DEPARTMENT OF ENVIRONMENT,
TECHNOLOGY AND TECHNOLOGY MANAGEMENT

Two-Level Designs of Strength 3 and up to 48 Runs

Eric D. Schoen & Robert W. Mee

UNIVERSITY OF ANTWERP
Faculty of Applied Economics

EARNED EXCELLENCE
AACSB ACCREDITED
EST. 1916
THE BEST BUSINESS SCHOOLS IN THE WORLD

Stadscampus
Prinsstraat 13, B.226
BE-2000 Antwerpen
Tel. +32 (0)3 265 40 32
Fax +32 (0)3 265 47 99
http://www.ua.ac.be/tew

FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ENVIRONMENT,
TECHNOLOGY AND TECHNOLOGY MANAGEMENT

Two-Level Designs of Strength 3 and up to 48 Runs

Eric D. Schoen & Robert W. Mee

RESEARCH PAPER 2012-005
FEBRUARY 2012

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium
Research Administration – room B.226
phone: (32) 3 265 40 32
fax: (32) 3 265 47 99
e-mail: joeri.nys@ua.ac.be

D/2012/1169/005

Two-Level Designs of Strength 3 and up to 48 Runs

Eric D. Schoen[1], Robert W. Mee[2]

[1]TNO, Zeist, Netherlands, and University of Antwerp, Belgium

[2]University of Tennessee, Knoxville TN, USA

June 17, 2011

Abstract

This article will help practitioners select strength-3 designs that are useful for screening both main effects and two-factor interactions. We calculated word-length patterns, correlations of four-factor interaction contrast vectors with the intercept, and ranks of the two-factor interaction matrices for all nonequivalent two-level orthogonal arrays of strength 3 and run sizes up to 48. Based on these characteristics, there are a limited number of designs that can be recommended for practical use.

KEY WORDS: orthogonal array; resolution IV; nonregular designs

1 Introduction

A type of problem regularly encountered in industrial research is the screening of factors to determine which ones are influential. Such a screening can be carried out fruitfully using an experimental design with all the factors included at two levels. Since Plackett and Burman (1946), strength-2 orthogonal arrays based on Hadamard matrices have been popular design options. Strength-2 arrays are such that each of the four level combinations of any set of two factors occur equally often (Rao, 1947). Therefore, the main effect estimators are orthogonal to each other.

The designs based on Hadamard matrices can accommodate up to $N - 1$ factors in N experiments, where N is a multiple of 4. They were initially labeled with a warning discouraging their use if two-factor interactions were expected. However, there are strength-2 arrays that do not completely confound main effects with two-factor interactions. Hamada and Wu (1992), Lin and Draper (1993), Wang and Wu (1995), and Cheng (1995) encouraged attempts to estimate one or two interactions from such arrays, provided that there are only a few main effects active.

This paper focuses on the situation where many of the main effects could be active, while there could also be interactions. We call factor screening for this situation *intensive screening*. An obvious design to consider in such a situation is a strength-4 array. Such an array has all estimators of main effects and two-factor interactions orthogonal to each other. Strength-4 arrays exist for run sizes equalling a multiple of 16. However, the run-sizes 16, 32, and 48 can accommodate only 5, 6, and 5 factors, respectively. Larger run-sizes are usually considered as inappropriate for factor screening. As there are often more than 5 factors included in a study, it is natural to turn to strength-3 arrays with $N \leq 48$ for intensive screening with 6 or more factors. In such arrays, the main effect estimators are orthogonal to each other and to the estimators of all the two-factor interactions. However, at least some of the two-factor interactions are fully or partially aliased with each other.

Cheng (1998) and Miller and Sitter (2001) considered strength-3 arrays constructed by folding over strength-2 arrays. A fold-over array consists of all the treatment combinations in the original array, and all their mirror images, where a mirror image is obtained by switching all the factor levels of a treatment combination. It is well known that folding over strength-2 arrays results in arrays of strength 3. The aforementioned authors showed that nonregular fold-over arrays permit estimation of many models including all the main effects and a good number of interactions. This has stimulated interest in the construction of strength-3 arrays by means other than folding over. Cheng et al. (2008) provided two ways of constructing strength-3 orthogonal arrays for which all the degrees of freedom can be used to estimate main effects and two-factor interactions. More recently, Schoen et al. (2010) completely enumerated all nonequivalent strength-3 designs of up to 48 runs. The numbers of such designs are summarized in Table 1.

A recent case requiring a two-level design involved the diamond turning of aluminium mirrors. In the diamond turning process, a diamond tool, mounted on a machine, cuts a rough workpiece in order to smooth its surface to get optical quality. The goal of the experiment, which was conducted at TNO

Table 1: Numbers of nonequivalent two-level designs of strength 3, 8 − 48 runs, and 4 − 24 factors

No of	Run size					
factors	8	16	24	32	40	48
4	1	2	2	3	3	4
5		2	1	5	3	10
6		1	2	10	9	45
7		1	1	17	25	397
8		1	1	33	105	8383
9			1	34	213	166081
10			1	32	353	1310006
11			1	22	260	3528089
12			1	23	235	4460865
13				12	132	3980095
14				10	96	3139653
15				5	36	2165144
16				5	26	1288460
17					7	629705
18					6	259346
19					3	84495
20					3	24012
21						4919
22						1129
23						130
24						60

Science and Industry, Delft, the Netherlands, was to identify factors that affect the smoothness of mirrors produced under various conditions. The conditions were defined by the settings of 13 controllable factors, as shown in Table 2. Besides the factors machine (**1**) and operator (**2**), there are four tool factors (**3 - 6**), two factors that define the workpiece (**7**, **8**), and two factors governing the action of the lubricant (**9**, **10**). The remaining three factors (**11 - 13**) control mechanical conditions of the diamond turning process itself.

The scientists were quite certain that most of the 13 factors would be active. In addition, they anticipated that there could be several active interactions. The identity of these was not known beforehand, however, so that the scientists faced a scenario of intensive screening. There was an initial budget of at most 50 runs due to the diamond tool expense. Because of the anticipated interactions, a strength-4 array would be an ideal choice. However, the smallest strength-4 array for 13 factors has 128 runs (Schoen et al., 2010). So it was natural to look for a suitable strength-3 array. Available options include 12 arrays of 32 runs, 132 arrays of 40 runs, and 3,980,095 arrays of 48 runs (see Table 1). The problem was to select the best arrays for each of the run sizes and to select a final design.

The purpose of this paper is to characterize and recommend the best designs

Table 2: Factors in diamond turning of mirrors

	Factor	Settings	
1	Machine	350	700A
2	Operator	B	P
3	Rake angle	neg	0
4	Rake face orientation [deg]	110	100
5	Rake nose radius [μm]	0.5	1.5
6	Rake sharpness	fresh	blunt
7	Workpiece material	RSA6061	RSA905
8	Workpiece shape	plano	sphere
9	Lubrication amount	low	high
10	Lubrication pressure	low	high
11	Feed rate	low	high
12	Depth of cut [μm]	low	high
13	Spindle speed [rpm]	low	high

for intensive screening with run-sizes 8, 16, 24, 32, 40, and 48. The rest of this paper is organized as follows. In Section 2, we discuss our criteria to classify the two-level strength-3 designs. We present the best designs for each of the run sizes in Section 3. Finally, in Section 4, we return to the diamond turning example and discuss some additional issues.

2 Classification criteria

Two-level designs can be either regular or nonregular. Regular designs exist for run sizes that are powers of 2. Recall that these designs are created by constructing a full 2^k factorial in $k = p - q$ 'basic' factors and then appending q additional factors defined by interaction contrast vectors of the basic factors. For regular strength-3 designs, main effects are not aliased with two-factor interactions, but there is at least one pair of completely aliased two-factor interactions. For a comprehensive treatment of regular designs, see textbooks such as Box et al. (2005) and Montgomery (2009).

Nonregular designs cannot be constructed by equating added factors to interactions of the basic factors. Nonregular strength-3 designs exist for run sizes that are a multiple of 8. So there is much more choice in run-sizes for nonregular designs. Another general advantage of nonregular designs is in the level of confounding. For strength-3 designs, a two-factor interaction contrast vector can be partially confounded with another such vector. Therefore, a model containing both contrasts can still be estimable, where this is not the case if the contrasts would be completely confounded, as in regular designs.

We ranked all strength-3 designs of up to 48 runs according to the available degrees of freedom to estimate two-factor interactions (Cheng et al., 2008), the G-aberration (Deng and Tang, 1999, 2002), and the G_2-aberration (Tang and Deng, 1999). For the best designs, we determined the fold-over status and the projection estimation capacity (PEC). The ranking criteria and additional properties will now be discussed.

2.1 Degrees of freedom for two-factor interactions

The degrees of freedom for two-factor interactions (df(2fi)) is the rank of the $N \times r$ matrix of two-factor interaction contrast vectors, where $r = p(p-1)/2$, and p is the number of factors. Since we utilize strength-3 arrays with the hope of estimating two-factor interactions, one is interested in designs with a large rank of this matrix.

2.2 G-aberration

G-aberration explores the so-called J-characteristics of contrast vectors. Any such vector has elements -1 or +1. The J characteristic is calculated by summing the vector's elements and taking the absolute values of the sum. For two-level strength-3 designs, any three-factor interaction contrast vector is orthogonal to the intercept. Deng and Tang (1999) showed that the J-characteristics for four-factor interaction contrast vectors in two-level strength-3 designs necessarily equal $N - 16k$, where k is a nonnegative integer. For instance, for $N = 32$, the possible values are 32, 16, and 0. A value of 32 would imply that the element-wise product of, say, factors A, B, C, and D either equals +1 for all the elements or -1 for all the elements. This results in three pairs of two-factor interaction contrast vectors with a correlation of +1 or -1 within each pair. A value of 16 for the J-characteristic would imply a correlation of +0.5 or -0.5 among interaction contrast vectors of a pair.

The frequencies of J-characteristics for four-factor interaction contrast vectors equalling $N - 16k$, for increasing values of k, are collected in a vector F_4. For example, there is a 10-factor 32-run design with $F_4(32, 16) = (1, 62)$. So there is 1 four-factor interaction contrast vector with $J = 32$, and there are 62 such vectors with $J = 16$. The remaining 147 four-factor interaction contrast vectors are orthogonal to the intercept.

A minimum G-aberration design sequentially minimizes the entries of the so-called confounding frequency vector $(F_4, F_5, F_6, \ldots, F_p)$ from left to right, where the vector F_i collects the frequencies of J-characteristics of i-factor contrast vectors. So the process of selecting a minimum G-aberration design initially only needs F_4. In many cases, there is a unique design that minimizes the entries of F_4, and one can conclude that this is the minimum G-aberration designs. In fact, for intensive screening in an industrial context, F_4 is the only part of the confounding frequency vector that is of practical interest.

2.3 G_2-aberration

G_2-aberration is based on sums of squared correlations of contrast vectors with the intercept. Take the 10-factor 32-run design with $F_4(32, 16) = (1, 62)$ as an example. There is one four-factor interaction contrast vector completely aliased with the intercept. So the squared correlation of this contrast vector with the intercept is 1. There are 62 contrast vectors with a squared correlation of 0.25. So the sum of the squared correlations is 16.5. This figure is called the generalized word count of length 4.

The generalized word counts of length 4 up to p are collected in a vector $(A_4, A_5, \ldots, A_p)$ called the generalized word-length pattern. The design that sequentially minimizes the elements in this vector is called the minimum G_2-aberration design. For a regular design, the correlations are either 0 or 1, and the generalized word-length pattern is the same as the usual word-length pattern.

To illustrate that both A_4 and F_4 could be useful to classify designs, compare the above 10-factor 32-run design with an alternative design having $A_4 = 10$ and $F_4(32, 16) = (10, 0)$. Clearly, there are only 10 out of 210 four-factor interaction contrast vectors correlated with the intercept, as opposed to 63 in the other design. However, this other design has just a single severe correlation, and 62 correlations that are only 0.5. We conclude that both designs have merit.

2.4 Fold-over status

The fold-over status of a strength-3 design reports whether or not it can be constructed by folding over a strength-2 design. If this is indeed the case, the treatment combinations can be grouped in $N/2$ mirror image pairs. In one such pair, each factor level in one treatment combination is switched to the other level in the second. Using -1/+1 notation to denote the factor levels, it is easy to see that the sign of an interaction contrast vector must be a constant within a pair. This implies that interactions can only be estimated by contrasts between the $N/2$ mirror image pairs, and, consequently, df(2fi) $\leq N/2 - 1$.

We determined the fold-over status for two reasons. First, it gives information on interaction contrast vectors involving more than four factors. For all fold-over designs, i-factor interaction contrast vectors for odd values of i are orthogonal to the intercept, and the corresponding A_i equals 0 (Cheng et al., 2008). All the designs in our tables that are not fold-over designs, have 5-factor interaction contrast vectors that are completely or partially aliased with the intercept.

Our second reason to determine the fold-over status is that fold-over designs can be easily blocked orthogonally to the main effects. The mirror image pairs constitute $N/2$ blocks of size 2. These can be combined to form blocks of larger size if required. While it it not a central theme of this paper, we thought it worthwhile to mention this simple way of blocking.

2.5 Projection Estimation Capacity

We included the PEC as an additional criterion to give the practitioner a rough idea on how well a design can estimate a range of models. Li and Aggarwal (2008) defined PEC as an integer q such that the two-factor interaction model is estimable for every subset of q factors, but not for every subset of $q + 1$ factors. Loeppky et al. (2007) defined a PEC sequence $(k_1, k_2, \ldots, k_p)$, where k_i is the proportion of models with i main effects and their associated two-factor interactions that is estimable. Here, we combine both approaches by augmenting the integer q as in Li and Aggarwal (2008) with the entry k_{q+1} as in Loeppky et al. (2007). Thus, a PEC of 5.9 means that every five factor projection and 90% of the six-factor projections support estimation of the full two-factor interaction model. Note that it will be extremely rare that *all* 15 interactions between the six factors in a particular projection will be active, and it could instead be useful to check how many six-factor projections have at least z out of 15 estimable interactions.

3 Selected designs

We now present the best strength-3 designs of run sizes 8, 16, 24, 32, 40 and 48. For each of the run sizes up to 40, and for the 48-run arrays for 4 to 8 factors and 24 factors, we denote individual arrays by $p.u$, where p is the number of factors, and u is the lexicographic ranking of the array among all the p-factor arrays in the enumeration of Schoen et al. (2010). For the 48-run arrays with 8 factors, the complete set of nonequivalent arrays was split into 10 subsets of about equal size, numbered from 0 up to 9. These subsets were processed separately to obtain all arrays for 9–23 factors. Accordingly, the individual arrays for 9–23 factors are denoted by $p.f - u$, where f numbers the initial subset of 8-factor arrays.

3.1 8–24 runs

All strength-3 designs for 8 or 16 runs are regular. The single strength-3 design for 8 runs is the well-known 2^{4-1} design of resolution IV. The two four-factor designs for 16 runs are the full factorial design and the replicated half fraction. The two five-factor designs for 16 runs are the strength-4 half-fraction and the four-factor half fraction crossed with a 2^1 design. The designs for 16 runs and 6-8 factors are the minimum aberration designs. Generators for all of the designs can be found in, e.g., Mee (2009, Table G.1 and G.2).

All but one of the 24-run designs can be generated by folding over the 12-run Plackett-Burman design and deleting columns arbitrarily. The one exception is the four-factor design that results from three replicates of the resolution-IV 2^{4-1}. This design is not recommended, because it has three pairs of fully aliased two-factor interactions. Instead, we recommend using a full factorial augmented with one replicate of a half-fraction, which permits estimation of all six two-factor interactions, which is the second nonequivalent design for 24 runs, 4 factors and strength 3.

The 7-factor design has two different projections into 6 factors. One of the columns, when omitted, results in a design with a replicated mirror image pair of runs and 10 degrees of freedom for interactions. Deleting any one of the other columns from the 7-factor design results in a design without replicated pairs, and 11 degrees of freedom for interactions. In view of the additional degree of freedom for interactions, this is the recommended design.

3.2 32–48 runs

Detailed results for the best designs with 32, 40, and 48 runs are presented in Tables 3, 4, and 5, respectively. The designs minimize i) G-aberration, or ii) they minimize G_2-aberration, and, subject to this, minimize G-aberration, or iii) they maximize the degrees of freedom for two-factor interactions. If there are multiple designs maximizing these degrees of freedom, we search for designs that minimize G- or G_2-aberration in this group. All recommended designs have been made available online at the website of this journal.

The results presented in the tables are the available degrees of freedom for estimating two-factor interactions, the fold-over status, the generalized word count of length 4, the frequencies of J-characteristics of four-factor interaction contrast vectors F_4, and the PEC.

32 runs

We first discuss the 32-run designs of Table 3. The best 4-factor and 5-factor designs are a replicated full factorial, and a single replicate of a full factorial; they are not given in the online file. The designs 6.10 and 10.20 are regular designs; all other recommended designs are nonregular.

For 11 up to 16 factors, all designs are fold-over designs. Therefore, the rank of the 2fi matrix is always 15. The PEC for these designs is almost 4. So as long as at most four factors are active, the vast majority of models with all six interactions among these factors are estimable.

For a 32-run design, the number of length-4 words producing full aliasing is given by $F_4(32)$. Each word defines three pairs of completely aliased two-factor interactions. For 11 up to 16 factors, the numbers of these words are all considerably smaller than for the corresponding regular minimum aberration designs. The best designs have 3, 6, 10, 14, 21, and 28 full length-4 words, respectively, compared to 25, 38, 55, 77, 105, and 140 for the regular minimum aberration designs.

The designs for 12 − 16 factors are minimum G-aberration designs. All of these designs also have minimum G_2-aberration. For 11 factors we present two designs. Design 11.22 has minimum G_2-aberration and, subject to this, minimum G-aberration. However, the global minimum G-aberration design for 11 factors is 11.20; it has three instead of four full-aliasing words of length 4.

The minimum G-aberration design for 10 factors is 10.32. The minimum G_2-aberration design for this number of factors, 10.20, is regular. The regular design has 21 degrees of freedom available for two-factor interactions. So all

Table 3: Characterization of the recommended 32-run designs

ID	df(2fi)	fold-over	A_4	$F_4(32,16)$	PEC
4.3	6	yes	0	0 0	4
5.5	10	yes	0	0 0	5
6.10	15	yes	0	0 0	6
7.16	20	no	1	0 4	5.86
8.32	21	no	3	0 12	5.75
9.34	22	no	6	0 24	5.67
10.20	21	no	10	10 0	3.95
10.32	15	yes	16.5	1 62	4.00
11.20	15	yes	25.5	3 90	3.99
11.22	15	yes	25	4 84	3.99
12.23	15	yes	38	6 128	3.99
13.10	15	yes	55	10 180	3.99
14.8	15	yes	77	14 252	3.99
15.4	15	yes	105	21 336	3.98
16.4	15	yes	140	28 448	3.98

available degrees of freedom are employed in estimating main effects and two-factor interactions. Cheng et al. (2008) call a design with this property second order saturated.

The minimum G-aberration design for 10 factors has only one length-4 word, but $A_4 = 16.5$ vs. 10 for the minimum aberration regular design. Note that the PEC of the regular design is smaller than the PEC of the non-regular design. There are 210 subsets of four factors. The regular design can estimate the two-factor interaction model for 200 of these, whereas the non-regular design can estimate such a model in 209 of the subsets.

Designs 9.34, 8.32, and 7.16 are minimum G-aberration designs for 9, 8, and 7 factors, respectively. The designs also have minimum G_2-aberration. These three designs were found earlier by Xu (2005) as the minimum aberration projections of an orthogonal array of 13 factors, 32 runs and strength 2. Our work shows that the designs have minimum G-aberration and minimum G_2 aberration among all 32-run two-level designs with 9, 8, or 7 factors. The worst correlation in a pair of two-factor interaction contrast vectors is 0.5. The PEC values of the three designs are well above 5.5. The 9-factor design is second order saturated, with 22 degrees of freedom for two-factor interactions. The designs for 8 and 7 factors have 21 and 20 degrees of freedom for two-factor interactions.

Finally, the tabulated 32-run designs include those with maximum PEC.

40 runs

We now turn to the recommended 40-run designs given in Table 4, which have no full-aliasing length-4 words, i.e., $F_4(40) = 0$. The recommended designs for 11 up to 20 factors are all fold-over designs. Butler (2004, 2007) showed that the maximum number of factors that could result in a design not consisting of fold-over pairs is strictly smaller than $N/3$. For the 40-run series, $N/3 = 13.33$. Our work shows that, for 40 runs, the maximum is 10.

When compared to the designs with 11 or more factors, those with 10 or fewer factors show a marked increase both in PEC and in df(2fi). The best arrays with 4 up to 6 factors permit estimation of all the 2fi. These designs are fold-over designs.

For 7 up to 10 factors, the arrays with the maximum number of estimable two-factor interactions are not fold-over designs. For fold-over designs, this maximum is 19. The best arrays for 7–10 factors have a df(2fi) of 21, 25, and 27, respectively. The 7-factor design permits all interactions to be estimated. (The other tabulated 7-factor design is a fold-over design that minimizes G- and G_2-aberration.) The recommended array 8.105 is only 3 degrees of freedom short of accommodating all two-factor interactions. Assuming that higher-order interactions can be ignored, there are 6 degrees of freedom available to estimate the error variance. Arrays 9.213 and 10.353 have 3 and 2 degrees of freedom, respectively, for estimating this variance.

The designs recommended for 4–11 factors and for 17–20 factors include those with maximum PEC. For 12–16 factors, the best PEC is at most 4%

Table 4: Characterization of the recommended 40-run designs

ID	df(2fi)	fold-over	A_4	$F_4(24, 8)$*	PEC
4.3	6	yes	0.04	0 1	4
5.3	10	yes	0.2	0 5	5
6.9	15	yes	0.6	0 15	6
7.24	21	no	1.4	0 35	7
7.25	19	yes	1.4	0 35	6
8.105	25	no	2.8	0 70	7
9.213	27	no	5.04	0 126	7
10.353	27	no	8.4	0 210	7
11.68	19	yes	18.96	18 312	5.83
12.157	19	yes	28.44	27 468	5.77
13.55	19	yes	41.72	41 674	5.77
14.46	19	yes	58.6	58 943	5.77
15.23	19	yes	80.2	80 1285	5.76
16.4	19	yes	107.04	107 1713	5.76
17.3	19	yes	140	140 2240	5.77
18.3	19	yes	180	180 2880	5.77
19.1	19	yes	228	228 3648	5.76
20.1	19	yes	285	285 4560	5.76

* $F_4(40) = 0$ for all designs

higher than in the tabulated designs, while the best PEC never reaches 7. We do not present explicit maximum-PEC designs, because we would then over-emphasize the importance of this criterion.

48 runs

We now turn to the recommended 48-run designs given in Table 5. For 14–19 factors, there are so many designs with a minimum value of A_4 that it was computationally infeasible to check on minimum G_2-aberration. For these cases,

Table 5: Characterization of the recommended 48-run designs

ID	df(2fi)	fold-over	A_4	$F_4(48, 32, 16)$	PEC
4.4	6	-	0	0 0 0	4
5.10	10	no	0	0 0 0	5
6.44	15	no	0.11	0 0 1	6
7.397	21	no	0.33	0 0 3	7
8.6609	28	no	1.44	0 0 13	8
8.8382	27	no	1	0 0 9	7
9.4-16240	29	no	2.44	0 0 22	5.95
9.6-15282	36	no	3.44	0 0 31	9
10.0-136170	35	no	6	0 0 54	5.98
10.0-136872	31	no	5.33	0 0 48	5.95
11.0-377572	32	no	9.11	0 0 82	5.94
11.3-351294	34	no	10.89	0 0 98	5.97
11.5-32061	34	no	9.56	1 0 77	(5.939)
12.0-541920	33	no	15.33	0 0 138	5.92
12.5-76810	34	no	15	1 0 126	(5.939)
13.0-594498	34	no	23	0 0 207	5.91
14.2-70173	33	no	34.33	1 0 300	(5.895)
14.4-221684	23	yes	46.78	0 0 421	5.91
15.3-42200	23	yes	64.11	0 7 549	5.95
15.4-136404	23	yes	64.33	0 0 579	5.91
16.3-6427	23	yes	85.67	0 17 703	5.95
16.4-78065	23	yes	86	0 0 774	5.91
17.3-885	23	yes	112.67	0 24 918	5.95
17.4-31172	23	yes	112.88	0 0 1016	5.90
18.4-7024	23	yes	145.33	0 0 1308	5.90
19.4-659	23	yes	184.44	0 0 1660	5.90
20.4-43	23	yes	230.67	0 0 2076	5.90
21.4-2	23	yes	285	0 0 2565	5.90
22.4-1	23	yes	348.33	0 0 3135	5.90
23.4-1	23	yes	421.67	0 0 3795	5.90
24.60	23	yes	506	0 0 4554	5.90

we searched for designs with minimum A_4 and, subject to this, minimum G-aberration.

The designs 14.2-70173, 12.5-76810, and 11.5-32061 in Table 5 have one full-aliasing length-4 word. Strictly speaking, their PEC equals 3.999, 3.998, and 3.997, respectively, since the two-factor interaction model is not estimable in one set of four factors. However, these designs allow estimation of the two-factor interaction model in nearly 100% of subsets of four or five factors. In Table 5 we display their PEC as (5.895), (5.939), and (5.939), using parentheses to indicate that nearly all of the four and five factor models are estimable. We believe this is more indicative of the designs' estimation capacity than the strict PEC.

For 4–9 factors, the tabulated designs include those with maximum PEC. For each factor number from 10 onward, the tabulated designs include a case with PEC ≥ 5.9; generally, the maximum PEC designs can estimate between 1% and 4% more of the six-factor all-two-factor-interaction models than these Table 5 designs.

Regular strength-3 fractions that are not fold-over designs exist for $p = 5N/16$ factors; for an example, see design 10.20 in Table 3. The 48-run series are almost exclusively nonregular, however. Our work on this series shows that the maximum number of factors in the designs that are not fold-over is 14. The tabulated minimum G_2 aberration design is of this type, while the minimum G-aberration design for 14 factors is a fold-over design. Note the extent to which the first design is superior over the second one: the fold-over design has df(2fi)=23, while the other design has df(2fi)=33.

For 5–13 factors, the minimum G-aberration designs and the minimum G_2-aberration designs are not fold-over designs. For 8–11 factors, the designs that maximize the degrees of freedom for two-factor interactions do not minimize G- or G_2-aberration.

Finally, for the minimum G_2-aberration designs with 13, 14, and 24 factors, all available degrees of freedom are employed in estimating main effects and two-factor interactions. All other designs have some degrees of freedom left for estimating the error variance.

4 Discussion

In this paper, we studied two-level designs that could be useful for intensive factor screening. We searched through the complete catalog of strength-3 designs for up to 48 runs to find the best designs in terms of G-aberration, G_2-aberration, and degrees of freedom for estimating two-factor interactions. The best designs were presented in Tables 3, 4, and 5. To illustrate how these tables might be used to find a suitable design, we return to the problem of finding a good design for the 13-factor experiment on diamond turning of mirrors.

There are no 13-factor strength-3 designs with 16 or 24 runs. We give a synopsis of the best 13-factor designs for 32, 40, and 48 runs in the upper panel of Table 6. The three designs presented there show a general increase in desirable features with run size. The best 32-run design has $F_4(32) = 10$. So there are 10 four-factor interaction contrast vectors that are completely aliased with the intercept. This implies that the worst correlation within pairs of two-factor interaction contrast vectors is 1.

For the best 40-run design, there are no four-factor interaction contrast vectors completely aliased with the intercept. There are 41 four-factor interaction contrast vectors with $J = 24$, so the worst correlation between pairs of two-factor interaction contrast vectors is 3/5. Finally, the worst correlation between pairs of two-factor interaction contrast vectors in the best 48-run design is 1/3, because 207 sets of four factors have $J = 16$ and there are no sets with a higher

Table 6: Design options for the diamond turning experiment

N	ID	df(2fi)	A_4	Worst correlation	PEC
32	13.10	15	55	1	3.99
40	13.55	19	41.72	3/5	5.77
48	13.0-594498	34	23	1/3	5.91
48	11.0-377572	32	9.11	1/3	5.94
48	11.5-32061	34	9.56	1	(5.939)
48	11.3-351294	34	10.89	1/3	5.97

value of J.

While the improvement in terms of worst correlations is gradual, the increase in PEC and df(2fi) is not. The PEC rises from almost 4 in the 32-run experiment to 5.77 for 40 runs. This is followed by a less marked increase to 5.91 for 48 runs. However, there is a striking increase in df(2fi) when going from 40 to 48 runs. Indeed, this is the most notable difference between the fold-over designs for 32 and 40 runs and the design for 48 runs, which is not a fold-over. In view of their interest in interactions, the experimenters adopted the 48-run option for their design. This choice proved beneficial in that serveral two-factor interactions were found to be active when the data were analyzed.

It is instructive to consider what would happen if the diamond turning experiment had 11 instead of 13 factors. For 32 runs, Table 3 gives two options that are very similar. Table 4 gives a single best option for a 40-run experiment. However, there are three different options for a 48-run experiment, and we focus our comparison on these three designs; refer to the lower panel of Table 6. The first option, design 11.0-377572, has 32 degrees of freedom available for 2fi. Therefore there is a total of 4 degrees of freedom to estimate error, assuming as is usual that higher-order effects can be ignored. The design minimizes G-aberration and G_2-aberration.

We like to note that minimizing aberration in strength-3 arrays is merely a computationally convenient vehicle for selecting designs that are likely to be useful. As a matter of fact, the options 11.5-32061 and 11.3-351294 do not minimize either type of aberration, but they do have two additional degrees of freedom for interactions. Design 11.5-32061 has a single four-factor interaction contrast vector that is fully confounded with the intercept. Thus it has three pairs of fully aliased two-factor interaction contrast vectors. The design permits estimation of all the main effects and all the 2fi among 6 factors in 93.9 % of the possible sets of 6 out of 11 factors. For option 11.3-351294, this percentage is 97; no four-factor interaction contrast vectors are completely confounded with the intercept.

For 11 factors and 48 runs we would recommend the first option (the minimum G-aberration design) when the likely interactions might involve a majority of the factors, whereas we would recommend the third option (which has higher PEC) when it is likely for the active two-factor interactions to be concentrated in a set of six (or fewer) factors, with most of these factors having active main effects.

The diamond turning experiment illustrates the practical importance of our search through the complete catalog of strength-3 designs for up to 48 runs. We believe that the designs could also be valuable as building blocks for larger designs, along the lines recorded by Cheng et al. (2008).

Acknowledgements

The authors are grateful to two anonymous referees and the associate editor, whose comments significantly increased the readability of the paper, and to Guido Gubbels of TNO Science and Industry, who supervised the diamond turning experiment. EDS was supported by a grant from the Flemish Foundation of Scientific Research FWO.

References

Box, G. E. P., Hunter, W. G., and Hunter, J. S. (2005). *Statistics for experimenters.* 2nd edition, New York: Wiley.

Butler, N. A. (2004). Minimum G_2-aberration properties of two-level foldover designs. *Statistics and Probability Letters*, 67:121–132.

Butler, N. A. (2007). Results for two-level fractional factorial designs of resolution IV or more. *Journal of Statistical Planning and Inference*, 137:317–323.

Cheng, C. S. (1995). Some projection properties of orthogonal arrays. *Annals of Statistics*, 23:1223–1233.

Cheng, C. S. (1998). Some hidden projection properties of orthogonal arrays with strength three. *Biometrika*, 85:491–495.

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

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.

Hamada, M. and Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, 24:130–137.

Li, Y. and Aggarwal, M. L. (2008). Projection estimation capacity of Hadamard designs. *Journal of Statistical Planning and Inference*, 138:154–159.

Lin, D. K. J. and Draper, N. R. (1993). Generating alias relationships for two-level Plackett and Burman designs. *Computational Statistics and Data Analysis*, 15:147–157.

Loeppky, J. L., Sitter, R. R., and Tang, B. (2007). Nonregular designs with desirable projection properties. *Technometrics*, 49:454–467.

Mee, R. W. (2009). *A comprehensive guide to factorial two-level experimentation.* Springer-Verlag, New York, NY, USA.

Miller, A. and Sitter, R. R. (2001). Using the folded-over 12-run Plackett-Burman design to consider interactions. *Technometrics*, 43:44–55.

Montgomery, D. C. (2009). *Design and analysis of experiments.* 7th edition, Wiley, New York.

Plackett, R. L. and Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, 33:305–325.

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.

Tang, B. and Deng, L. Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Annals of Statistics*, 27:1914–1926.

Wang, J. C. and Wu, C. J. F. (1995). A hidden projection property of Plackett-Burman and related designs. *Statistica Sinica*, 5:235–250.

Xu, H. (2005). A catalogue of three-level regular fractional factorial designs. *Metrika*, 62:259–281.