belgium

TNO, Zeist, Netherlands

Peter Goos

Faculty of Bioscience Engineering and Leuven Statistics Research Centre (LSTAT), KU Leuven, Belgium

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

August 10, 2018

Abstract

Row-column designs involve two crossed blocking factors. We propose three combinatorial criteria and a criterion based on D-efficiency to identify optimal row-column arrangements of fractional factorial two-level designs. We search for such arrangements in complete catalogs of 16-run and 24-run orthogonal arrays of strength 2 with two-level treatment factors and two multi-level blocking factors. One of the combinatorial criteria is a good surrogate for the efficiency-based criterion. For searching similar catalogs of strength-3 orthogonal arrays, we adapt the criteria and conduct a complete search for optimal row-column arrangements in catalogs with 64 and 72 runs.

KEY WORDS: Aliasing; Confounding; Confounding Frequency Vector; Crossed Blocking Factors; Generalized Word-Length Pattern; Orthogonal Array.

Table 1: Names of the 12 treatment factors studied in the car tire experiment.

Factor	Name
X1	rubber compound
X2	number of ribs
X3	shoulder block position
X4	center line cut depth
X5	center line cut width
X6	shoulder cut width
X7	center line cut angle
X8	shoulder cut angle
X9	center line cut through
X10	shoulder cut through
X11	center line additional sipe
X12	shoulder additional sipe

1 Introduction

In this paper, we study orthogonal blocking arrangements for the experimental tests described by regular and nonregular orthogonal two-level treatment designs when there are two crossed blocking factors. Bailey (2008) refers to these kinds of arrangements as row-column designs, because the arrangements can be represented in a table in which the levels of the first blocking factor correspond to the table's rows and the levels of the second blocking factor correspond to the table's columns. A cell in the table's i th row and j th column then corresponds to one or more runs conducted at the i th level of the first blocking factor and the j th level of the second blocking factor. We call the blocking arrangements we study here orthogonal because the main effects of the blocking factors and the treatment factors are all orthogonal. By an n -factor regular two-level treatment design, we mean a 2^{n-p} fractional factorial design constructed using p interactions as generators (Box et al., 2005). By an n -factor nonregular two-level treatment design, we mean any n -factor two-level design that cannot be constructed using this method. Any treatment design whose number of runs is not a power of two is inevitably nonregular. When its number of runs is a power of two, a treatment design can be regular or nonregular.

The experiment that motivated our work was a car tire experiment involving the 12 two-level treatment factors shown in Table 1. The total number of experimental runs was 24. The experiment was conducted on four consecutive days. Every day, six runs were performed by three different drivers. The same set of three drivers was used each day. Since, usually, there is day-to-day and driver-to-driver variability, two blocking factors are involved in the experiment. The first, day, has four levels, while the second, driver, has three levels. Consequently, the required design is a row-column design involving four rows and three columns.

The design we recommend for the car tire experiment is shown in Table 2. One attractive feature of the design is that the main effects of the two blocking factors and the 12 treatment factors are orthogonal. Another attractive feature is that the average D-efficiency over the 2145 models with the main effects of all the treatment factors and a set of two two-factor interactions is 0.895, after correcting for the two blocking factors. The design therefore allows the detection of a few active two-factor interactions.

We obtained the row-column arrangement in Table 2 by searching through a catalog of orthogonal designs with 24 runs, one four-level factor, one three-level factor and 12 two-level factors. In doing so, we assumed that the two blocking factors do not interact, the effects of the four-level blocking factor and the three-level blocking factor enter the model in an additive fashion, and that estimating the block effects requires $(4 - 1) + (3 - 1) = 5$ degrees of freedom. This is in line with the traditional statistical analysis of data from experiments with two crossed blocking factors (see, for example, Bailey, 2008). Given that a row-column arrangement involving four rows and three columns defines 12 groups of treatments, an alternative approach would be to seek an arrangement of all 24 treatments in 12 blocks of two treatments. One weakness of this alternative approach is that it does not assign the treatments to the rows and the columns of the design. Of course, this assignment could be done at random, but, generally, such a random assignment will not be optimal. Another weakness of the alternative approach is that it may be infeasible to find an orthogonal blocking arrangement. This is because it is generally harder to create an orthogonal blocking arrangement involving many small blocks than to create orthogonal blocking arrangements involving a few larger blocks. Therefore, if there are r rows and c columns in an experiment with N runs, it may be possible to create orthogonal arrangements in r rows of N/r runs and c columns of N/c runs even in situations where it is infeasible to find an orthogonal blocking arrangement involving rc blocks of $N/(rc)$ runs.

The purpose of our paper is to propose general criteria for selecting arrangements in rows and columns for regular and nonregular two-level designs, and to demonstrate the usefulness of these criteria by identifying optimal arrangements in r rows and c columns of N -run designs that appear in existing complete catalogs of $r \times c \times 2^n$ orthogonal designs. We distinguish between strength-2 designs and strength-3 designs. These types of designs require a separate treatment, because they differ in the aliasing between main effects and two-factor interactions. In two-level designs of strength 2, all main effects are orthogonal to each other, but they are not orthogonal to the two-factor interactions. In two-level designs of strength 3, the main effects are orthogonal to each other and to the two-factor interactions, but the two-factor interactions are not orthogonal to each other. In $r \times c \times 2^n$ orthogonal designs of strength 2, the two-factor interactions of the two-level factors are generally also confounded with the main effects of the r -level row factor and the c -level column factor. In catalogs of strength-3 arrays, this is no longer the case.

Two of the criteria we propose minimize the aliasing among the treatment factors' effects and the con-

Table 2: Recommended treatment design for the 12-factor car tire experiment. Factor names are given in Table 1. The 24 treatments are orthogonally blocked with respect to days as well as to drivers.

Day	Driver	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1
1	2	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1
1	3	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1
1	3	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1
2	1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1
2	1	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1
2	2	-1	-1	1	1	-1	1	1	-1	1	-1	1	-1
2	2	1	1	-1	-1	1	-1	-1	1	-1	1	-1	1
2	3	-1	1	-1	1	1	-1	-1	1	1	-1	1	-1
2	3	1	-1	1	-1	-1	1	1	-1	-1	1	-1	1
3	1	-1	-1	1	1	1	-1	1	1	-1	1	-1	-1
3	1	1	1	-1	-1	-1	1	-1	-1	1	-1	1	1
3	2	-1	1	-1	-1	1	1	1	-1	1	1	-1	-1
3	2	1	-1	1	1	-1	-1	-1	1	-1	-1	1	1
3	3	-1	1	1	1	-1	-1	-1	-1	1	1	-1	1
3	3	1	-1	-1	-1	1	1	1	1	-1	-1	1	-1
4	1	-1	1	1	-1	-1	1	-1	1	-1	1	1	-1
4	1	1	-1	-1	1	1	-1	1	-1	1	-1	-1	1
4	2	-1	1	1	-1	1	-1	1	-1	-1	-1	1	1
4	2	1	-1	-1	1	-1	1	-1	1	1	1	-1	-1
4	3	-1	1	-1	1	-1	1	1	1	-1	-1	-1	1
4	3	1	-1	1	-1	1	-1	-1	-1	1	1	1	-1

founding of the treatment factors' interactions with the two blocking factors. Cheng et al. (2002) showed that, for completely randomized designs (i.e., designs without any blocking factor), these kinds of criteria are computationally cheap surrogates for the average D-efficiency over all possible models including all main effects and a limited number of two-factor interactions. This computational advantage is important when exploring large catalogs of designs. In this paper, we demonstrate that the two criteria we propose are also good surrogates for the average D-efficiency over all possible models with limited numbers of two-factor interactions, in the presence of two blocking factors. We also propose a third criterion for selecting row-column arrangements. That criterion differs from the other two because it maximizes the confounding between the two-factor interactions and the two blocking factors. This is a useful strategy for experimenters who start the identification of active factors by fitting a main-effects model in the treatment factors and the two blocking factors.

The rest of this paper is structured as follows. In Section 2, we review the existing optimality criteria for two-level strength-2 designs in the absence of any blocking factor and in the presence of a single blocking factor. **Next, in Section 3, we adapt these optimality criteria to deal with two crossed blocking factors so as to find optimal row-column arrangements of strength-2 designs.** In Section 4, we explore catalogs of strength-2 orthogonal designs with 16 and 24 runs in order to illustrate our approach. In Section 5, we simplify the criteria for strength-2 designs to identify optimal row-column arrangements in catalogs of 64- and 72-run strength-3 designs. We return to the car tire example in Section 6 and end the paper with a discussion of the strengths and weaknesses of our approach, along with possible extensions.

2 Optimality of strength-2 designs in the absence and in the presence of a single blocking factor

2.1 Criteria for two-level treatment designs

Criteria for evaluating N -run orthogonal two-level treatment designs in the absence of blocking are based on so-called J -characteristics. The most important J -characteristics are based on sets of three or four factors. The J -characteristic corresponding to a three-factor set is called a J_3 -characteristic and equals the absolute value of the sum of the N element-wise products of the three factors' levels in the N experimental runs. The J -characteristic of a set of four factors is called a J_4 -characteristic and equals the absolute value of the sum of the N element-wise products of the four factors' levels in the N runs of the experiment. The importance of the J_3 -characteristics lies in the fact that they measure the extent to which the main effects are aliased with the two-factor interactions. The importance of the J_4 -characteristics lies in the fact that they measure the extent to which the two-factor interactions are aliased with other two-factor interactions. Ideally, the J_3 - and J_4 -characteristics are all zero, in which case there is no aliasing among the main effects and two-factor interaction effects. Whenever J_3 - and J_4 -characteristics equal N , the number of runs, this indicates complete aliasing. Clearly, this is undesirable.

Deng and Tang (2002) showed that the only possible absolute values for the J -characteristics in N -run strength-2 two-level orthogonal designs equal $(N - 8j)$, where j is a non-negative integer of at most $N/8$. So, the only possible values for the J_3 - and J_4 -characteristics for strength-2 16-run designs are 16, 8 and 0, while those for strength-2 24-run designs are 24, 16, 8 and 0.

The F_3 and F_4 vectors, which are called confounding frequency vectors, list the frequencies of the J_3 - and J_4 -characteristics, starting with the frequencies for the largest possible J_3 - and J_4 -characteristics, and ending with the frequencies for the smallest possible J_3 - and J_4 -characteristics. Generally, the frequencies for the J_3 - and J_4 -characteristic value of zero are dropped from the F_3 and F_4 vectors.

As computationally attractive alternatives to the F_3 and F_4 vectors, Tang and Deng (1999) proposed generalized word counts of length 3 and length 4, respectively. To calculate the generalized length-3 word count, all J_3 -characteristics are first converted into **absolute** correlations between a main-effect contrast vector and a two-factor interaction contrast vector by dividing them by N . Next, the generalized length-3 word count, denoted by A_3^t (where the superscript t refers to the word 'treatment'), is the sum of the squares of all resulting correlations. The generalized word count of length 4, denoted by A_4^t , is defined in a similar fashion. Whenever all J_3 - and J_4 -characteristics are zero, A_3^t and A_4^t are also zero. Therefore, good treatment designs minimize the A_3^t and A_4^t values. The minimization of the A_3^t value is prioritized, to ensure

that the main effects are aliased with the two-factor interaction effects to the smallest possible extent.

2.2 Criteria for blocking arrangements involving a single blocking factor

While criteria for treatment designs only quantify the aliasing of the treatment factors' main effects with their two-factor interactions and the aliasing among pairs of two-factor interactions, criteria for blocking arrangements also have to take into account the confounding between the two-factor interactions and the blocking factor. Note that, since we only consider designs for which the treatment factors' main effects are orthogonal to the blocking factor, there is no confounding between the treatment factors' main effects and the blocks.

The confounding of two-factor interactions with a single blocking factor can be expressed using a scalar, which we refer to as A_3^b . To determine the A_3^b value for a given blocking arrangement, we first define a set of orthogonal contrast vectors for the blocking factor. Next, we calculate the correlations between these contrast vectors, on the one hand, and the two-factor interaction contrast vectors, on the other hand. The A_3^b value is the sum of all squared correlations obtained this way. Xu and Wu (2001) show that the A_3^b value does not depend on the particular set of orthogonal contrast vectors employed. Minimizing the A_3^b value leads to a blocking arrangement that minimizes the overall confounding of two-factor interactions with the blocking factor.

To select a design for a blocked experiment, both the treatment design and its arrangement in blocks have to possess desirable properties. More specifically, the A_3^t and A_4^t values (for the treatment design) as well as the A_3^b value (for the blocking arrangement) should be small. In general, these three objectives are conflicting. There is, however, a general agreement that minimizing the A_3^t value is more important than minimizing the A_4^t and A_3^b values, and should be prioritized. It is unclear, though, whether minimizing the A_4^t value should be prioritized over minimizing the A_3^b value, or vice versa.

For this reason, when discussing the regular blocking of regular two-level designs, Cheng and Wu (2002) propose two criteria to rank blocked regular two-level designs. In simplified form, the first criterion, called the W_1 criterion, sequentially minimizes the A_3^t , A_4^t and A_3^b values. So, the W_1 criterion prioritizes the minimization of the A_4^t value over the A_3^b value. The second criterion they propose, called the W_2 criterion, sequentially minimizes the A_3^t , A_3^b and A_4^t values. So, the W_2 criterion prioritizes the minimization of the A_3^b value over the A_4^t value. **Assuming random block effects, the differences between the blocks provide information on the variance of the block effects as well as on two-factor interactions confounded with the blocks. If only a few degrees of freedom are available, both kinds of effects cannot be separated well. Therefore, it is important to ensure that interactions are confounded with the blocks as little as possible. For this reason, Schoen et al. (2018), who study blocking arrangements of 24- and 28-run two-level designs, argue that the W_1 criterion is not suitable when the number of blocks is small. For this reason, we extend the W_2 criterion to scenarios with two crossed blocking factors in this paper, rather than the W_1 criterion.**

Cheng et al. (2004) address the blocking of orthogonal two-level designs, including nonregular designs. These authors use two versions of both the W_1 criterion and the W_2 criterion. One version is based on the A_3^t and A_4^t values, just like the original criteria of Cheng and Wu (2002). The other version uses the F_3 and F_4 vectors instead of the A_3^t and A_4^t values.

Schoen et al. (2013) address the blocking of general orthogonal designs (i.e., two-level, multi-level and mixed-level designs). In addition to using the W_1 and W_2 criteria of Cheng and Wu (2002), they introduce the W_1^- and W_2^- criteria. The difference between the W_1^- and W_2^- criteria and the W_1 and W_2 criteria is that the former two criteria maximize the A_3^b value, while the latter two criteria minimize this value. The benefit of maximizing the A_3^b value is that it leads to smaller root mean squared errors for models containing only the main effects of the treatment factors and the blocking factor. This implies that the power of significance tests for detecting active main effects is larger. Recently, Schoen et al. (2018) argue that the W_2^- criterion ought to be preferred over the W_1^- criterion if a main-effects model will be fitted first. We therefore focus on an extension of the W_2^- criterion to scenarios with two crossed blocking factors in this paper, rather than on an extension of the W_1^- criterion.

In addition to the W_1^- and W_2^- criteria, Schoen et al. (2013) also introduce the W_3 criterion. For two-level designs, this criterion can be considered as a simplified version of the W_2^G criterion of Cheng et al. (2004), in which the F_3 vector replaces the A_3^t value and a confounding frequency vector for the two-factor interactions and the blocks, **called the FA_{21} vector**, replaces the A_3^b value.

All blocking criteria discussed so far favor strength-3 designs over strength-2 designs, because $A_3^t = 0$ for strength-3 designs. For experiments with 24 runs, strength-3 designs exist with up to 12 factors. However, there exist strength-2 designs with up to 12 factors which outperform the strength-3 designs in terms of the number of estimable two-factor interactions. Because the number of estimable two-factor interactions is generally considered a useful criterion, Schoen et al. (2018) modified the original W_2 criterion so that it prioritizes the maximization of the rank R of the model matrix including an intercept, main effect contrast vectors and two-factor interaction contrast vectors. In other words, they modified the criterion so that it prioritizes the maximization of the number of estimable two-factor interactions, given that the intercept and the main effects (of the treatment factors and the blocking factor) need to be estimated.

3 Criteria for blocking arrangements involving two crossed blocking factors

In this section, we introduce our new criteria to select row-column arrangements of two-level designs for scenarios involving two crossed blocking factors. We illustrate the different criteria for a scenario with two four-level blocking factors, five treatment factors and 16 runs.

3.1 Literature review

For scenarios involving two crossed blocking factors, Cheng and Mukerjee (2003) provide a methodology to construct s^{n-k} regular fractional factorial designs and arrange them in s^r rows and s^c columns. This methodology has two major limitations. First, the number of runs and the numbers of rows and columns must be powers of s , where s has to be a prime number. Second, because Cheng and Mukerjee (2003) use regular designs, the resulting row-column designs may involve complete aliasing and complete confounding of two-factor interactions (when $s = 2$) or two-factor interaction components (when $s > 2$).

Based on a weighted mean efficiency factor criterion, Gilmour and Trinca (2003) proposed an algorithm for creating factorial row-column designs for quantitative factors and a response surface model. Their construction involves an interchange algorithm. Goos and Donev (2006a,b, 2007) used point-exchange algorithms to construct D-optimal row-column designs. However, both the approach of Gilmour and Trinca (2003) and that of Goos and Donev (2006a,b, 2007) require the number of experimental runs to be large enough to estimate the model involving all main effects and all two-factor interaction effects. In this paper, we focus on screening experiments which, generally, do not allow all two-factor interaction effects to be estimated at the same time. For this reason, for most scenarios studied in this paper, the approaches of Gilmour and Trinca (2003) and Goos and Donev (2006a,b, 2007) are not applicable. The criteria we present are intended to select row-column arrangements that provide as much information about the interaction effects as possible, and they are especially useful when a full interaction model is not estimable and the experimenter has no prior information concerning the importance of individual interaction effects.

3.2 Building blocks of the new criteria

We build on the existing criteria for scenarios with just one blocking factor to define four criteria to select row-column arrangements of orthogonal two-level treatment designs. Three of the four criteria are adaptations of the W_2 , W_2^- and W_3 criteria discussed in Section 2.2 for experiments with a single blocking factor. We call these criteria combinatorial criteria. Our fourth criterion, the IC_q criterion, explicitly deals with the estimation efficiency of models with a limited number of interaction effects. We start by presenting our three combinatorial criteria, and then discuss the IC_q criterion. Throughout, we assume that treatment factor interaction effects of orders larger than two are negligible.

The combinatorial criteria all address the confounding of two-factor interactions with a particular blocking factor. The confounding can be expressed using scalars, which we refer to as A_3^r for confounding with the row factor and A_3^c for confounding with the column factor, or using a confounding frequency vector called the FA_3^{c22} vector. The scalars A_3^r and A_3^c are calculated in the same way as the A_3^t value for arrangements involving a single blocking factor; see Section 2.2. Minimizing the A_3^r and A_3^c values leads to a blocking arrangement that minimizes the overall confounding of the two-factor interactions with the two blocking

factors. The FA_3^{rc22} vector, which we define in detail in Section 3.5, provides a more detailed picture of the confounding of two-factor interactions with the blocking factors. It is akin to the F_3 and F_4 vectors discussed above, and generalizes the FA_{21} vector utilized by Schoen et al. (2013).

Our combinatorial criteria to deal with two blocking factors are called W_2 , W_2^- and W_3 criteria, because they generalize similar criteria for the case of a single blocking factor (see Section 2.2). They all involve a vector the elements of which have to be minimized from left to right. The minimization vectors for the three criteria in the presence of two blocking factors are shown in Table 3. The minimization vectors include

- the rank R of the model matrix for a model including the intercept, the treatment factors' main effect contrast vectors and their two-factor interaction contrast vectors;
- the A_3^t value summarizing the aliasing between the main effects and the two-factor interactions in the treatment design;
- the A_4^t value summarizing the aliasing among the two-factor interactions in the treatment design;
- the A_3^r value measuring the confounding between, on the one hand, the two-factor interactions of the treatment factors and, on the other hand, the first blocking factor, corresponding to the rows of the row-column arrangement;
- the A_3^c value measuring the confounding between, on the one hand, the two-factor interactions of the treatment factors, and, on the other hand, the second blocking factor, corresponding to the columns of the row-column arrangement;
- the confounding frequency vectors F_3 , F_4 and FA_3^{rc22} , which provide more details concerning the various kinds of aliasing and confounding measured by the A_3^t , A_4^t , A_3^r and A_3^c values.

We always minimize the leftmost entry of any minimization vector first. Next, we minimize the second entry of the vector, while ensuring that the leftmost entry retains its minimum value. We continue this sequential minimization until we have minimized the rightmost entry of the minimization vector. We call a design that sequentially minimizes a minimization vector W_2 , W_2^- or W_3 optimal, depending on which vector was utilized.

Table 3: Minimization vectors for the W_2 , W_2^- and W_3 optimal row-column arrangements of strength-2 designs.

Criterion	Minimization vector					
W_2	$-R$	A_3^t	F_3	$A_3^r + A_3^c$	A_4^t	F_4
W_2^-		A_3^t	F_3	$-A_3^r - A_3^c$	A_4^t	F_4
W_3			F_3	FA_3^{rc22}		

3.3 W_2 criterion

The first entry of the minimization vector for the W_2 criterion is minus the rank R of the model matrix including the intercept, the main effect contrast vectors and the two-factor interaction contrast vectors for the two-level treatment design. As we are interested in designs with high ranks, we seek to maximize the rank or, equivalently, to minimize minus the rank. Among all treatment designs which maximize the rank R , the W_2 criterion first seeks to minimize the A_3^t value. This ensures that the global aliasing between the main effects of the treatment factors and the two-factor interactions is as small as possible. The next entry we minimize is the F_3 vector. In doing so, we try to avoid that certain main effects are severely aliased with certain two-factor interactions. In other words, for a given global amount of aliasing, we prefer having many effects that are aliased to a small extent over a few effects that are severely, possibly completely, aliased. For similar reasons, the F_4 vector enters the W_2 criterion's minimization vector immediately after the A_4^t value.

A key feature of the W_2 criterion is that it prioritizes the minimization of the confounding between the blocking factors and the treatment factors' interactions, before addressing the aliasing among the two-factor interactions in the treatment design. This can be seen from the fact that the sum $A_3^r + A_3^c$, which

measures the global confounding between the interactions and both blocking factors, precedes the A_4^t value in the minimization vector. That approach is especially useful when the numbers of rows and columns in the desired row-column arrangement are small. **In that case, it is not feasible to separate the available inter-block information on the variance of the block effects from that on the treatment factors' interactions. Therefore, it is better to make sure that the inter-block information on the treatment factors' interactions is minimized, by minimizing the confounding of these interactions with the blocks.**

3.4 W_2^- criterion

The main goal of the W_2^- criterion is to maximize the confounding between the two-factor interaction effects and the block effects, while still minimizing the aliasing of the main effects and the two-factor interactions of the treatment factors. The confounding is maximal if all two-factor interaction effects are completely confounded, in which case $A_3^r + A_3^c = 0.5n(n-1)$. Since maximizing the sum $A_3^r + A_3^c$ is the same as minimizing $-A_3^r - A_3^c$, we use the entry $-A_3^r - A_3^c$ in the W_2^- criterion's minimization vector in Table 3.

The W_2^- criterion is useful when a model including the main effects of all the treatment factors and the blocking factors is used to detect active treatment factors. Any active two-factor interaction effect then potentially causes an inflation of the mean squared error, and, as a result, an increase in the estimated standard errors of the main effects and a smaller power for the significance tests of these effects. One approach to limit this inflation of the mean squared error is to ensure that as many two-factor interaction effects as possible are confounded with the blocks. This is exactly what the W_2^- criterion seeks to achieve. All other things being equal, W_2^- optimal row-column arrangements will therefore result in a larger power for detecting **factors with large main effects**.

Note that, because the W_2^- criterion concentrates on the detection of active main effects, the rank of the interaction model matrix is not relevant for this criterion.

3.5 W_3 criterion

The W_3 criterion first sequentially minimizes the entries of the F_3 vector, which quantifies the aliasing between the main effects and the two-factor interactions of the treatment factors, starting with the most severe kind of aliasing and ending with the least severe kind. Due to the absence of the A_3^t value as well as the rank R from its minimization vector, the W_3 criterion prioritizes the minimization of the most severe aliasing between main effects and two-factor interactions over the total amount of aliasing. Therefore, we expect different designs to be optimal in terms of the W_3 criterion than in terms of the previous two criteria.

Among all designs that sequentially minimize the entries of the F_3 vector, the W_3 criterion seeks those that minimize the confounding of the two-factor interactions with the two blocking factors. The minimization of the most severe confounding with the two blocking factors is prioritized over the minimization of less severe confounding. To achieve this, for each two-factor interaction, we first determine the sum of the squared correlations of its contrast vector and each of the contrast vectors for the rows. The frequencies of the possible values of these sums of squared correlations form the FA_3^{r22} vector. The FA_3^{c22} vector is a similar vector with the column factor's contrast vectors replacing those of the row factor. Next, we combine the FA_3^{r22} and FA_3^{c22} vectors into a single vector, FA_3^{rc22} , so that this vector's entries quantify all the confounding of the two-factor interactions of the treatment factors with the two blocking factors. The leftmost entries of the FA_3^{rc22} vector correspond to the largest sums of squared correlations and thus to the most severe kinds of confounding with the blocking factors. For notational convenience, we omit the frequencies of the zero sums, which correspond to two-factor interactions that are orthogonal to the rows or columns, from the FA_3^{rc22} vector.

3.6 IC_q criterion

As mentioned above, our fourth criterion for selecting blocking arrangements explicitly deals with the estimation efficiency of models with a limited number of interaction effects. More specifically, our fourth criterion is the average D-efficiency over all models containing all main effects and q two-factor interactions, after correcting for the intercept and the effects of the two blocking factors. In other words, our fourth criterion is an average D_s -efficiency (Atkinson et al., 2007) for a specific subset of parameters in a model with an intercept, all main effects, all block effects and q treatment factor interactions. The specific subset

of parameters includes the n main effects and q two-factor interactions of the treatment factors. For a given model with q interactions, the D_s -efficiency is calculated as

$$D_s = |\mathbf{U}^T(\mathbf{I}_N - \mathbf{C}\mathbf{C}^T)\mathbf{U}|^{1/p}/N,$$

where \mathbf{U} is an $N \times p$ matrix containing all main-effect contrast vectors and q specific two-factor interaction contrast vectors, \mathbf{C} is an $N \times b$ matrix consisting of a column of ones and the normalized orthogonal contrast vectors for the two blocking factors with r and c levels, $p = n + q$, $b = 1 + (r - 1) + (c - 1) = r + c - 1$, and \mathbf{I}_N is the $N \times N$ identity matrix. Our fourth criterion is known as the IC_q criterion in the literature, where IC stands for information capacity. It was proposed initially by Sun (1993), and used later by Li and Nachtsheim (2000) in the context of completely randomized designs (i.e., in scenarios without any need for blocking). Cheng et al. (2002) point out that there is a negative relation between the IC_q value and the A_3^t and A_4^t values. In other words, they point out that designs with a large IC_q value tend to have small A_3^t and A_4^t values, and vice versa. A drawback of the IC_q criterion is that it is computer-intensive, especially for designs with many factors and large values of q . In the strength-2 examples in this paper, we use $q = 1$ for 16-run row-column designs and $q = 2$ for 24-run row-column designs. In doing so, we favor designs that allow a precise estimation of a limited number of interaction effects.

One of the goals of this paper is to compare our results for the combinatorial criteria with those for the computationally intensive IC_q criterion. The combinatorial criteria have the advantage that they are computationally less demanding than the IC_q criterion, but they are not directly linked to any statistical property such as estimation efficiency.

3.7 Tied designs

On many occasions, we identified more than one design that is optimal with respect to a given combinatorial criterion. On those occasions, we used the IC_q criterion as a tie breaker. For the W_2 and W_3 criteria, we report the optimal design with the best (largest) value for the IC_q criterion. For the W_2^- criterion, we report the optimal design with the worst (smallest) value for the IC_q criterion. We do so because a small value for the IC_q criterion indicates that the corresponding blocking arrangement does not involve much information on the two-factor interactions within the blocks and that the interactions are confounded with the blocks to a considerable extent. Consequently, the inflation of the mean squared error when estimating a main-effects model should be limited when that blocking arrangement is used. Hence, small values for the IC_q criterion are consistent with the rationale for using the W_2^- criterion.

3.8 Illustration

Table 4 presents four different row-column arrangements of five-factor 16-run designs for a scenario with two four-level blocking factors. The performance of the four row-column arrangements in terms of the three combinatorial criteria and in terms of the IC_1 criterion is shown in Table 5. The row-column arrangements labeled D_1 and D_2 involve the same treatment design, namely a regular half fraction of the 2^5 full factorial design. The design has a resolution of V, meaning that the intercept, all five main effects and all ten two-factor interactions are orthogonal to each other. Therefore, the rank R equals 16 for that design.

The treatment combinations in row-column arrangement D_3 involve a duplicated regular half fraction of the 2^4 full factorial design for the first four factors. The first half fraction is run at one level of the fifth factor, while the second half fraction is run at the other level of the fifth factor. The resulting treatment design is regular, but it is not the regular fractional factorial design that is considered to be the best. For row-column arrangement D_3 , the rank R is only 13.

Finally, row-column arrangement D_4 involves a nonregular treatment design. For three out of the ten sets of three factors, the main effect contrast vector for one factor has an absolute correlation of 0.5 with the contrast vector of the interaction between the other two factors. Hence, the treatment design involves partially aliased effects. It has a rank R of 16.

Since the R values for the row-column arrangements D_1 , D_2 and D_4 are all 16 and thus larger than that for arrangement D_3 , arrangement D_3 is inferior to the others in terms of the W_2 criterion. Because, unlike D_1 and D_2 , row-column arrangement D_4 has a non-zero A_3^t value, it is not optimal in terms of the W_2 criterion. Row-column arrangements D_1 and D_2 do not only have the same value for R and for A_3^t , but they also have

Table 4: Four 16-run designs with five two-level factors arranged in four rows and four columns.

Row	Column	D ₁					D ₂					D ₃					D ₄							
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	0	0	0	1	1	0	0
0	2	1	1	1	0	1	1	1	1	0	1	1	1	0	0	1	1	1	1	0	0	1	0	1
0	3	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	1	1	1	0	1	0	0
1	0	0	1	1	0	0	0	1	1	0	0	0	1	0	1	0	0	1	1	0	1	1	1	1
1	1	0	1	1	1	1	0	1	1	1	1	1	0	1	0	0	0	0	1	1	0	0	0	0
1	2	1	0	0	1	0	1	0	0	0	1	0	1	1	0	1	1	0	0	1	0	0	1	0
1	3	1	0	0	0	1	1	0	0	1	0	1	0	0	1	1	1	0	0	0	1	0	0	1
2	0	1	0	1	1	1	1	0	1	1	1	1	1	0	1	0	1	1	1	0	0	1	0	0
2	1	1	0	1	0	0	1	0	1	0	0	0	1	0	1	1	1	1	0	1	0	1	0	0
2	2	0	1	0	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1	1	1	1	1	1
2	3	0	1	0	1	0	0	1	0	0	1	0	1	1	0	0	0	0	1	0	0	0	1	1
3	0	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
3	1	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	0	0	1	0	1	0	1	1
3	2	0	0	1	1	0	0	0	1	1	0	0	0	1	1	0	0	0	1	0	0	0	0	0
3	3	0	0	1	0	1	0	0	1	0	1	0	0	0	0	1	0	0	1	1	0	1	0	0

the same A_3^r , A_3^c and A_4^t values, and the same confounding frequency vectors F_3 and F_4 . Therefore, in terms of the combinatorial properties appearing in the minimization vector of the W_2 criterion, the row-column arrangements D_1 and D_2 are alike. However, they do differ in terms of their IC_1 value. Arrangement D_1 has an IC_1 value of 0.56, whereas arrangement D_2 has an IC_1 value of 0.40 only.

In terms of the W_2^- criterion, row-column arrangement D_4 is outperformed by the other three because of its non-zero A_3^t value. Arrangement D_3 is outperformed by arrangements D_1 and D_2 because of its non-zero A_4^t value. Finally, when using the W_2^- criterion, we prefer arrangement D_2 over D_1 because it has the worst IC_1 value of the two. We therefore expect row-column arrangement D_2 to cause a smaller inflation of the mean squared error when estimating a main-effects model. **Row-column arrangement D_2 is a regular two-level row-column designs reported by Cheng and Mukerjee (2003).**

Because the F_3 vector of arrangement D_4 is nonzero, that arrangement is less attractive than the others in terms of the W_3 criterion. The other three row-column arrangements all have a zero F_3 vector, but they differ in the confounding of the two-factor interactions with the blocking factors. When using arrangement D_1 , for example, four two-factor interactions are completely confounded with the blocking factors. When using arrangement D_2 , six two-factor interactions are completely confounded with the blocking factors. Finally, when using arrangement D_3 , only two two-factor interactions are completely confounded with the blocking factors. This can be seen from the first entries of the FA_3^{rc22} vectors for D_1 , D_2 and D_3 , which equal 4, 6 and 2, respectively, and which correspond to perfect correlations between two-factor interactions and contrast vectors of the blocking factors. In other words, these entries indicate how many two-factor interactions are completely confounded with the blocking factors. Because, **among the designs with a zero F_3 vector**, it has the fewest completely confounded two-factor interactions, row-column arrangement D_3 is the best in terms of the W_3 criterion.

Finally, row-column arrangement D_4 is the best in terms of our fourth criterion, because it has the largest IC_1 value. The difference in IC_1 value with arrangement D_3 is, however, small. **The resolution-V treatment design is the best treatment design when there are no blocking factors. However, the IC_1 values that can be reached using that design when there are two blocking factors are smaller than the value for the IC_1 optimal row-column arrangement because of the substantial confounding of the two-factor interactions with the blocks in the event the resolution-V treatment design is used. This is witnessed by row-column arrangements D_1 and D_2 , each of which are optimal arrangements of the resolution-V treatment design and produce small IC_1 values.**

Table 5: Performance of the row-column arrangements in Table 4 in terms of the optimality criteria defined in Table 3 and in terms of the IC_1 value. The entries of the F_3 and F_4 vectors are the frequencies of J_3 - and J_4 -characteristics of 16 and 8. The entries of the FA_3^{rc22} vector are the frequencies with which interactions are completely confounded and semi-confounded with the blocking factors.

W_2 criterion:	$-R$	A_3^t	F_3	$A_3^r + A_3^c$	A_4^t	F_4	IC_1
D ₁	-16	0	(0,0)	6	0	(0,0)	0.56
D ₂	-16	0	(0,0)	6	0	(0,0)	0.40
D ₃	-13	0	(0,0)	6	1	(1,0)	0.71
D ₄	-16	0.75	(0,3)	5	0	(0,0)	0.73
W_2^- criterion:		A_3^t	F_3	$-A_3^r - A_3^c$	A_4^t	F_4	IC_1
D ₁		0	(0,0)	-6	0	(0,0)	0.56
D ₂		0	(0,0)	-6	0	(0,0)	0.40
D ₃		0	(0,0)	-6	1	(1,0)	0.71
D ₄		0.75	(0,3)	-5	0	(0,0)	0.73
W_3 criterion:			F_3		FA_3^{rc22}		IC_1
D ₁			(0,0)		(4,4)		0.56
D ₂			(0,0)		(6,0)		0.40
D ₃			(0,0)		(2,8)		0.71
D ₄			(0,3)		(1,8)		0.73

4 Optimal row-column arrangements of strength-2 designs

Our approach to find optimal row-column arrangements of two-level designs of strength 2 starts from the set of all non-isomorphic orthogonal arrays (OAs) of the type $OA(N; r \times c \times 2^n; 2)$, where N is the number of experimental runs, n is the number of two-level treatment factors, and r and c represent the numbers of levels of the first and second blocking factor, respectively. The last ‘2’ in the notation $OA(N; r \times c \times 2^n; 2)$ indicates the minimum strength of the arrays contained within the set.

Any orthogonal array of the type $OA(N; r \times c \times 2^n; 2)$ defines an acceptable row-column arrangement, in the sense that the treatment factors’ main effects are orthogonal to the two blocking factors. The remaining challenge is to find the best possible row-column arrangement of an n -factor two-level treatment design of the type $OA(N; 2^n; 2)$ within the complete $OA(N; r \times c \times 2^n; 2)$ catalog of acceptable row-column arrangements. Throughout the paper, we assume that $r \geq c$, without loss of generality.

In this section, we use the W_2 , W_2^- , W_3 and IC_q criteria to evaluate complete catalogs of non-isomorphic orthogonal arrays of the types $OA(16; 4 \times 4 \times 2^n; 2)$, $OA(24; 4 \times 3 \times 2^n; 2)$ and $OA(24; 6 \times 4 \times 2^n; 2)$ (Schoen et al., 2010) to find optimal row-column arrangements of 16- and 24-run two-level designs with n factors. The numbers of non-isomorphic arrays in the catalogs are shown in Table 6. In order to find the optimal row-column arrangements, we had to explore all these non-isomorphic orthogonal arrays and evaluate them in terms of the W_2 , W_2^- , W_3 and IC_q criteria, which was a major computational effort.

Supplementary materials to this paper, available at <http://www.asq.org/pub/jqt/>, include all optimal row-column arrangements of strength-2 treatment designs as well as tables with their characterization in terms of the F_3 , F_4 , FA_3^{rc22} and FA_3^{c22} vectors, the IC_q value and the rank R . In this section, we discuss our main findings.

4.1 Designs with 16 runs

Figure 1 shows the main results for the optimal row-column arrangements of 16-run two-level designs for scenarios with four rows and four columns. The circles in the figure indicate designs that are optimal in terms of the combinatorial W_2 , W_2^- or W_3 criteria, while the black dots represent the arrangements that are optimal in terms of the IC_1 value. Finally, the crosses correspond to the benchmark designs of Cheng and Mukerjee (2003).

The left panel of Figure 1 shows the total amount of confounding of the two-factor interactions with the rows and the columns, as measured by the sum $A_3^r + A_3^c$. Obviously, this total amount of confounding increases with the number of factors. Optimal designs for a given number of factors do not vary much in

Table 6: Numbers of non-isomorphic orthogonal arrays of the types $OA(16; 4 \times 4 \times 2^n)$, $OA(24; 4 \times 3 \times 2^n)$ and $OA(24; 6 \times 4 \times 2^n)$.

n	$OA(16; 4 \times 4 \times 2^n)$	$OA(24; 4 \times 3 \times 2^n)$	$OA(24; 6 \times 4 \times 2^n)$
1	2	5	3
2	7	131	38
3	17	6,412	400
4	27	226,330	2,060
5	30	2,360,583	3,911
6	25	6,467,858	2,200
7	14	5,404,183	1,357
8	6	2,341,404	689
9	3	963,413	283
10		350,559	64
11		94,228	14
12		16,238	
13		1,282	

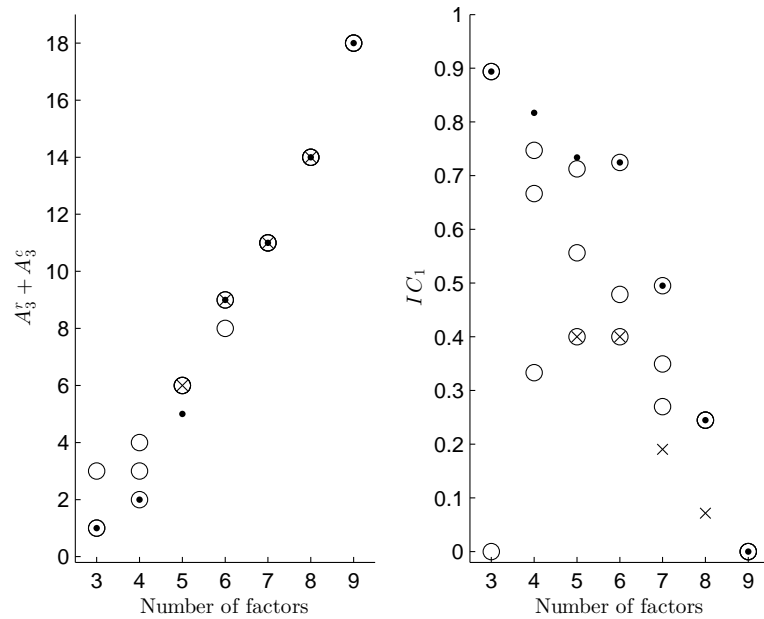


Figure 1: $A_3^r + A_3^c$ and IC_1 values of the optimal arrangements of 16-run two-level designs in four rows and four columns. Circles represent W_2 , W_2^- and W_3 optimal row-column arrangements, while black dots represent arrangements that are optimal in terms of the IC_1 value and crosses correspond to the benchmark designs of Cheng and Mukerjee (2003).

terms of their $A_3^r + A_3^c$ value. The designs with the largest $A_3^r + A_3^c$ values in the left panel of the figure are W_2^- optimal. The four row-column designs identified by Cheng and Mukerjee (2003) exhibit the same total amount of confounding as certain designs that are optimal in terms of at least one combinatorial criterion. More specifically, the five- and six-factor row-column designs of Cheng and Mukerjee (2003) match one of our five- and six-factor optimal designs. However, the seven- and eight-factor row-column designs of Cheng and Mukerjee (2003) are different from the optimal designs we found. For instance, their IC_1 values are smaller than those of our optimal designs. This can be seen in the right panel of Figure 1, which shows that all the optimal designs we obtained outperform the seven- and eight-factor benchmark designs in terms of the IC_1 value. Of all the W_2 , W_2^- and W_3 optimal five- and six-factor designs we identified, those corresponding to the benchmark designs are the poorest in terms of the IC_1 value.

An interesting feature of Figure 1's right panel is the fact that, for five of the seven numbers of factors shown on the horizontal axis, the row-column arrangement that is optimal in terms of the IC_1 value is also optimal in terms of at least one of the combinatorial criteria. In particular, the W_3 optimal arrangements for six and seven factors, the W_2 and W_3 optimal arrangement for three factors and all combinatorially optimal arrangements for eight and nine factors are IC_1 optimal. For these numbers of factors, the black dots appear in the center of the highest circle (for more details, see the supplementary tables). For four and five factors, the designs that are optimal in terms of the IC_1 value are different from the designs that perform best in terms of the combinatorial criteria. For these numbers of factors, the black dots lie higher than the circles in the right panel.

For five factors, the performance of the four optimal designs is discussed in detail in Section 3.8 and summarized in Table 5. The difference in IC_1 value between the row-column arrangement that is optimal in terms of the IC_1 value and the arrangement that is optimal in terms of the W_3 criterion, which corresponds to the highest circle, is about 0.02. The row-column arrangement corresponding to the optimal IC_1 value exhibits less confounding with the blocking factors than any of the designs that are optimal in terms of a combinatorial criterion. This can be seen in the left panel of Figure 1, where, for five factors, the black dot appears below the circle. For four factors, the largest IC_1 value we encountered is 0.82. Of the W_2 , W_2^- and W_3 optimal row-column arrangements involving four treatment factors, it is again the W_3 optimal arrangement which has the largest IC_1 value. Its IC_1 value equals 0.75, which is substantially lower than the optimal IC_1 value. In summary, except for the case of four factors, 16-run designs that are optimal in terms of the W_3 criterion also performs well in terms of the IC_1 value.

One row-column arrangement of a three-factor two-level treatment design has an IC_1 value of 0.89. So, it allows models with one interaction effect to be estimated efficiently. However, it turns out that it also permits the model including all three two-factor interactions to be estimated quite efficiently. As a matter of fact, the D-efficiency of the row-column arrangement for that model is 0.79. In other words, $IC_3 = 0.79$ for that row-column arrangement. Therefore, this row-column arrangement is suitable for identifying active interaction effects, on top of active main effects. The W_2^- optimal row-column arrangement for three treatment factors does not allow any two-factor interaction to be estimated. Therefore, its IC_1 value is zero. The $A_3^r + A_3^c$ value for that row-column arrangement is maximal and equal to 3, the number of two-factor interactions.

The largest IC_1 values for row-column arrangements involving 4, 5 and 6 treatments factors are 0.82, 0.73 and 0.72, respectively. So, these design options allow models with one interaction to be estimated with a reasonable efficiency. Now, to discriminate between two models with a single interaction, we have to be able to estimate models including two interactions efficiently. The average D-efficiencies for models with two interactions turn out to be 0.63, 0.49 and 0.47 for the IC_1 optimal row-column arrangements with 4, 5 and 6 factors, respectively. As these efficiencies are rather low, model discrimination using these row-column arrangements might be hard. Therefore, the 16-run row-column arrangements for 4–6 factors should only be used for experiments where there is a strong prior belief that at most one interaction will be active.

All row-column arrangements with 7–9 factors have IC_1 values of at most 0.5. Therefore, we only recommend these arrangements for experiments in which no active two-factor interactions are expected. If at least one active interaction is anticipated and 7–9 factors need to be studied, we recommend using one of the 24-run row-column arrangements discussed in the next section.

Finally, for eight factors, there are two IC_1 optimal row-column arrangements. One of them is W_2 and W_2^- optimal, and one of them is W_3 optimal. Both have the same F_3 vector. The former arrangement has a better F_4 vector, while the latter arrangement has a better FA_3^{rc22} vector.

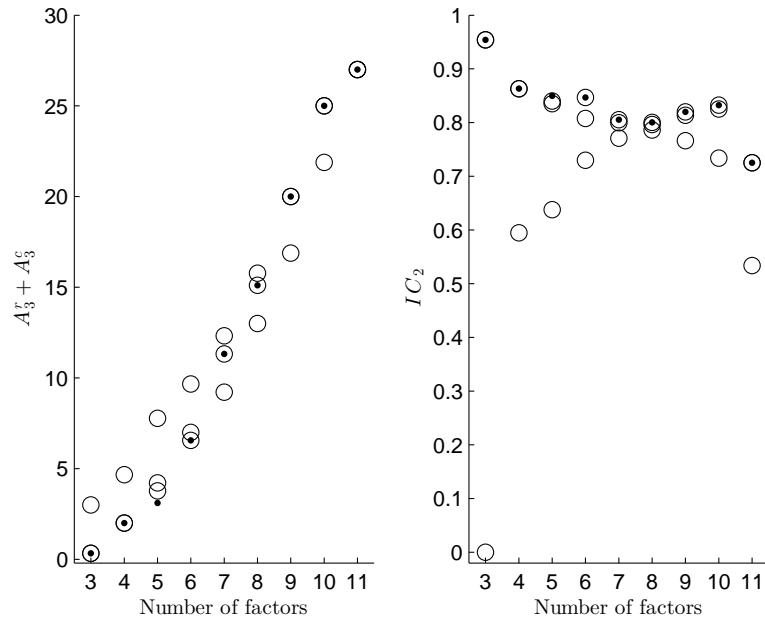


Figure 2: $A_3^r + A_3^c$ and IC_2 values of the optimal arrangements of 24-run two-level designs in six rows and four columns. Circles represent W_2 , W_2^- and W_3 optimal row-column arrangements, while black dots represent arrangements that are optimal in terms of the IC_2 value.

4.2 Designs with 24 runs

Row-column arrangements for 24-run designs can accommodate up to 11 treatment factors in six rows and four columns or up to 13 treatment factors in four rows and three columns. **Orthogonal row-column arrangements involving 24 runs, eight rows and three columns do not exist, because there are no orthogonal arrays involving 24 runs, one eight-level factor, one three-level factor and one or more two-level factors (i.e., there are no orthogonal arrays of the type $OA(24; 8 \times 3 \times 2^n)$). Likewise, arrangements of 20-run two-level designs in five rows and four columns do not exist because there are no orthogonal arrays of the type $OA(20; 5 \times 4 \times 2^n)$.** If a two-level 24-run row-column design with eight rows and three columns or a two-level 20-run row-column design with five rows and four columns is desired, then, inevitably, the main effects of the treatment factors will be partially confounded with the row factor and with the column factor, respectively, and the design will not be orthogonally blocked. These kinds of row-column design are outside the scope of this paper. The optimal design construction algorithms in statistical software allow the construction of such designs, without there being a guarantee that they are truly optimal.

Since the 24-run row-column arrangements allow larger numbers of interactions to be estimated than their 16-run counterparts, we use IC_2 values here, rather than IC_1 values, to identify arrangements that are optimal in terms of our fourth criterion.

4.2.1 Six rows and four columns

Figure 2 shows the main properties of the optimal 24-run row-column arrangements we identified for one blocking factor with six levels and another blocking factor with four levels, i.e., arrangements involving six rows and four columns. Again, the circles in the figure represent row-column designs that are optimal in terms of at least one of the combinatorial criteria (W_2 , W_2^- or W_3), while the black dots represent the designs that optimize the IC_2 value.

The left panel of Figure 2 shows the total amount of confounding between the two-factor interactions, on the one hand, and the blocking factors, on the other hand, as measured by the sum $A_3^r + A_3^c$. For most numbers of factors, there is a substantial difference in $A_3^r + A_3^c$ value between the designs we identified as being

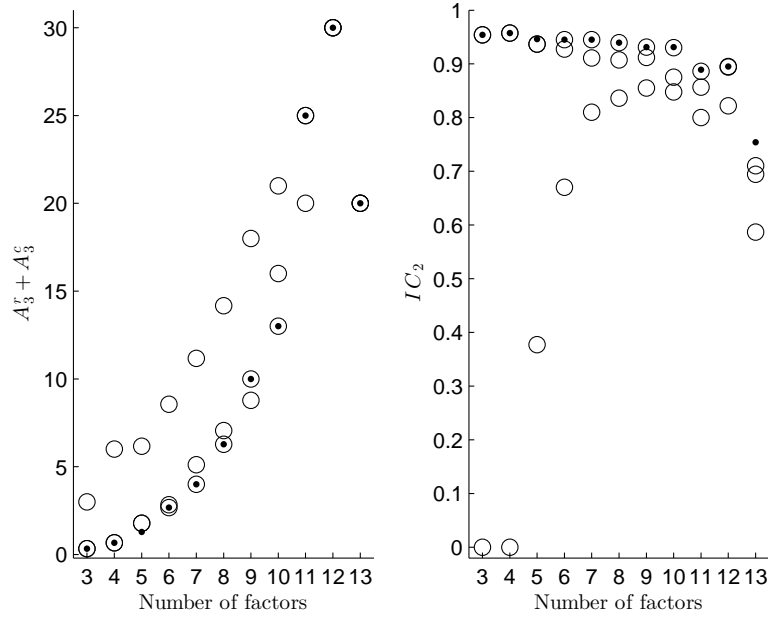


Figure 3: $A_3^r + A_3^c$ and IC_2 values of the optimal arrangements of 24-run two-level designs in four rows and three columns. Circles represent W_2 , W_2^- and W_3 optimal row-column arrangements, while black dots represent arrangements that are optimal in terms of the IC_2 value.

optimal. In each case, the design exhibiting the largest amount of confounding is optimal in terms of the W_2^- criterion. This makes sense, since that criterion maximizes the confounding of the two-factor interactions with the blocking factors, after minimizing the A_3^t value and sequentially minimizing the F_3 vector. **The consequence is that W_2^- optimal arrangements have the smallest IC_2 values of the cases considered here.**

For the 5-factor case, we identified four different optimal arrangements. The IC_2 values for the W_2 , W_2^- , W_3 and IC_2 optimal arrangements, shown in the right panel of Figure 2, are 0.840, 0.638, 0.835 and 0.850, respectively. For numbers of factors other than 5, the W_3 and IC_2 optimal row-column arrangements are identical. The W_2 optimal arrangements differ from the IC_2 optimal ones for the cases with 5–10 factors. For 5 factors, both the W_2 optimal arrangement and the IC_2 optimal arrangement have the same rank R for the interaction model matrix. In this particular case, the IC_2 optimal arrangement has a slightly worse F_3 vector for the treatment design, but that design permits a smaller amount of confounding of the interactions with the blocks. For 6–10 factors, the W_2 optimal arrangements have a larger rank R for the interaction model matrix. Clearly, this does not guarantee a more efficient estimation of models with just a few interactions.

We conclude that, for the scenario involving six rows and four columns, the W_3 criterion leads to the selection of row-column arrangements that perform well in terms of the IC_2 criterion. In contrast, the W_2 criterion, whose minimization vector prefers designs with a large rank for the interaction model matrix, does not tend to select the best row-column arrangement in terms of the IC_2 value.

4.2.2 Four rows and three columns

Figure 3 shows the main properties of the optimal row-column arrangements for 24-run treatment designs when there are four rows and three columns. The properties for the arrangements with four rows and three columns are similar to those of the arrangements involving six rows and four columns. Again, for almost all numbers of factors, the W_2^- optimal arrangements exhibit considerably more confounding between the two-factor interactions and the blocking factors than the arrangements that are optimal in terms of the other criteria. For the cases involving three and four factors, the IC_2 values are zero and the $A_3^r + A_3^c$ values are maximal for the W_2^- optimal row-column arrangements, indicating that these row-column arrangements

do not allow any model with two two-factor interactions to be estimated. The W_2^- optimal arrangements generally have the smallest values for the IC_2 criterion. For the 10-factor case, however, the smallest IC_2 value corresponds to the W_2 optimal design. This is mainly due to the difference in treatment design. As a matter of fact, the IC_2 value for the treatment design used for the W_2 optimal arrangement is 1.8% smaller than that for the treatment design used for the W_2^- optimal arrangement.

For 3, 4, 6, 7, 8, 9 and 12 factors, using the W_3 criterion leads to a row-column arrangement that is also optimal in terms of the IC_2 criterion. For the 13-factor case, considering only the designs that optimize the combinatorial criteria would lead to a loss of 0.044 in IC_2 value. For 5, 10 and 11 factors, however, the optimal IC_2 value exceeds those of the best combinatorial row-column arrangement by 0.01 or less. So, we observe once more that W_3 optimal designs generally perform well in terms of the IC_2 criterion.

5 Optimal row-column arrangements based on catalogs of strength-3 designs

5.1 Criteria and approach

When using two-level orthogonal arrays of strength 3, the main effects are orthogonal to each other, and they are also orthogonal to the two-factor interaction contrast vectors. As a result, $A_3^t = 0$ and the F_3 vector is a zero vector for all strength-3 designs. For this reason, the A_3^t value and the F_3 vector, which are major components of the combinatorial criteria proposed for the strength-2 cases, are no longer useful to classify strength-3 designs. The components that remain useful for strength-3 designs are the rank R , the $A_3^r + A_3^c$ value, the A_4^t value, the F_4 vector, and the FA_3^{rc22} vector.

The results of Deng and Tang (1999) imply that the only possible absolute values for the J_4 -characteristics in N -run strength-3 two-level orthogonal designs equal $(N - 16j)$, where j is a non-negative integer of at most $N/16$. So, the only possible values for the J_4 -characteristics for strength-3 64-run designs are 64, 48, 32, 16 and 0, while those for strength-3 72-run designs are 72, 56, 40, 24 and 8.

As for the strength-2 two-level designs, our desire was to evaluate complete catalogs of suitable orthogonal arrays to identify row-column arrangements that are optimal in terms of our four criteria. By suitable orthogonal arrays, we mean strength-2 orthogonal arrays including a two-level treatment design of strength 3. However, complete catalogs of such suitable orthogonal arrays of the type $OA(N; r \times c \times 2^n; 2)$ are currently unavailable. This is primarily due to the fact that the number of non-isomorphic orthogonal arrays of strength 2 with more than 32 runs is too large to construct complete catalogs. For this reason, we restrict our attention to strength-3 two-level treatment designs that are embedded in a strength-3 orthogonal array of the type $OA(N; r \times c \times 2^n; 3)$. Due to this restriction, the $A_3^r + A_3^c$ value is zero for all row-column arrangements we consider here, and the FA_3^{rc22} vector is a zero vector. As a result, the only remaining relevant components of our combinatorial criteria are the rank R , the A_4^t value and the F_4 vector. Rather than seeking row-column arrangements that optimize the (simplified) W_2 , W_2^- and W_3 criteria, we identify Pareto optimal row-column arrangements based on the rank R , the A_4^t value and the F_4 vector. A row-column arrangement is called Pareto optimal (or admissible; see Sun et al., 1997) with respect to a set of optimality criteria (such as the rank R , the A_4^t value and the F_4 vector) if there exists no other row-column arrangement that scores strictly better on at least one criterion and at least as well on all other criteria. We identify Pareto optimal row-column arrangements for two scenarios. The first one involves 64 runs, four rows and four columns. The second scenario involves 72 runs, three rows and three columns. We also identify row-column arrangements that maximize the IC_5 value in addition to the Pareto optimal designs based on the rank R , the A_4^t value and the F_4 vector.

When performing experiments with 64 or 72 runs, we can, of course, estimate many interactions. Therefore, we should be interested in IC_q values, where q is substantially larger than 1 or 2. However, evaluating large catalogs of orthogonal arrays in terms of the IC_q value is computationally very cumbersome when q is large. As a compromise, for each row-column arrangement under consideration, we calculated the IC_5 value based on 10,000 models including all main effects and five randomly selected two-factor interactions, or based on all models including all main effects and five two-factor interactions whenever there are fewer than 10,000 such models.

We start our search from the complete catalogs of strength-3 orthogonal arrays of the types $OA(64; 4 \times$

Table 7: Numbers of non-isomorphic orthogonal arrays of the types $OA(64; 4 \times 4 \times 2^n; 3)$ and $OA(72; 3 \times 3 \times 2^n; 3)$.

n	$OA(64; 4 \times 4 \times 2^n; 3)$	$OA(72; 3 \times 3 \times 2^n)$
2	12	9
3	267	465
4	13,903	380,253
5	104,949	32,565,625
6	175,297	10,620,268
7	151,708	953,145
8	138,825	653,569
9	83,409	207,970
10	35,807	22,833
11	10,030	69
12	2,159	27

Table 8: Properties of the optimal arrangements of 64-run two-level designs in four rows and four columns based on orthogonal arrays of the type $OA(64; 4 \times 4 \times 2^n; 3)$. The entries of the F_4 vector are the frequencies of J_4 -characteristics of 64, 32 and 16. All designs except for design 12.2 (which maximizes the IC_5 value) are Pareto optimal when considering the A_4^t value, the F_4 vector and the rank R of the interaction model matrix.

ID	A_4^t	F_4			R	D	IC_5
6.1	0	0	0	0	22	1	1
7.1	0	0	0	0	29	1	1
8.1	0	0	0	0	37	1	1
9.1	1	0	4	0	46	0.91172	0.99618
9.2	1.5	0	2	16	46	0.86400	0.99437
10.1	2	0	8	0	50	0	0.99509
11.1	14	8	12	48	38	0	0.84702
11.2	14	10	4	48	41	0	0.80533
11.3	14	14	0	0	46	0	0.75520
11.4	15	10	0	80	41	0	0.80405
12.1	23	15	0	128	42	0	0.79566
12.2	23	15	32	0	38	0	0.81047

$4 \times 2^n; 3)$ and $OA(72; 3 \times 3 \times 2^n; 3)$. Table 7 shows the numbers of non-isomorphic designs we considered for identifying Pareto optimal row-column arrangements and for identifying arrangements that maximize the IC_5 value. Since some of the 64-run treatment designs with up to nine factors and some of the 72-run treatment designs with up to eight factors allow the estimation of a model including all main effects and all two-factor interactions, we also identify designs that maximize the D-efficiency for that full interaction model. Supplementary materials to this paper, available at <http://www.asq.org/pub/jqt/>, include all optimal row-column arrangements of strength-3 treatment designs that we identified.

5.2 Designs with 64 runs

Table 8 shows the properties of the optimal row-column arrangements for 6–12 treatment factors involving 64 runs, four rows and four columns. The first column shows the IDs of the designs. The remaining columns contain the A_4^t value, the F_4 vector, the rank R , the D-efficiency of the full interaction model and the IC_5 value. All row-column arrangements listed in the table for 6–11 treatment factors are Pareto optimal when considering the A_4^t value, the F_4 vector and the rank R of the interaction model matrix. They include the arrangements with the overall best D-efficiencies for the full interaction model and with the best IC_5 values. For 12 factors, design 12.1 is the only Pareto optimal design. Design 12.2 is included in Table 8 because it is the row-column arrangement with the largest IC_5 value for 12 factors.

Table 9: Properties of the optimal arrangements in three rows and three columns of 72-run two-level designs based on orthogonal arrays of the type $OA(72; 3 \times 3 \times 2^n; 3)$. The entries of the F_4 vector are the frequencies of J_4 -characteristics of 24 and 8. All designs are Pareto optimal when considering the A_4^t value, the F_4 vector, the rank R of the interaction model matrix, the D -efficiency for the full interaction model and the IC_5 criterion.

ID	A_4^t	F_4		R	D	IC_5
6.1	0.19	0	15	22	0.97330	0.99518
7.1	0.43	0	35	29	0.95034	0.99475
8.1	0.86	0	70	37	0.91226	0.99460
9.1	4.62	31	95	41	0	0.98304
10.1	7.93	54	156	42	0	0.98269
11.1	12.96	90	240	43	0	0.98231
12.1	19.44	135	360	44	0	0.98266

For six, seven and eight factors, there is only one Pareto optimal design. In each of these cases, the treatment part of the design happens to be a strength-4 design, since $A_4^t = 0$ and the F_4 vector is a zero vector. Therefore, for 6–8 factors, the row-column design we identified allows us to estimate the full interaction model with maximum D-efficiency. When fewer than six factors are required, we recommend dropping factors from the Pareto optimal 6-factor design 6.1. The resulting design will also be Pareto optimal, because the main effects and two-factor interactions of the treatment design are all orthogonal to each other. Therefore, for any row-column arrangement obtained by dropping some of the factors, the rank of the model matrix will be maximal, the A_4^t value will be zero, the F_4 vector will be a zero vector, and the D-efficiency as well as the IC_5 value will be 1.

For nine two-level factors, Table 8 shows two Pareto optimal designs. Design 9.2 is Pareto optimal because it has a slightly better F_4 vector than design 9.1. We prefer design 9.1 over design 9.2, however, because its D-efficiency for the full interaction model is about 5% higher.

For ten factors, there is again a single Pareto optimal row-column arrangement. That design does not allow the full interaction model to be estimated: the rank of the model matrix for the model including the intercept, all main effects and all two-factor interactions is 50 rather than the 56 which is required for the estimability of the full model. Consequently, the D-efficiency for the full interaction model is zero. On the positive side, the large IC_5 value suggests that it should be possible to estimate more than five interactions efficiently at the same time.

For 11 two-level treatment factors, the table shows four designs. The first three of these minimize the A_4^t value. Design 11.1 has the best F_4 vector, and design 11.3 has the largest rank R . Designs 11.2 and 11.4 are compromise designs in terms of the F_4 vector and the rank criterion. We recommend design 11.1 because it has the largest IC_5 value, along with the best F_4 vector.

The case with 12 two-level treatment factors is the only one where the set of Pareto optimal row-column arrangements, constructed by considering the A_4^t value, the F_4 vector and the rank R , did not contain the best row-column arrangement in terms of the IC_5 value. The arrangement with the largest IC_5 value is labeled design 12.2 in Table 8. The difference in IC_5 value between the Pareto optimal design 12.1 and design 12.2 is, however, as small as 0.01481.

5.3 Designs with 72 runs

Table 9 shows the properties of the 72-run row-column arrangements with three rows, three columns and 6–12 factors that are Pareto optimal when considering the A_4^t value, the F_4 vector and the rank R of the interaction model matrix. The structure of the table is similar to that of Table 8. Remarkably, there is just a single Pareto optimal design for each number of treatment factors. In addition, the designs are optimal in terms of the IC_5 value and, when the number of factors is smaller than 9, in terms of the D-efficiency for the full interaction model. For this reason, the designs involving 6–12 treatment factors are also Pareto optimal when we include the D -efficiency and the IC_5 value in the list of criteria to establish Pareto optimality.

The 72-run Pareto optimal row-column arrangements with 6, 7 and 8 treatment factors do allow the full

Table 10: Properties of the two optimal arrangements of 24-run two-level designs in four rows and three columns. The elements of the FA_3^{r22} vector correspond to A_3^r values of 1, 5/9, 1/3 and 1/9. The elements of the FA_3^{c22} vector correspond to A_3^c values of 2/3, 1/2 and 1/6. The elements of the F_4 vector correspond to J_4 -characteristics of 24, 16 and 8 (and correspond to correlations of 1, 2/3 and 1/3).

ID	FA_3^{r22}			FA_3^{c22}			F_4		IC_2		
12.1	0	18	12	36	0	12	36	0	0	495	0.89522
12.2	3	0	36	27	3	6	42	0	0	495	0.82201

interaction model to be estimated, since the rank R is maximal for these numbers of factors. The D-efficiency for the full interaction model is larger than 0.91, but not equal to 1. The D-efficiency for the full interaction model is zero for nine factors or more, indicating that, for these numbers of factors, that model is no longer estimable when using a design from the catalog of strength-3 arrays of the type $OA(72; 3 \times 3 \times 2^n; 3)$. For nine or more factors, it is nevertheless possible to estimate many interactions efficiently, as witnessed by the large IC_5 values.

6 Car tire experiment revisited

In the introduction, we discussed an experiment concerning the wear of car tires which motivated us to conduct the research described in this article. The experiment involved the 12 two-level treatment factors in Table 1 and the total number of experimental runs was 24. The experiment was conducted on four consecutive days. Every day, six runs were performed by three different drivers. The same set of three drivers was used each day. Clearly, two blocking factors are involved in the experiment. The first, day, has four levels, while the second, driver, has three levels. Consequently, the required design is a row-column design involving four rows and three columns.

The design actually used by the experimenters was a replicated 12-run Plackett-Burman design for 11 factors. The 12th factor was set at one level for the first replicate and at the other level for the second replicate. As the experimenters were unaware of the possibility to create an orthogonal blocking arrangement for a 24-run design, they randomly assigned the 24 runs to the twelve combinations of day and driver. The result is that, in their design, the main effects of the 12 factors were not orthogonal to the blocks. Due to the fact that the 12-run Plackett-Burman design was replicated rather than folded over, the main effects of the 11 treatment factors accommodated by that design were partially aliased with the two-factor interactions. That is not the case for the main effect of the 12th treatment factor.

Table 6 shows that there are 16,238 different orthogonal blocking arrangements with four rows (corresponding to the four levels of the first blocking factor, day), three columns (corresponding to the three levels of the second blocking factor, driver) and 12 two-level treatment factors involving 24 runs. When optimizing our three combinatorial criteria and the IC_2 criterion for this type of row-column arrangement, we identified two different optimal row-column arrangements. Their properties are quantified in Table 10 and visualized in Figure 3. The first arrangement, labeled 12.1, is W_2 and W_3 optimal, and has the largest IC_2 value. The second one, labeled 12.2, is W_2^- optimal. For both row-column arrangements, the F_3 vector is zero, indicating that the main effects of the 12 treatment factors are orthogonal to the two-factor interactions and that the design for the 12 treatment factors has strength 3. Each J_4 -characteristic equals 8 for both designs, which means that the contrast vectors for every pair of two-factor interactions exhibit a correlation of $8/24 = 1/3$. The treatment design in both optimal row-column arrangements turns out to be a folded-over (rather than a replicated) 12-run Plackett-Burman design for the first 11 factors. The 12th factor is set at one level for the runs of the original Plackett-Burman design, and at the other level for the negative of that design. The only difference between the row-column arrangements 12.1 and 12.2 is in the assignment of the 24 treatment combinations to the rows and columns, causing the FA_3^{r22} and FA_3^{c22} vectors and the IC_2 values to differ.

As pointed out above, the treatment design actually used for the experiment involved two replicates of a 12-run Plackett-Burman design for 11 factors. This implies that, for these 11 factors, none of the models with all 11 main effects plus any additional interaction is estimable. So, if these factors are considered, $IC_k = 0$ for $k > 0$. Treatment factor 12 defines the two replicates of the Plackett-Burman design. This

implies that the main effect of that factor and all interactions between that factor and the initial 11 factors are orthogonal to each other. Therefore, in the absence of blocking, $\binom{11}{2} = 55$ models involving all main effects and two two-factor interactions are estimable with a D-efficiency of 1. Since there are 66 two-factor interactions for the total of 12 factors, the IC_2 value of the 12-factor treatment design in the absence of blocking is $55/\binom{66}{2} = 0.02564$. Even though this value is too optimistic because it ignores blocking, it is already much poorer than the IC_2 values of the two row-column arrangements in Table 10. To correct the IC_2 value of 0.02564 for rows and columns, we need to find a row-column arrangement of the treatment design actually used. As it turns out that no orthogonal row-column arrangement exists for that design, we resorted to optimal design software to obtain a D-optimal row-column arrangement. The IC_2 value of this row-column arrangement is merely 0.02021.

Since the researchers who performed the car tire experiment were interested in quantifying interactions, row-column arrangement 12.1 is the most attractive one. This is because it has the largest IC_2 value, and the most attractive FA_3^{r22} and FA_3^{c22} vectors. The row-column arrangement 12.1, which we would have recommended for the car tire experiment, is shown in Table 2 in the introduction to this paper. **The original treatment design can not be recommended, except when the main interest of the researchers would be in the interaction effects of the 12th factor.**

Figure 4 shows a color map of all the correlations between the contrast vectors for the rows, the columns, the main effects of the 12 treatment factors and the second-order interactions of the treatment factors. The white in the parts of the color map corresponding to the main effects shows three things. First, it shows that the main effects are orthogonal to the rows and to the columns. Second, the main effects are orthogonal to each other. These two things are due to the fact that the entire row-column arrangement forms an orthogonal array with a four-level factor for the rows, a three-level factor for the columns and 12 two-level factors for the treatments. Third, the main effects are orthogonal to the two-factor interactions. This is due to the fact that the treatment design (i.e., the design for the 12 two-level factors) is a strength-3 orthogonal array (and so, is a resolution-IV design). The white cells in the bottom right part of the color map show that the contrast vector of any interaction of a factor i is orthogonal to the contrast vector of any other interaction of that factor i . The gray cells in the bottom right part of the color map show that the contrast vectors for any other pair of distinct interactions involving four different factors have correlations of $\pm 1/3$. This is in line with the fact that the only nonzero entry of the F_4 vector corresponds to J_4 -characteristics of 8. Finally, the different shades of gray in the top right and bottom left parts of the color map show that the two-factor interaction contrast vectors are not orthogonal to the rows and the columns. This is in line with the nonzero entries for the various A_3^r and A_3^c values in the FA_3^{r22} and FA_3^{c22} vectors, which quantify the confounding between the rows and the columns, on the one hand, and the two-factor interactions, on the other hand.

7 Discussion

In this paper, we proposed general criteria for selecting row-column arrangements of regular and nonregular two-level orthogonal treatment designs, to cope with experimental scenarios that involve two crossed blocking factors. We demonstrated the usefulness of the criteria by identifying attractive row-column arrangements from complete catalogs of non-isomorphic orthogonal arrays with n two-level columns for the treatment factors, an r -level column for the first blocking factor and a c -level column for the second blocking factor. In our search, we distinguished between strength-2 and strength-3 treatment designs, because they differ substantially in the aliasing between the main effects and the two-factor interactions.

Our three combinatorial criteria are based on vectors describing the combinatorial properties of the row-column arrangements, along with the rank of the model matrix corresponding to the full interaction model. We compared the row-column arrangements that are optimal in terms of the combinatorial criteria with those obtained by maximizing the computationally intensive IC_q criterion. We did not find major differences between the row-column arrangements that optimize the combinatorial W_2 and W_3 criteria and those that optimize the IC_q criterion. This suggests that the computationally cheap W_2 and W_3 criteria are good surrogates for the computationally expensive IC_q criterion.

While the W_2 and W_3 criteria are good surrogates for the IC_q criterion, we did identify the row-column arrangements with the best IC_q values for the 16-run and 24-run cases we studied. If the desire is to detect two-factor interactions on top of the main effects, we recommend the designs that have the best IC_q values. If the emphasis is on detecting significant factors by just fitting a main-effects model, we recommend the W_2^-

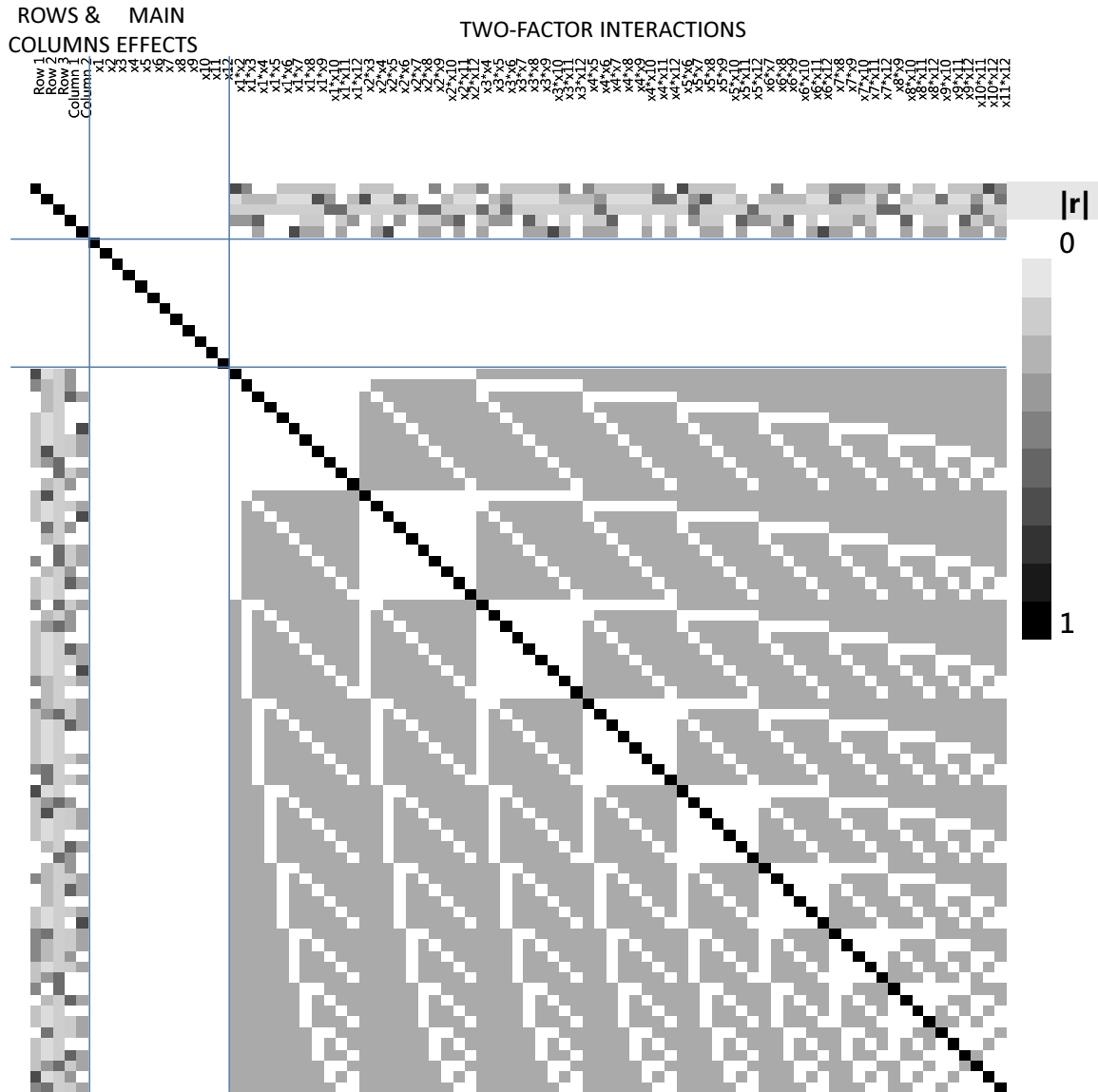


Figure 4: Color map showing the correlations between the contrast vectors for the two blocking factors (the rows and the columns), the main effects of the 12 treatment factors and the second-order interactions of the treatment factors for the row-column arrangement in Table 2 for the car tire experiment.

optimal row-column arrangements, as the mean squared error of a main-effects model based on these designs is less likely to be inflated by active two-factor interactions than that obtained from alternative row-column arrangements.

The approach we take in this paper requires the availability of complete catalogs of orthogonal arrays. Inspecting the website of Eendebak (2016) suggests that the row-column arrangements we studied here are the most practically relevant ones that can be obtained from the currently available catalogs. However, row-column arrangements of two-level strength-3 treatment designs need not be embedded in a strength-3 design to obtain a sensible row-column design. Our search of the complete series of 24-run orthogonal arrays revealed several two-level strength-3 designs embedded in a strength-2 design with an r -level and a c -level factor for the two crossed blocking factors. For example, the recommended row-column arrangement for the car tire experiment in Table 2 involves a strength-3 two-level treatment design, but the entire design, including the treatment factors and the two blocking factors, has a strength of 2 only. Unfortunately, from 32 runs onward, which is the next run size for which two-level strength-3 designs exist, no complete catalogs of suitable strength-2 orthogonal arrays, involving a strength-3 treatment design, are available. To handle this problem when there is a single blocking factor, Sartono et al. (2015) and Vo-Thanh et al. (2018) developed a mixed integer linear programming approach that results in orthogonal blocking arrangements that minimize the confounding between the second-order interaction effects of the treatment factors and the blocking factor. It would be useful to extend this approach to find optimal row-column arrangements for strength-3 two-level treatment designs. A similar approach could be used to arrange any given strength-2 design in rows and columns.

One may wonder whether nonorthogonal blocking of a given orthogonal array would result in better IC_q values. To investigate nonorthogonal blocking in cases with a single blocking factor, Sartono et al. (2015) compared their orthogonal blocking approach based on integer linear programming with an approach using an algorithm for D-optimal blocking. For models with four interaction contrasts, 82% of the orthogonal arrays had a better average efficiency when arranged in blocks using the integer linear programming approach than when using the optimal blocking algorithm. The average difference in D-efficiency was 1.8%. This is caused by the fact that the D-optimal blocking approach used by Sartono et al. (2015) does not penalize the confounding of two-factor interactions with blocks. We would expect this difference in D-efficiency to be larger when there are two blocking factors instead of a single one. On the other hand, optimal design algorithms do not suffer from the restrictions in run size required for orthogonal blocking. For example, it is perfectly feasible to create a 20-run two-level row-column design involving four rows of five runs and five columns of four runs.

Finally, an interesting alternative to both the present approach and the mixed integer linear programming approach would be to develop a procedure that optimizes the IC_q value directly. Designing for average D-efficiency has long been computationally challenging, but recent work by Smucker and Drew (2015) provides a computationally efficient method to construct such designs. Adapting that work to deal with experimental designs involving one or more blocking factors would be an interesting area for future research.

Supplementary materials

Supplementary electronic files, available at <http://www.asq.org/pub/jqt/>, include

Tables.pdf: tables showing the F_3 , F_4 and FA_3 vectors along with the ranks of the interaction model matrices and IC_1 or IC_2 values for the optimal row-column arrangements involving 16 and 24 runs.

Designs.zip: all optimal row-column designs.

Acknowledgements

The research leading to this paper was supported by the Research Foundation – Flanders FWO. The authors are grateful to the referees for their constructive comments.

References

- Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS*. Oxford: Oxford University Press.
- Bailey, R. A. (2008). *Design of comparative experiments*. Cambridge: University Press.
- Box, G. E. P., Hunter, W. G., and Hunter, J. S. (2005). *Statistics for experimenters*. 2nd edition, New York: Wiley.
- Cheng, C. S., Deng, L. J., and Tang, B. (2002). Generalized minimum aberration and design efficiency for nonregular fractional factorial designs. *Statistica Sinica*, 12:991–1000.
- Cheng, C. S. and Mukerjee, R. (2003). On regular-fractional factorial experiments in row-column designs. *Journal of Statistical Planning and Inference*, 114:3–20.
- Cheng, S. W., Li, W., and Ye, K. Q. (2004). Blocked nonregular two-level factorial designs. *Technometrics*, 46:269–279.
- Cheng, S. W. and Wu, C. F. J. (2002). Choice of optimal blocking schemes in two-level and three-level designs. *Technometrics*, 44:269–277.
- Deng, L. Y. and Tang, B. (1999). Generalized resolution and minimum aberration criteria for plackett-burman and other nonregular factorial designs. *Statistica Sinica*, 9:1071–1082.
- Deng, L. Y. and Tang, B. (2002). Design selection and classification for Hadamard matrices using generalized minimum aberration criteria. *Technometrics*, 44:173–184.
- Eendebak, P. T. (2016). Complete series of non-isomorphic orthogonal arrays; <http://pietereendebak.nl/oapage/>.
- Gilmour, S. G. and Trinca, L. A. (2003). Row-column response surface designs. *Journal of Quality Technology*, 35:184–193.
- Goos, P. and Donev, A. N. (2006a). Blocking response surface designs. *Computational Statistics & Data Analysis*, 51:1075–1088.
- Goos, P. and Donev, A. N. (2006b). The D-optimal design of blocked experiments with mixture components. *Journal of Quality Technology*, 38:319–332.
- Goos, P. and Donev, A. N. (2007). D-optimal minimum support mixture designs in blocks. *Metrika*, 65:53–68.
- Li, W. and Nachtsheim, C. J. (2000). Model-robust factorial designs. *Technometrics*, 42:345–352.
- Sartono, B., Schoen, E. D., and Goos, P. (2015). Blocking orthogonal designs with mixed integer linear programming. *Technometrics*, 57:428–439.
- 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., Sartono, B., and Goos, P. (2013). Optimal blocking for general resolution-3 designs. *Journal of Quality Technology*, 45:166–187.
- Schoen, E. D., Vo-Thanh, N., and Goos, P. (2018). Orthogonal blocking arrangements for 24-run and 28-run two-level designs. *Journal of Quality Technology*. To appear.
- Smucker, B. J. and Drew, N. M. (2015). Approximate model spaces for model-robust experiment design. *Technometrics*, 57:54–63.
- Sun, D. X. (1993). Estimation capacity and related topics in experimental design. PhD dissertation, University of Waterloo, Department of Statistics and Actuarial Science, Waterloo ON, Canada.

- Sun, D. X., Wu, C. F. J., and Chen, Y. (1997). Optimal blocking schemes for 2^n and 2^{n-p} designs. *Technometrics*, 39:298–307.
- Tang, B. and Deng, L. Y. (1999). Minimum G_2 -aberration for nonregular fractional factorial designs. *Annals of Statistics*, 27:1914–1926.
- Vo-Thanh, N., Jans, R., Schoen, E. D., and Goos, P. (2018). Symmetry breaking in mixed integer linear programming formulations for blocking two-level orthogonal experimental designs. *Computers & Operations Research*, 97:96–110.
- Xu, H. and Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, 29:549–560.