Two-level orthogonal screening designs with 24, 28, 32 and 36 runs

Eric D. Schoen

Faculty of Applied Economics, University of Antwerp, Belgium

TNO, Zeist, Netherlands

Nha Vo-Thanh

Faculty of Applied Economics, University of Antwerp, Belgium

Peter Goos

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

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

November 30, 2016

Abstract

The potential of two-level orthogonal designs to fit models with main effects and two-factor interaction effects is commonly assessed through the correlation between contrast vectors involving these effects. We study the complete catalog of non-isomorphic orthogonal two-level 24-run designs involving 3–23 factors and we identify the best few designs in terms of these correlations. By modifying an existing enumeration algorithm, we identify the best few 28-run designs involving 3–14 factors and the best few 36-run designs in 3–18 factors as well. Based on a complete catalog of 7570 designs with 28 runs and 27 factors, we also seek good 28-run designs with more than 14 factors. Finally, starting from a unique 31-factor design in 32 runs that minimizes the maximum correlation among the contrast vectors for main effects and two-factor interactions, we obtain 32-run designs that have low values for this correlation. To demonstrate the added value of our work, we provide a detailed comparison of our designs to the alternatives available in the literature.

KEY WORDS: G -Aberration; G_2 -Aberration; Hadamard Matrix; Orthogonal Array; Plackett-Burman Design.

1 Introduction

In the early stages of process optimization, experimenters seek controllable factors that really affect the process. The recommended approach to this end is to create a list of candidate factors, vary their settings according to a well-chosen experimental plan, and relate the changes observed in the process to the changes in the factor settings. This procedure is called factor screening. Orthogonal experimental designs in which all factors have two levels are commonly used plans for screening experiments. This is due to the fact that each of the four level combinations of any pair of factors is tested equally often in such plans. Therefore, in a statistical model that includes only main effects, each of these effects is estimated with maximum precision and independently from other main effects. Two-level orthogonal experimental designs are therefore ideal when it comes to detecting main effects of factors, especially when these are the only active effects.

Some two-level orthogonal designs are regular, while others are nonregular. Regular two-level designs (see, e.g., Mee, 2009; Wu and Hamada, 2009) exist for run sizes that are powers of 2. The smallest regular designs have 4, 8, 16 or 32 runs and can accommodate up to 3, 7, 15 and 31 factors, respectively. In these designs, main effects can be completely aliased with two-factor interactions and two-factor interactions can be completely aliased with other two-factor interactions. Therefore, there exists ambiguity when interpreting the experimental results. On the positive side, this ambiguity is readily recognized. On the negative side, completely aliased effects cannot be included in a model simultaneously. Therefore, the number of models that can be fitted to data from regular experimental designs is limited.

Many nonregular designs do not involve completely aliased effects. For this reason, they permit a larger number of statistical models to be fitted to the data. In addition, when interactions are active and a main-effects model is estimated, the estimates of the main effects usually have a smaller bias. The smallest nonregular designs have 12, 16 or 20 runs and can accommodate up to 11, 15 and 19 factors, respectively. These designs have been studied extensively by Lin and Draper (1993), Deng and Tang (2002), and Sun et al. (2008).

In this paper, we study designs with 24, 28, 32 and 36 runs. Compared to designs with 12–20 runs, these designs allow more factors to be studied and offer more precise estimates of the factor effects. Also, for any given number of factors, the aliasing between main effects and two-factor interactions and the aliasing among two-factor interactions are less severe for designs with 24–36 runs than for designs with 12, 16 or 20 runs. For example, for the best 13-factor 20-run designs listed in Deng and Tang (2002), the generalized word count of length 3, which measures the correlation between main effect contrast vectors and two-factor interaction contrast vectors, equals 15.92. For the best 24-run designs, this generalized word count is only 6, indicating a substantially smaller amount of aliasing.

The purpose of the present paper is to identify designs with 24–36 runs with a minimum or a near-minimum degree of aliasing between main effects and two-factor interactions and, subject to this, a minimum or near-minimum degree of aliasing among two-factor interactions. The paper therefore reports designs that perform well in terms of the G -aberration (Deng and Tang, 1999, 2002) and G_2 -aberration (Tang and Deng, 1999) criteria.

In Section 2, we review G - and G_2 -aberration as well as additional criteria that are relevant for comparing the performance of designs in terms of model fitting. In Section 3, we review the literature on designs with 24–36 runs and identify some shortcomings. In Section 4, we study 24-run designs based on the complete enumeration by Schoen et al. (2010). Our list of recommended 24-run designs includes 18 designs that perform slightly better than the best designs available in the literature. Our contributions for the run sizes 28, 32 and 36 are more substantial. We are the first to identify the best 28-run designs involving 3–14 factors and the best 36-run designs involving 3–18 factors in terms of the G -aberration criterion. This required the enumeration of 97,590 different 28-run designs and 245,886,095 different 36-run designs. Based on a complete catalog of 7570 designs with 28 runs and 27 factors, we also identify good 28-run designs with more than 14 factors. In addition, we identify new sets of interesting 32-run designs. For this run size, all minimum G -aberration designs and all minimum G_2 -aberration designs with up to 16 factors are known (Schoen

and Mee, 2012). However, for 8–16 factors, all these designs have correlations of 0.5 or 1 among the two-factor interaction contrast vectors. Based on an enumeration of more than 3.5 million designs, we provide a series of designs with 7–31 factors in which all correlations among the contrast vectors for the main effects and the two-factor interactions are at most 0.25. Our designs with 17–30 factors all have a better G -aberration than the best designs in the literature. The best 32-run design with 31 factors was identified earlier in Mee (2009). Our results for 28-, 32- and 36-run designs are reported in Sections 5, 6 and 7, respectively. In the Section 8, we discuss the strengths and weaknesses of our approach and we summarize our findings.

2 Evaluation criteria

In this paper, we identify designs with 24, 28, 32 and 36 runs with a limited amount of aliasing between the main effects and the two-factor interactions and, subject to this, a limited amount of aliasing among two-factor interactions. To evaluate designs, we consider their strength (Hedayat et al., 1999), their confounding frequency vector (CFV; Deng and Tang, 1999, 2002) and their generalized word-length pattern (GWLP; Tang and Deng, 1999). The generalized resolution and the G -aberration criterion are based on the CFV, while the G_2 -aberration criterion is based on the GWLP. We also evaluate designs in terms of their ability to estimate models involving interactions. To this end, we calculate the rank of the model matrix corresponding to a model with an intercept, all main effects and all two-factor interactions. We call that model the interaction model. Finally, we calculate the fraction of 5-factor subsets with an estimable interaction model, and we compute the average D -efficiency over all these models.

2.1 Strength

All designs considered here are two-symbol orthogonal arrays. These arrays are rectangular arrangements of the symbols, where the rows correspond to the experimental runs, the columns correspond to the factors, and the symbols correspond to the factor levels. An orthogonal array with two symbols and a strength of t is such that, in every set of t columns, each of the 2^t combinations of symbols occurs equally often. A strength of 3 implies that main effects can be estimated independently from two-factor interaction effects, while a strength of 2 implies that main effects can be estimated independently from other main effects, but not independently from two-factor interaction effects. Any 24-run design either has strength 3 or strength 2. The designs with 28 or 36 runs all have strength 2. They cannot have strength 3, because the run size of 28 or 36 is not divisible by $2^3 = 8$. Therefore, it is impossible for the eight level combinations of any three factors to occur equally often. Finally, the 32-run designs we discuss all have strength 2, because all strength-3 designs have been studied by Schoen and Mee (2012).

2.2 Confounding frequency vector and generalized resolution

The CFV is based on so-called J_s -characteristics of s -factor interaction contrast vectors. For a two-level design, any contrast vector has elements -1 or $+1$. The J_s -characteristic of an s -factor interaction contrast vector is the absolute value of the sum of the vector's elements. Deng and Tang (1999) showed that the J_3 -characteristics in two-level strength-2 designs are of the form $N - 8k$, where k is a nonnegative integer, N is the run size, and $k \leq N/8$. So, J_3 -characteristics for 24-run or 32-run designs must be multiples of 8 (including 0), while those for the 28-run and 36-run designs must be odd multiples of 4. A J_3 -characteristic of N in an N -run design means that the main effect contrast vector of each factor involved and the two-factor interaction contrast vector of the remaining two factors are perfectly positively or negatively correlated. In that case, there is complete aliasing between main effects and two-factor interaction effects. A J_3 -characteristic of $N - 8k$ implies correlations among the

corresponding contrast vectors of $\pm(1 - 8k/N)$. Ideally, a J_3 -characteristic is as close to zero as possible. A zero J_3 -characteristic means that the main effect contrast vectors of the factors involved are orthogonal to the two-factor interaction contrast vectors. For 28-run and 36-run designs, a zero J_3 -characteristic for a three-factor interaction contrast vector is impossible, as a result of which at least some aliasing between main effects and two-factor interaction effects is unavoidable. This is not the case for 24-run or 32-run designs.

When J_3 -characteristics are calculated for all possible three-factor interaction contrast vectors, the frequencies of the different outcomes can be collected in a vector F_3 . For example, there is a 14-factor 24-run design with an F_3 vector equal to $F_3(24, 16, 8, 0) = (0, 4, 92, 268)$. So, no three-factor interaction contrast vectors have J_3 -characteristics of 24, four have a J_3 -characteristic of 16, 92 have a J_3 -characteristic of 8 and 268 have a J_3 -characteristic of 0. It is customary to drop the last element of the F_3 vector when it corresponds to a zero J_3 -characteristic, and to use $F_3(24, 16, 8) = (0, 4, 92)$ instead of $F_3(24, 16, 8, 0) = (0, 4, 92, 268)$.

In the same way as with the F_3 vector, frequency vectors $F_4, F_5, \dots, F_n$ of the J_4 -, J_5 -, $\dots$, J_n -characteristics can be calculated, where n is the total number of factors. The concatenation of all these vectors, $(F_3, F_4, F_5, \dots, F_n)$, is the CFV. A minimum G -aberration design sequentially minimizes the entries of the CFV from left to right. Such a design therefore sequentially minimizes the frequencies with which the worst correlations occur. In this paper, to limit the computational burden, we restrict ourselves to the sub-vectors (F_3, F_4) , (F_3, F_4, F_5) and (F_3, F_4, F_5, F_6) . We refer to the corresponding design evaluation criteria as the GA4, GA5 and GA6 classifiers. Note that, if there is a unique best design according to any of these criteria, then that design has minimum G -aberration. Note also that strength-3 designs have a zero F_3 vector.

The generalized resolution of a strength-2 design equals 4 minus the largest absolute correlation between a main effect contrast vector and a two-factor interaction contrast vector. For example, the design mentioned above includes main effect contrast vectors that have a correlation of $\pm 2/3$ with two-factor interaction contrast vectors. Therefore, its generalized

resolution equals $3^{1/3}$. In general, designs of strength t have a generalized resolution of $t+2-S/N$, where S is the largest non-zero J_{t+1} -characteristic. A high generalized resolution is preferable to a small one.

2.3 Generalized word-length pattern

The GWLP is a vector of scalars $A_3, A_4, \dots, A_n$ called generalized word counts of length $3, 4, \dots, n$. These word counts summarize the vectors $F_3, F_4, F_5, \dots, F_n$ that constitute the CFV. Calculating an entry A_i of the GWLP requires the division of all J_i characteristics by N , squaring the results and summing the squares. For example, the 14-factor 24-run design with $F_3(24, 16, 8, 0) = (0, 4, 92, 268)$ has an A_3 value equal to $0 \times (24/24)^2 + 4 \times (\pm 16/24)^2 + 92 \times (\pm 8/24)^2 + 268 \times (0/24)^2 = 12$. The word counts of length 3 are proportional to the sum of the squared correlations between the main effect contrast vectors and the two-factor interaction contrast vectors. Similar observations can be made for word counts of length 4 up to n . A design that minimizes the elements of the GWLP from left to right is a minimum G_2 -aberration design.

There may be many designs with the same generalized word counts. To distinguish between designs with the same A_3 value, we use the F_3 vector, and to distinguish between designs with the same A_4 value, we utilize the F_4 vector. So, we seek designs that sequentially minimize the vector (A_3, F_3, A_4, F_4) , which we refer to as the mixed A4 classifier. The second half of this vector, (A_4, F_4) , was used by Schoen and Mee (2012) to classify strength-3 designs.

Using the mixed A4 classifier instead of the pure GWLP has two advantages. First, it emphasizes the desire to avoid strong correlations between main-effect contrast vectors and two-factor interaction contrast vectors. This is because sequentially minimizing the A_3 value and the elements of the F_3 vector results in the smallest frequencies possible for large J_3 -characteristics among all designs that minimize the A_3 value. A second advantage of using the mixed A4 classifier is that it reduces the number of designs to consider substantially. As a matter of fact, there may be large numbers of designs which minimize the A_3 value, but,

generally, only a few of them possess the most attractive F_3 vectors.

To illustrate the usefulness of the mixed A4 classifier, consider two 16-factor 24-run designs. The best design in terms of the mixed A4 classifier has an A_3 value of 24.44 and $F_3(24, 16, 8) = (0, 4, 204)$. Another design, ranked second in terms of the mixed A4 classifier, has the same A_3 value, but $F_3(24, 16, 8) = (0, 6, 196)$. Therefore, it involves two more J_3 -characteristics of 16 (and thus two more correlations of $\pm 2/3$) than the first design. This feature makes the second design less suitable when the intention is to fit a main-effects model.

2.4 Rank of the interaction model matrix

For a design with n factors, the interaction model matrix is an $N \times q$ matrix with a column for the intercept, n main-effect contrast vector columns and $n(n-1)/2$ two-factor interaction contrast vector columns. Hence, $q = 1 + n(n+1)/2$. The difference between the rank r of the interaction model matrix and $1 + n$ (the number of main effects plus one for the intercept) indicates the number of estimable two-factor interaction effects. Therefore, designs with a large rank r of the interaction model matrix are attractive.

As an illustration of the usefulness of the rank criterion, the three best 24-run 7-factor designs in terms of the GA4 classifier have ranks of 19, 24 and 22 for the interaction model matrix. Therefore, the design ranked second in terms of the GA4 classifier might be preferred if the detection of two-factor interaction effects is deemed very important.

2.5 Estimability and efficiency in subsets of five factors

The designs we consider are intended for screening experiments. Often, only a limited number of factors turn out to have significant effects. Therefore, it makes sense to study the projections of the designs obtained by dropping columns. This results in designs with fewer factors. Ideally, these designs allow estimation of an interaction model in the remaining factors.

To quantify the potential of a design in this context, Loepky et al. (2007) defined the

projection estimation capacity (PEC) vector $(\text{PEC}_3, \text{PEC}_4, \dots, \text{PEC}_n)$, where PEC_x is the fraction of x -factor subsets for which the interaction model can be estimated. For 24- and 28-run designs, interaction models with any seven or more factors are not estimable, so that $\text{PEC}_x = 0$ for $x \geq 7$. For 32- and 36-run designs, $\text{PEC}_x = 0$ for $x \geq 8$. In addition to the PEC_x value, Loepky et al. (2007) defined the projection information capacity for x factors (PIC_x) as the average D -efficiency with which all interaction models in x -factors can be estimated. The vector of all PIC_x values is the PIC vector. In this paper, we use the PEC_5 and PIC_5 values rather than the complete PEC and PIC vectors for two reasons. First, scalar criteria are simpler to handle. Second, because $\text{PEC}_3 \geq \text{PEC}_4 \geq \text{PEC}_5$, designs with a large PEC_5 value also have large PEC_4 and PEC_3 values. Similarly, designs with large PIC_5 values also possess large PIC_4 and PIC_3 values.

3 Literature review

Most designs with 24–36 runs in the literature have been constructed using projections from Hadamard designs with 23, 27, 31 or 35 factors or using a bottom-up design enumeration. We review these techniques in this section. We briefly discuss two additional methods for 32-run designs in Section 6.

3.1 Projection-based approach

The papers that discuss projections from Hadamard designs start from a catalog of non-isomorphic Hadamard matrices such as the one from Sloane (2015a). A Hadamard matrix of order N , denoted by H_N , is an $N \times N$ matrix satisfying the equality $H_N' H_N = N I_N$. As a result, the columns of a Hadamard matrix are orthogonal to each other. This explains why Hadamard matrices can be used as orthogonal designs. Note that the rows of a Hadamard matrix are also orthogonal to each other.

The Hadamard matrices relevant for this article have dimensions 24, 28, 32 and 36. There

are 60 non-isomorphic Hadamard matrices of order 24 (Ito et al., 1981; Kimura, 1989), 487 non-isomorphic Hadamard matrices of order 28 (Kimura and Ohmori, 1986; Kimura, 1994a,b) and 13,710,027 non-isomorphic Hadamard matrices of order 32 (Kharagani and Tayfeh-Rezaie, 2010, 2013). The number of non-isomorphic Hadamard matrices of order 36 is unknown.

A Hadamard design is obtained from a normalized Hadamard matrix (whose first column and first row consist of +1s only) by dropping the first column. In this way, Evangelaras et al. (2004), Deng and Tang (2002), Ingram and Tang (2005), Belcher-Novosad and Ingram (2003) and Loepky et al. (2007) obtain 60 Hadamard designs involving 24 runs and 23 factors and 487 Hadamard designs involving 28 runs and 27 factors. Similarly, Angelopoulos and Koukouvinos (2007) and Evangelaras et al. (2006) start their search by dropping the first column from 15 normalized Hadamard matrices of order 32 and 12 normalized Hadamard matrices of order 36. The designs recommended by all these authors involve selections of columns of these starting designs. So, these authors study projections of 23-, 27-, 31- and 35-factors designs onto fewer factors.

One major problem with this projection approach is that it overlooks many alternative designs in 24, 28, 32 and 36 runs involving 23, 27, 31 and 35 factors, respectively. Therefore, potentially interesting projections of these designs obtained by dropping columns may be missed. This is because any given Hadamard matrix may give rise to several non-isomorphic designs, possibly with different statistical properties. We explain this in detail in the supplementary Section A. It is, for example, known that there are 130 non-isomorphic 23-factor designs with 24 runs (Hedayat et al., 1999) and 7570 non-isomorphic 27-factor designs in 28 runs (Sloane, 2015c, A048885), while there are 60 Hadamard matrices of order 24 and 487 of order 28. Any search involving only 60 23-factor Hadamard designs and 487 27-factor Hadamard designs is therefore incomplete and may lead to suboptimal designs. This is also the case for the searches based on 15 31-factor and 12 35-factor Hadamard designs, especially given the fact that, in a personal communication, Brendan McKay mentioned to Kharagani

and Tayfeh-Rezaie (2013) that the number of non-isomorphic 32-run 31-factor designs equals 355,293,682.

Yamamoto et al. (1995) and Yumiba et al. (1997) enumerate all non-isomorphic projections of all 130 non-isomorphic designs with 23 factors and 24 runs into five and six factors, but the focus of these authors was on enumeration rather than on identifying good experimental designs. A positive aspect, however, is that they started their search from the complete set of 130 non-isomorphic orthogonal designs. Evangelaras et al. (2004) discuss non-isomorphic projections of the 60 Hadamard designs involving 24 runs and up to five factors, and report their projection properties, generalized resolution and GWLP. Starting from the same 60 Hadamard designs, Deng and Tang (2002) list the best three projections with up to eight factors according to the GA5 classifier, whereas Ingram and Tang (2005) specify a stepwise algorithm to search for minimum G -aberration and minimum G_2 -aberration designs among all projections, and provide a list of recommended designs with 3–23 factors. Belcher-Novosad and Ingram (2003) modify the algorithm of Ingram and Tang (2005) to overcome the computational difficulties when addressing 28-run designs and report a near-minimum G -aberration design with 28 runs and 17 factors. Bulutoglu and Ryan (2015) used a heuristic approach to find 28-run designs with 15–26 factors with near minimum G_2 -aberration based on all 7570 27-factor designs. Starting from the 60 Hadamard designs with 24 runs and the 487 Hadamard designs with 28 factors, Loeppky et al. (2007) propose a systematic approach to search for 24-run designs with 6–23 factors and 28-run designs with 6–27 factors with good PEC values.

Angelopoulos and Koukouvinos (2007) obtained 15 32-run 31-factor Hadamard designs from 15 Hadamard matrices of order 32 constructed from circulant matrices. They classified projections of these Hadamard designs into 3–6 factors in terms of the D -efficiency for the interaction model. Likewise, Evangelaras et al. (2006) obtained 12 36-run 35-factor Hadamard designs from 12 Hadamard matrices of order 36 constructed from circulant matrices. They classified projections into 3–7 factors in terms of the generalized resolution, the

GWLP and the number of distinct runs. Angelopoulos and Koukouvinos (2007) identified all isomorphism classes based on the projections of the 12 35-factor designs into up to 5 factors. Finally, Bulutoglu and Ryan (2015) used a heuristic approach to find 36-run designs with 10–18 factors with near minimum G_2 -aberration based on the two Hadamard matrices of order 36 given by Sloane (2015a).

3.2 Enumeration-based approach

Vijayan (1976) proved that any N -run orthogonal two-level design with more than $N - 5$ factors can be extended to a Hadamard matrix, while Bulutoglu and Kaziska (2009) conjectured that this is even the case for any design with more than $N - 8$ factors. Therefore, all designs with $N - 5$ or perhaps even with $N - 8$ factors can be obtained by projections from Hadamard matrices. For smaller numbers of factors, however, orthogonal designs exist that cannot be obtained from projections. For example, Yumiba et al. (1997) report 1317 non-isomorphic six-factor projections from all 130 23-factor designs in 24 runs. Evangelaras et al. (2007) and Schoen et al. (2010), however, report 1350 non-isomorphic six-factor designs. As a result, 33 six-factor designs cannot be found by projection from 23-factor designs. A bottom-up enumeration of all possible designs would include both the latter 33 designs as well as the former 1317 designs.

Various authors have used a bottom-up enumeration approach, in which all designs involving small numbers of factors are extended factor by factor to obtain all possible designs with large numbers of factors. The enumeration of all regular designs with 8, 16, 32 and 64 runs by Chen et al. (1993) is the earliest one relevant to this paper. A good summary of the literature on regular designs can be found in Wu and Hamada (2009) and Mee (2009). Li et al. (2004) introduce a bottom-up enumeration method to generate regular and nonregular designs for 3–6 factors with up to 36 runs. Starting from a complete set of three-factor designs, they generated all possible extensions with one additional factor. Each newly obtained set with an additional factor was then sorted according to the GA5 classifier, and up to ten

of the designs in each class were extended further.

Evangelaras et al. (2007) express the design generation problem as a linear system of equations. For 24 runs, they generate all non-isomorphic designs with up to six factors, while, for 28, 32 and 36 runs, they generate all non-isomorphic designs with up to five factors. They classify the designs in terms of G -aberration and in terms of their D -efficiency for the interaction model. Angelopoulos et al. (2007) generate all non-isomorphic 7-factor designs with 24 runs and all non-isomorphic 6-factor designs with 28 and 32 runs. Their primary classification criteria are the G_2 -aberration criterion for the 24-run designs, the D -efficiency for the interaction model for the 28-run designs and the GWLP for the 32-run designs. Finally, Bulutoglu and Margot (2008) generate all 7-factor 24-run designs and all strength-3 24-run designs with up to 11 factors.

As opposed to the authors named above, Schoen et al. (2010) completely enumerate all non-isomorphic 24-run designs for any number of factors. For 28 runs, they generate all non-isomorphic designs with up to seven factors. For 32 runs, they generate all strength-3 designs. These 32-run designs involve at most 16 factors and were studied in detail by Schoen and Mee (2012). Recently, Eendebak and Schoen (2017) enumerated all 32-run 7-factor designs and identified those with the best D -efficiency for the interaction model.

Bulutoglu and Ryan (2015) list the distance distributions of the minimum G_2 -aberration 24-run designs. These distance distributions can be converted into GWLPs using Krawtchouk polynomials. They also enumerate subsets of 28-run designs for 3–14 factors and 36-run designs for 3–9 factors that are guaranteed to include the minimum G_2 -aberration designs, and they identify the GWLP of the minimum G_2 -aberration designs.

3.3 Gaps in the literature

The literature review reveals that a substantial amount of knowledge concerning minimum G -aberration designs with 24, 28, 32 and 36 runs is lacking. For these run sizes, minimum G -aberration or near minimum G -aberration designs are unknown for more than 12, 7, 16

Table 1: Numbers of non-isomorphic 24-run two-level designs

Factors	Designs	Factors	Designs	Factors	Designs
3	4	10	38,592,861	17	4,385,567
4	10	11	52,912,678	18	1,502,242
5	63	12	51,154,497	19	409,478
6	1,350	13	43,092,737	20	86,725
7	57,389	14	31,833,387	21	13,833
8	1,470,157	15	19,960,039	22	1,604
9	12,952,435	16	10,351,396	23	130

and 7 factors, respectively. We fill this gap in the literature by identifying minimum G -aberration of near minimum G -aberration 24-, 28-, 32- and 36-run designs with up to 23, 14, 31 and 18 factors, respectively. For the case of 28 runs, we further provide good projections of 27-factor designs into 15–26 factors in terms of G -aberration.

Minimum G_2 -aberration designs with 28 runs and more than 14 factors, 32 runs and more than 16 factors and 36 runs and more than 9 factors are also unknown. We do not contribute to filling this gap because we used restricted enumeration techniques directed to the F_3 and F_4 vectors rather than the A_3 and A_4 counts needed to minimize G_2 -aberration. However, we do provide designs in 24 runs and up to 23 factors and designs in 28 runs and up to 14 factors that are good in terms of the mixed A4 classifier, which have an advantage over designs that sequentially optimize A_3 and A_4 counts.

4 Designs with 24 runs

The numbers of non-isomorphic 24-run designs generated by Schoen et al. (2010) for 3–23 factors are given in Table 1. The large numbers of designs necessitate the use of computationally cheap criteria to search for optimal designs. In this section, we present a selection from the top-3 designs in terms of the GA4 and mixed A4 classifiers and a comparison with the best designs available in the literature.

4.1 Selected designs

We identified the top-3 designs according to the GA4 and mixed A4 classifiers. For 4–12 factors and 17–23 factors, the top-3 designs according to the GA4 classifier match the top-3 designs according to the mixed A4 classifier. For 13–16 factors, the classifiers produce different top-3 designs. As there are sometimes multiple designs with the same ranking in terms of the GA4 classifier or the mixed A4 classifier, we also evaluated all top-3 designs using the GA6 classifier, the PIC_5 value and the PEC_5 value. All top-3 designs are available in the supplementary materials to this paper, as are detailed tables describing their features. Here, we only present the most important designs we identified and their key features. A top-3 design is presented in the tables in this article if it is best according to (i) the GA4 classifier (a simplified version of the CFV on which the G -aberration criterion is based), (ii) the mixed A4 classifier (which is related to the G_2 -aberration criterion), (iii) the rank of the interaction model matrix, (iv) the PIC_5 value, or (v) the PEC_5 value. When there is a tie for one criterion, we use the other criteria as tie breakers. The selected designs with 4–12 factors and their main characteristics are given in Table 2, while the selected designs with 13–23 factors and their main characteristics are shown in Table 3.

The first columns in the tables identify individual designs and contain labels of the form $n.i$ or $n - i$, where n is the number of factors, the i in the label $n.i$ is the ranking of the corresponding design in terms of the GA6 classifier and the i in the label $n - i$ is the ranking of the corresponding design in terms of the mixed A6 classifier (for designs with 13–16 factors only). If multiple designs exist with the same ranking according to the GA6 or mixed A6 classifier, the designs are reported in descending order of their PIC_5 value and letters are included in the labels of these designs, starting with an a for the best design.

The tables' next few columns contain details about the generalized word counts A_3 and A_4 and the frequency vectors F_3 and F_4 . Note that J_3 - and J_4 -characteristics not reported have zero frequencies. The tables' last column presents PIC_5 values. Table 2 also shows the rank of the interaction model matrix. This rank is not shown in Table 3 because it equals

Table 2: Selected 24-run designs with 4–12 factors. Designs identified previously are indicated by superscripts. For all designs in the table, $F_3(24, 16) = (0, 0)$ and $F_4(24, 16) = (0, 0)$.

ID	A_3	$F_3(8)$	A_4	$F_4(8)$	Rank	PIC_5
4.1 ^{1,3,4}	0.00	0	0.11	1	11	-
5.1 ^{1,3,4}	0.00	0	0.56	5	16	0.86812
5.4	0.22	2	0.11	1	16	0.93901
6.1 ^{1,3,4,5}	0.00	0	1.67	15	18	0.86812
6.3 ^{4,5}	0.22	2	1.00	9	22	0.90049
7.1 ^{1,3,4,6}	0.00	0	3.89	35	19	0.86812
7.2 ^{4,6}	0.44	4	2.33	21	24	0.89668
8.1 ^{1,3,4}	0.00	0	7.78	70	20	0.86812
8.2 ⁴	0.67	6	5.78	52	23	0.87690
9.1 ^{1,2,3}	0.00	0	14.00	126	21	0.86812
9.2	1.11	10	10.89	98	24	0.86782
10.1 ^{1,2,3}	0.00	0	23.33	210	22	0.86812
10.2	1.56	14	18.67	168	24	0.86493
11.1 ^{1,2,3}	0.00	0	36.67	330	23	0.86812
11.2	2.00	18	30.00	270	24	0.86428
12.1 ^{1,2,3}	0.00	0	55.00	495	24	0.86812

¹Ingram and Tang (2005); ²Loeppky et al. (2007); ³Bulutoglu and Ryan (2015); ⁴Deng and Tang (2002);

⁵Evangelaras et al. (2007); ⁶Angelopoulos et al. (2007).

24 for all selected designs with 12 or more factors. Instead, Table 3 shows the PEC_5 values of the selected designs. We do not show the PEC_5 value in Table 2, because it equals one for all selected designs with up to 12 factors. Designs available from the literature are indicated in Tables 2 and 3 using superscripts.

4–12 factors

The selected designs with 4–12 factors in Table 2 include strength-3 designs. These strength-3 designs have a zero entry in the column labeled A_3 in the table, as a result of which the entire $F_3(24, 16, 8)$ vector is also zero. As a consequence, the main effects can be estimated independently from the two-factor interactions when these designs are used. For four as well as six factors, two different strength-3 designs exist. Table 2 only shows the better of the two designs in each case. For other numbers of factors, there is only one strength-3 design.

Table 3: Selected 24-run designs with 13–23 factors. Designs identified previously are indicated by superscripts. For design 23.3, $F_3(24) = 1$. All other designs have $F_3(24) = 0$.

ID	A_3	$F_3(16)$	$F_3(8)$	A_4	$F_4(24)$	$F_4(16)$	$F_4(8)$	PEC_5	PIC_5
13.1	10.00	0	90	41.67	1	0	366	0.96503	0.82092
13-1 ^{1,2}	6.00	6	30	55.00	0	0	495	0.98834	0.84833
13-3	8.00	2	64	48.33	0	0	435	0.99456	0.84857
14.1	15.11	0	136	53.22	1	4	454	0.96054	0.80565
14-1	12.00	4	92	61.00	1	0	540	0.98701	0.83853
15.1a	20.00	0	180	67.67	2	21	507	0.95738	0.80178
15-1 ¹	18.11	3	151	73.00	3	0	630	0.97269	0.81988
15-2a	18.11	8	131	73.00	1	4	632	0.97303	0.81937
16.1 ¹	24.89	0	224	90.22	4	32	648	0.94872	0.79405
16-1	24.44	4	204	91.11	6	0	766	0.96062	0.80489
16-3a	24.44	6	196	91.11	3	3	781	0.96268	0.80614
17.1a	31.11	0	280	115.11	2	83	686	0.95669	0.79808
18.1a ¹	37.33	0	336	148.00	9	81	927	0.94993	0.79346
19.1a ¹	45.33	0	408	185.33	4	38	1480	0.93455	0.77885
19.2a	45.33	0	408	185.33	4	40	1472	0.93627	0.78006
20.1 ¹	53.33	0	480	231.67	5	48	1848	0.93421	0.77857
20.2a	53.33	0	480	231.67	5	160	1400	0.95227	0.79290
21.1 ¹	63.33	0	570	285.00	0	0	2565	0.90226	0.75259
21.3a	63.33	0	570	285.00	5	160	1880	0.93489	0.77852
22.1 ^{1,2}	73.33	0	660	348.33	0	0	3135	0.90226	0.75259
22.2a	73.33	0	660	348.33	10	99	2649	0.94087	0.78134
23.1 ^{1,2,P}	84.33	0	759	421.67	0	0	3795	0.90226	0.75259
23.3	83.33	12	702	421.67	5	60	3510	0.92781	0.77243

¹Ingram and Tang (2005); ²Loeppky et al. (2007); ^PPlackett and Burman (1946).

13–23 factors

None of the designs with 13 or more factors has a strength of 3. As a result, when 13 or more factors are studied, the main effects are aliased to some extent with the two-factor interactions and the generalized resolution is smaller than 4. The generalized resolution is $3^{2/3}$ for most of the selected designs, as they have no J_3 -characteristics of 16 and 24. For the designs 13-1, 13-3, 14-1, 15-1, 15-2a, 16-1, and 16-3a, the generalized resolution is $3^{1/3}$, due to their nonzero entry in the column labeled $F_3(16)$. Nevertheless, these designs have better PEC_5 and PIC_5 values than the corresponding minimum G -aberration designs. Similarly, design 23.3 has $F_3(24) = 1$ and thus a generalized resolution of 3 only, but its PEC_5 and PIC_5

values are better than those of the minimum G -aberration design. The design involves some complete aliasing between the two-factor interactions and the main effects in one subset of three factors.

4.2 Comparison with designs from the literature

A detailed discussion of 24-run orthogonal designs available in the literature can be found in the supplementary Section B. In that section, we compare these designs to our newly found designs in terms of the CFV, the GWLP and the PIC_5 value. In this section, we highlight 18 designs that were not identified by the extensive searches of Ingram and Tang (2005) (IT), Loepky et al. (2007) (LST), Bulutoglu and Ryan (2015) (BR) and Deng and Tang (2002) (DT). These 18 designs were included in Tables 2 and 3 for the following reasons:

1. Design 5.4 has a better D -efficiency for the interaction model than design 5.1 (found by IT, BR and DT, and also included in the selected designs).
2. Design 9.2 has a higher rank for the interaction model matrix than design 9.1 (found by IT, LST and BR, and also included in the selected designs).
3. Design 10.2 has a higher rank for the interaction model matrix than design 10.1 (found by IT, LST and BR, and also included in the selected designs).
4. Design 11.2 has a higher rank for the interaction model matrix than design 11.1 (found by IT, LST and BR, and also included in the selected designs).
5. Design 13.1 has a better F_6 vector than design 13.2 (found by IT).
6. Design 13-3 has a better PEC_5 value than design 13-1 (found by IT and LST, and also included in the selected designs).
7. Design 14.1 has $F_4(24, 16, 8) = (1, 4, 454)$, which is better than the vector $(1, 7, 438)$ of design 14.2 (obtained by IT).
8. Design 14-1 has $F_3(24, 16, 8) = (0, 4, 92)$, which is better than the vector $(0, 12, 60)$ of 14-2 (IT, LST). Both designs has the same A_3 and A_4 values.
9. Design 15.1a has $F_4(24, 16, 8) = (2, 21, 507)$, while the best design of IT has $F_4(24, 16, 8) =$

- (4, 8, 541). That design is not a top-3 design in terms of the GA4 classifier.
10. Design 15-2a has a better PEC_5 value than design 15-1 (IT).
 11. Design 16-1 has $F_3(24, 16, 8) = (0, 4, 204)$, which is better than the vector $(0, 6, 196)$ of design 16-2a (IT). Both designs have the same A_3 and A_4 values.
 12. Design 16-3a has a better PIC_5 value than design 16-1.
 13. Design 17.1a has a better PIC_5 value than design 17.1b (IT).
 14. Design 19.2a has a better PIC_5 value than design 19.1a (IT).
 15. Design 20.2a has a better PIC_5 value than design 20.1 (IT).
 16. Design 21.3a has a better PIC_5 value than design 21.1 (IT).
 17. Design 22.2a has a better PIC_5 value than design 22.1 (IT, LST).
 18. Design 23.3 has a better PIC_5 value than design 23.1 (IT, LST, Plackett and Burman (1946)).

While our complete search produced some improvements over the best known designs so far, it also confirms that the best designs available in the literature were very good. Finally, all 24-run designs we listed in Tables 2 and 3 perform well in terms of the design selection criteria we use in this paper. Therefore, we can recommend all of these designs.

5 Designs with 28 runs

In this section, we discuss 28-run designs involving 3–27 factors. For 3–7 factors, catalogs with all non-isomorphic designs are available. The numbers of non-isomorphic designs for these numbers of factors are shown in the rightmost column of Table 4. At present, no complete catalog exists with all non-isomorphic 28-runs designs with 8–26 factors, but it is known that there are 7570 non-isomorphic designs in 27 factors (Sloane, 2015c, sequence A048885). To the best of our knowledge, the complete set of non-isomorphic 27-factor designs has not yet been explored in terms of G -aberration.

In Section 5.1, we outline a step-by-step enumeration procedure to identify top-3 designs

with 3–14 factors according to the GA4 and mixed A4 classifiers, and we characterize the resulting designs. In Section 5.2, we adopt a projection approach to find good designs involving 15–26 factors, starting from the 7570 non-isomorphic 27-factor designs, which we also investigate in terms of G -aberration.

5.1 3–14 factors

5.1.1 Enumeration approach

We obtained the 28-run designs for 3–14 factors using a procedure similar to the one adopted by Bulutoglu and Ryan (2015). There are four non-isomorphic 3-factor designs involving 28 runs. These four designs have $F_3(28, 20, 12, 4)$ vectors equal to $(1, 0, 0, 0)$, $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ and $(0, 0, 0, 1)$. The best of the four three-factor designs is the one with $(0, 0, 0, 1)$ as F_3 vector. We extended this design by adding one extra factor column at a time, using the enumeration procedure of Schoen et al. (2010). Each time a factor column has been added, we discard the designs with J_3 -characteristic values greater than 4 and designs that are not in lexicographically minimal form. In other words, we construct sets of non-isomorphic designs with an $F_3(28, 20, 12, 4)$ vector of the form $(0, 0, 0, a)$, where $a = \binom{n}{3}$. The resulting sets of designs only involve the smallest possible J_3 -characteristic. Therefore, these sets contain the designs that are optimal in terms of the GA4 classifier.

Using our step-by-step procedure, we were able to enumerate the complete set of 28-run designs with $F_3(28, 20, 12, 4) = (0, 0, 0, a)$. In the second column of Table 4, we show the numbers of non-isomorphic designs we found with this property. Our results match those of Bulutoglu and Ryan (2015) and indicate that no designs with a maximum J_3 -characteristic of 4 exist for 15 or more factors. The numbers in the enumerated class are very small compared to the total number of non-isomorphic designs for 3–7 factors. The new enumeration approach thus saves substantial computing time, as fewer designs need to be extended and evaluated.

We also studied all designs with J_3 -characteristics larger than 4 that can be obtained by

Table 4: Numbers of non-isomorphic 28-run designs

n	$F_3(28, 20, 12, 4)$	
	$(0,0,0,a)$	General
3	1	4
4	3	7
5	15	127
6	320	17,826
7	12,194	5,882,186
8	63,606	
9	20,552	
10	841	
11	45	
12	10	
13	2	
14	1	
15–27	0	

adding extra columns to the 12-, 13- and 14-factor designs with $F_3(28, 20, 12, 4) = (0, 0, 0, a)$. The extensions from the 12- and 13-factor design all have $F_3(12) \geq 36$ and $F_4(20) \geq 1$. It is unlikely that these designs are optimal in terms of G -aberration. The extensions from the 14-factor design have $F_3(20, 12) = (1, 18)$. This is certainly not optimal, because any sub-design from the 28-run Plackett-Burman design has a better G -aberration. Therefore, for 15 or more factors, we need to use a different procedure to find designs that perform well in terms of the GA4 classifier, such as the one described in Section 5.2.

The 28-run designs with $F_3(28, 20, 12, 4) = (0, 0, 0, a)$ have the smallest possible A_3 values for any given number of factors. To find 4–14 factor designs that optimize the mixed A4 classifier, we therefore only have to rank the designs with $F_3(28, 20, 12, 4) = (0, 0, 0, a)$ according to their A_4 values, followed by the F_4 vector.

5.1.2 Results

For 4–14 factors, we identified 77 groups of designs that are best, second best or third best according to the GA4 classifier or the mixed A4 classifier, and which differ according to the GA6 classifier. All designs in the 77 groups permit estimation of the interaction models

Table 5: Selected 28-run designs with 4–14 factors. For all designs in the table, $F_3(28, 20, 12) = (0, 0, 0)$ and $F_4(28) = 0$.

ID	A_3	$F_3(4)$	A_4	$F_4(20, 12, 4)$			PIC_5
4.1.1	0.08	4	0.02	0	0	1	-
5.1.3	0.20	10	0.1	0	0	5	0.94091
6.1.3	0.41	20	0.31	0	0	15	0.93726
7.1.1	0.71	35	0.88	0	1	34	0.93102
8.1.1	1.14	56	2.9	0	9	61	0.91075
9.1.1	1.71	84	5.51	0	18	108	0.90738
10.1.1	2.45	120	11.47	0	44	166	0.89314
10-1.1	2.45	120	10.49	6	20	184	0.89621
10-2.1	2.45	120	10.82	2	34	174	0.89642
11.1.1	3.37	165	21.43	0	90	240	0.87942
11-1.1	3.37	165	18.82	3	65	262	0.88890
12.1.1	4.49	220	32.14	0	135	360	0.87934
12-1.1	4.49	220	28.22	6	93	396	0.88859
13.1.1	5.84	286	46.43	0	195	520	0.87933
14.1.1	7.43	364	65	0	273	728	0.87931

involving any subset of five factors. Hence, $PEC_5 = 1$ for each of the designs.

In supplementary tables, we show the designs' GA6 classifiers, along with their $A_3, \dots, A_6$ values and their PIC_5 values. The A_3 and A_4 values, the F_3 and F_4 vectors and the PIC_5 values of a selection of the designs are shown in Table 5. The first column in the table identifies individual designs using labels of the form $n.i.j$ or $n-i.j$, where n is the number of factors. The design with label $n.i.j$ is ranked j th in terms of the GA6 classifier, among all designs with the i th best GA4 classifier. The design with label $n-i.j$ is ranked j th in terms of the mixed A6 classifier, among all designs with the i th best mixed A4 classifier.

For each number of factors, we report the design with the best PIC_5 value we found across all design groups, the design with the best PIC_5 values among those that are optimal in terms of the GA4 classifier and those with the best PIC_5 values among those that are optimal in terms of the mixed A4 classifier. All designs listed have the maximum possible rank of 28 for the interaction model matrix.

Because all designs in Table 5 have $F_3(28, 20, 12, 4) = (0, 0, 0, a)$, we only show the value of a in the table, in the column labeled $F_3(4)$. That value always equals $\binom{n}{3}$. This result implies

that the correlation between any main-effect contrast vector and any two-factor interaction contrast vector involving two other factors equals $\pm 1/7$. As a result, when estimating a main-effects model based on any of the designs in the table, the main effects estimates are biased only to a small extent by active two-factor interactions.

The table shows that the minimum G -aberration designs for 28 runs and 7–14 factors all have an F_4 vector of the form $F_4(28, 20, 12, 4) = (0, 0, b, a)$, with $a + b = \binom{n}{4}$ and $b \geq 0$. The fact that b is generally much smaller than a implies that the designs generally perform well when a few two-factor interactions need to be estimated. This is confirmed by the high PIC_5 values: the smallest values in the table are all close to 0.88. Table 5 shows four designs that are not optimal in terms of the GA4 classifier, namely designs 10-1.1, 10-2.1, 11-1.1 and 12-1.1. The designs 10-1.1, 11-1.1 and 12-1.1 are included because they are optimal in terms of the mixed A4 classifier. Their A_4 values are smaller than those of the designs that optimize the GA4 classifier. Design 10-2.1 is included in Table 5 because it has the best PIC_5 value among the top-3 10-factor designs we identified.

The designs with 8–14 factors identified by Loeppky et al. (2007) as well as those consisting of the first 6–14 columns of the popular 28-run Plackett-Burman design all have a generalized resolution of $3^{4/7}$, while our designs have the best possible generalized resolution, $3^{6/7}$. The best 7-factor design of Loeppky et al. (2007) matches our best 7-factor design in terms of generalized resolution ($3^{6/7}$). However, our design has $F_4(12, 4) = (1, 34)$, while the competing design has a slightly worse F_4 vector of $(4, 31)$. For six factors, our best design is equivalent to the best one of Loeppky et al. (2007) in terms of the generalized resolution and the GA4 classifier, but better in terms of the GA5 classifier.

Figure 1 compares the PIC_5 values of the designs with 6–14 factors found by Loeppky et al. (2007) with those of our designs and the designs obtained by taking the first 6–14 columns from the 28-run 27-factor Plackett-Burman design. The latter designs are clearly inferior in terms of PIC_5 values. The PIC_5 values of our best six-factor designs and those in Loeppky et al. (2007), Angelopoulos et al. (2007) and Li et al. (2004) coincide. For 7–14

Figure 1: Best PIC_5 values of our 28-run designs with 6–14 factors (crosses), those obtained by Loeppky et al. (2007) (squares) and designs derived from the Plackett-Burman design (asterisks).

factors, our best designs outperform those of Loeppky et al. (2007) (Angelopoulos et al. (2007) and Li et al. (2004) do not study designs with more than six factors).

Finally, Bulutoglu and Ryan (2015) present the GWLP of minimum G_2 -aberration designs. Our list of selected designs with 4–14 factors includes minimum G_2 -aberration designs, but we report the best of the minimum G_2 -aberration designs in terms of the mixed A4 classifier. Our results are therefore more refined than those of Bulutoglu and Ryan (2015).

5.2 15–27 factors

5.2.1 Approach

For 15 or more factors, no 28-run designs exist with an F_3 vector of the form $F_3(28, 20, 12, 4) = (0, 0, 0, a)$. The next best form of F_3 vector is $F_3(28, 20, 12, 4) = (0, 0, b, a)$ with $b > 0$. By exploring the complete set of non-isomorphic seven-factor designs, we identified 5,434,439

designs with such an F_3 vector. Extending all the designs with $F_3(28, 20, 12, 4) = (0, 0, b, a)$ by adding one factor at a time would allow us to obtain a complete catalog of designs with the same kind of F_3 vector. Exploring this catalog would allow us to identify the minimum G -aberration designs for 15 factors and more. Unfortunately, this approach is computationally infeasible due to the extremely large numbers of non-isomorphic designs involving eight or more factors. For example, extending 7,000 non-isomorphic seven-factor designs for which $F_3(28, 20, 12, 4) = (0, 0, b, a)$ resulted in 1,792,801 non-isomorphic eight-factor designs for which $F_3(28, 20, 12, 4) = (0, 0, b, a)$. So, a rough estimate of the number of eight-factor designs of the type we are interested in is $1,792,801 \times 5,434,439/7,000 = 1,391,838,239$. Therefore, except in the special case of 27 factors, for which all non-isomorphic designs are known, identifying the minimum G -aberration 28-run designs for 15 or more factors is infeasible.

The infeasibility of the enumeration approach prompted us to adopt a projection approach for finding good designs involving 15–26 factors, starting from the complete catalog of non-isomorphic 27-factor designs. Before doing so, however, we first identified top-3 designs with 27 factors in terms of the GA4 classifier. To this end, we constructed all 7570 non-isomorphic 27-factor designs starting from the complete catalog of all 487 non-isomorphic Hadamard matrices of order 28. The 7570 designs all have the same GWLP. There are twelve 27-factor designs with $F_3(28, 20, 12, 4) = (0, 0, 351, 2574)$. These twelve designs are best in terms of the GA4 classifier and in terms of the mixed A4 classifier.

As the best F_3 vector for 27 factors is of the form $F_3(28, 20, 12, 4) = (0, 0, b, a)$, any minimum G -aberration design with fewer than 27 factors must also have an F_3 vector of the form $F_3(28, 20, 12, 4) = (0, 0, b, a)$. The smaller the number of factors, the smaller the anticipated b value of the corresponding minimum G -aberration design. For 20–26 factors, we identified the top-3 designs in terms of the GA4 classifier among all possible projections from these twelve designs. In doing so, we considered more than 10,000,000 projections into 20-factor designs. Due to this large number of relevant 20-factor designs, it was computationally

infeasible to study all projections of the best 12 27-factor designs onto 15–19 factors. For this reason, for the cases of 15–19 factors, we studied all projections from the six best 20-factor designs we obtained in terms of the GA4 classifier.

Using the projection approach to identify the best designs in terms of the mixed A4 classifier is computationally infeasible. This is due to the fact that all 7570 27-factor designs have the same GWLP. So, none of 7570 designs stand out in terms of their A_3 and A_4 values. As a result, there is no limited subset of designs that is the most promising for studying projections that perform well in terms of the mixed A4 classifier. The only sensible thing would be to study projections from all 7570 designs.

Every time we encountered a large class of designs with the same performance in terms of the GA4 classifier in the course of our projection approach, we restricted attention to the top-3 designs in terms of the GA6 classifier within the class. For example, the 16-factor designs that perform second best according to the GA4 classifier all have $F_3(12, 4) = (47, 513)$ and $F_4(20, 12, 4) = (21, 199, 1600)$. The three best F_5 vectors for these designs are $F_5(16, 8) = (54, 1686)$, $(56, 1678)$ and $(60, 1662)$, while all three designs have $F_6(24, 16, 8) = (0, 15, 3168)$.

5.2.2 Results for 20–27 factors

Our list of top-3 designs according to the GA4 classifier for 20–27 factors includes 42 groups of designs in total. The groups differ according to the GA6 classifier. In supplementary tables, we show the GA6 classifier of a representative of each of the groups, along with the PEC_5 and PIC_5 values. Properties for a selection of the designs are shown in Table 6. The labels of the form $n.i.j$ now apply to the projections from the 12 27-factor designs. For each number of factors, designs were initially included in Table 6 only if they have the best PIC_5 value of all top-3 designs, or if they have the best PIC_5 value among the designs that perform best in terms of the GA4 classifier. Remarkably, this procedure did not result in designs involving 20–22 factors with a PEC_5 value of 1, even though design 23.1 involves a larger number of factors and yet has a PEC_5 value equal to 1. For this reason, we added

Table 6: Selected designs with 28 runs and 20–27 factors. For all designs in the table, $F_3(28, 20) = (0, 0)$ and $F_4(28) = 0$.

ID	A_3	$F_3(12)$	$F_3(4)$	A_4	$F_4(20)$	$F_4(12)$	$F_4(4)$	PEC_5	PIC_5
20.1.1	44.00	127	1013	196.84	30	510	4305	0.99542	0.85923
20.2.1	44.16	128	1012	196.02	12	559	4274	0.99845	0.86129
20.x	44.82	132	1008	195.04	0	589	4256	1	0.86037
21.1.1	52.12	153	1177	240.67	34	624	5327	0.99568	0.85815
21.3.1	52.29	154	1176	240.67	8	702	5275	0.99833	0.86000
21.x	52.61	156	1174	240.18	0	723	5262	1	0.85980
22.1.1	60.98	181	1359	293.12	10	851	6454	0.99848	0.85949
22.2.3	60.98	181	1359	293.12	11	848	6456	0.99859	0.85957
22.x	61.31	183	1357	292.80	0	879	6436	1	0.85919
23.1.1	70.59	211	1560	354.43	0	1064	7791	1	0.85904
23.3.2	70.59	211	1560	354.43	8	1040	7807	0.99878	0.85916
24.1.1	80.82	242	1782	425.18	0	1276	9350	1	0.85885
24.3.1	80.82	242	1782	425.18	11	1243	9372	0.99885	0.85899
25.1.1	92.00	276	2024	506.00	0	1518	11132	1	0.85866
25.3.1	92.00	276	2024	506.00	15	1473	11162	0.99881	0.85872
26.1.1	104.00	312	2288	598.00	0	1794	13156	1	0.85866
27.1.1 ^P	117.00	351	2574	702.00	0	2106	15444	1	0.85866

^PPlackett and Burman (1946).

the designs 20.x, 21.x and 22.x, which are projections from design 23.1, to the table. While these designs are neither optimal in terms of the GA4 classifier nor in terms of the PIC_5 value, they are good design options for 20–22 factors with $PEC_5 = 1$.

Each of the designs in the table has a PIC_5 value greater than 85% and a PEC_5 value greater than 99.5%. Designs with a given number of factors only show very small differences in PEC_5 and PIC_5 values. Therefore, whenever at most five factors are expected to be active, we recommend a design with $PEC_5 = 1$ over designs with a slightly higher PIC_5 value and $PEC_5 < 1$. When more than five factors are expected to be active, we recommend the best tabulated designs in terms of the GA4 classifier.

Design 27.1.1 corresponds to the 28-run Plackett-Burman design, so that we recommend this design when 27 factors are studied. This design permits estimation of interaction models involving any subset of five factors, so that $PEC_5 = 1$. The average D -efficiency with which all these models can be estimated is 0.85866.

Figure 2: Differences in $F_3(12)$ frequencies between the best 28-run designs with 15–26 factors obtained by Loeppky et al. (2007) (squares), the 17-factor design from Belcher-Novosad and Ingram (2003) (circle) and those derived from the first 15–26 columns of the Plackett-Burman design (asterisks), on the one hand, and our designs (reference line), on the other hand.

The differences in PIC_5 values between our best designs and those of Loeppky et al. (2007) and the designs constructed by taking the first 20–26 factors from the Plackett-Burman design are smaller than 0.3 %. Figure 2b shows the difference in frequency of the J_3 -characteristic of 12 between the best of the competing designs and our best designs. If this difference is positive for a certain number of factors, our design outperforms the alternative design. If the difference is negative, our design is inferior. The figure shows that our designs are better than those of Loeppky et al. (2007) for 20–22 factors, worse for 23 and 24 factors, and equally good for 25 and 26 factors. The designs derived from the Plackett-Burman design are inferior to ours for 20–24 factors and equally good for 25 and 26 factors. We recommend our designs for 20–22 factors, those of Loeppky et al. (2007) for 23 and 24 factors and designs from any of the three approaches for 25 and 26 factors.

Bulutoglu and Ryan (2015) reported $A_3, \dots, A_6$ values from a heuristic search for good 28-run designs in terms of G_2 -aberration. Our designs involving 20–24 factors turn out to have slightly higher A_3 values (difference at most 0.49) and slightly lower A_4 values (difference at most 1.15) than theirs. For 25–27 factors, our best designs have the same A_3 and A_4 values as those of Bulutoglu and Ryan (2015). Generally, our best designs are more desirable than

Table 7: Selected 28-run designs with 15–19 factors. For all designs in the table, $F_3(28, 20) = (0, 0)$ and $F_4(28) = 0$.

ID	A_3	$F_3(12)$	$F_3(4)$	A_4	$F_4(20)$	$F_4(12)$	$F_4(4)$	PEC_5	PIC_5
15.1.1	14.02	29	426	65.57	21	168	1176	0.99767	0.87134
15.x	15.98	41	414	57.41	0	181	1184	1	0.87081
16.1.1	17.31	36	524	87.43	28	224	1568	0.99725	0.87075
16.x	20.73	57	503	74.69	0	230	1590	1	0.86631
17.1.1	23.18	57	623	105.06	28	262	2090	0.99628	0.86704
17.x	25.96	74	606	96.57	0	294	2086	1	0.86342
18.1.3	29.71	80	736	128.73	28	322	2710	0.99568	0.86324
18.x	31.67	92	724	123.35	0	373	2687	1	0.86184
19.1.2	36.59	103	866	159.10	28	406	3442	0.99579	0.86107
19.3.1	36.76	104	865	158.78	28	404	3444	0.99596	0.86078
19.x	37.90	111	858	156.00	0	471	3405	1	0.86102

those of these authors in terms of the GA4 classifier, because this was the primary criterion in our heuristic search.

5.2.3 Results for 15–19 factors

Our top list for 15–19 factors includes 32 designs. In supplementary tables, we show the GA6 classifier of the designs along with their PEC_5 and PIC_5 values. Properties for a selection of the designs are shown in Table 7. Initially, designs were included in Table 7 only if they have the best PIC_5 value of all top-3 designs, or if they have the best PIC_5 value among the designs that perform best in terms of the GA4 classifier. This procedure did not result in designs with a PEC_5 value of one. For this reason, we included the additional designs 15.x, 16.x, 17.x, 18.x and 19.x, which are projections from design 20.x, in Table 6. While these designs are neither optimal in terms of the GA4 classifier nor in terms of the PIC_5 value, they are good designs with 15–19 factors and with $PEC_5 = 1$.

Each of the designs in Table 6 has a PIC_5 value greater than 86% and a PEC_5 value greater than 99.5%. Designs with a given number of factors only show very small differences in PEC_5 and PIC_5 values. Therefore, whenever a practitioner anticipates that at most five factors will be active, we recommend the designs with $PEC_5 = 1$ over designs with a slightly higher PIC_5 value and $PEC_5 < 1$. If more than five factors are expected to be active, we

recommend designs that are best in terms of the GA4 classifier (i.e., designs whose labels end with a 1).

The best GWLPs identified by Bulutoglu and Ryan (2015) for 15- and 18-factor designs differ slightly from ours. The A_3 values for our best 15- and 18-factor designs are 1.31 and 0.32 units larger than those of Bulutoglu and Ryan (2015), while our A_4 values are 6.86 and 0.33 units smaller. For 16, 17, and 19 factors, our designs and the designs provided by Bulutoglu and Ryan (2015) have equal A_3 and A_4 values.

The designs derived from the 28-run Plackett-Burman design are inferior to ours. The difference in PIC_5 value is more than 1% in a few cases. The PIC_5 values of our best designs in 15–19 factors differ at most 0.44% from those of Loeppky et al. (2007), so that both their designs and ours are good options when no more than 5 factors are expected to be active.

Figure 2a shows the differences in frequency of the J_3 -characteristic of 12 between the designs of Loeppky et al. (2007), those constructed from the first 15–19 columns of the Plackett-Burman design and the single 17-factor design from Belcher-Novosad and Ingram (2003), on the one hand, and our designs, on the other hand. Clearly, our best designs in terms of the GA4 classifier are better than the competing designs in terms of these frequencies. Therefore, when more than five factors are expected to be active, identifying active factors using a main-effects model should be done using our designs.

6 Designs with 32 runs

In this section, we discuss designs with 32 runs involving 3–31 factors. Catalogs of all non-isomorphic strength-2 designs with 3–7 factors as well as all strength-3 designs for all possible numbers of factors (i.e., up to 16 factors) are available from the literature. The numbers of these designs are shown in Table 8. The most useful 5- and 6-factor designs are the full factorial 2^5 design, the replicated regular strength-4 half-fraction of this design, and the regular strength-5 half fraction of the 2^6 design. The most useful 7-factor designs are those with the highest D -efficiency for the interaction model reported by Eendebak and Schoen

Table 8: Numbers of non-isomorphic 32-run n -factor designs

n	Strength 2	Strength 3	n	Strength 3
3	5	1	10	32
4	19	3	11	22
5	491	5	12	23
6	266,217	10	13	12
7	530,469,996	17	14	10
8		33	15	5
9		34	16	5

(2017). For up to 16 factors, minimum G -aberration and minimum G_2 -aberration designs have a strength of 3. However, all strength-3 designs in 7–16 factors have correlations of 0.5 or 1 among the two-factor interaction contrast vectors.

The run size of 32 allows for special design constructions that work only if the run size is a power of 2. Xu and Wong (2007) constructed nonregular designs based on quaternary linear codes. Their 32-run designs for $n \geq 17$ factors all have a generalized resolution of at most 3.5. Consequently, these designs have at least one correlation of at least 0.5 between a main effect contrast vector and a two-factor interaction contrast vector. Lin et al. (2012) constructed 32-run designs by concatenating two 16-run designs. For $n \geq 17$ factors, their designs also have a generalized resolution of at most 3.5. In contrast, the designs we present below for 7–31 factors all have a generalized resolution of 3.75, meaning that their J_3 -characteristics are at most 8 and the correlations between main effect contrast vectors and two-factor interaction contrast vectors are at most 0.25. In addition, their J_4 -characteristics are also at most 8. This implies that none of the two-factor interactions is severely aliased with any main effect or with another two-factor interaction.

6.1 Enumerating designs with resolution 3.75

Kharagani and Tayfeh-Rezaie (2013) show that the Paley Hadamard matrix of order 32 is the only type-3 Hadamard matrix of this order. Normalizing this matrix and dropping the column of ones results in a 31-factor design with J_3 - and J_4 -characteristics of at most 8.

It turns out that all 31-factor designs based on the type-3 Hadamard matrix of order 32 are isomorphic. Therefore, up to isomorphism, the minimum G -aberration design in 31 factors is unique and has a generalized resolution of 3.75. Combined with the fact that no strength-3 32-run designs exist for more than 16 factors, this implies that any minimum G -aberration design with 17–31 factors must also have a generalized resolution of 3.75.

Our first attempt to find minimum G -aberration designs with 17–30 factors started from the set of 530,469,996 7-factor designs enumerated by Eendebak and Schoen (2017). Of these designs, 119,594,116 have a generalized resolution of 3.75. We extended a stratified sample of these designs by one extra factor; the stratification was based on the frequency of the J_3 -characteristic of 8. Based on that extension exercise, we estimated the number of 8-factor designs with a generalized resolution of 3.75 to be about 83 billion. As a result, it is computationally infeasible to identify minimum G -aberration designs with 17–30 factors by a complete bottom-up enumeration.

Designs with a generalized resolution of 3.75 are particularly attractive if their maximum J_4 -characteristic is 8. In our second attempt to enumerate good 32-run designs, we started from the set of 119,594,116 7-factor designs with a generalized resolution of 3.75. Of these designs, 24,156,320 designs have a maximum J_4 -characteristic of 8. Based on an extension of a sample of these designs, we estimate the number of designs with J_3 - and J_4 -characteristics of at most 8 to be more than 3.9 billion. Therefore, it is not feasible to completely enumerate the resolution 3.75 designs with J_4 -characteristics of at most 8 and to identify the minimum G -aberration designs in this subclass.

The designs that we did collect with generalized resolution 3.75 are all based on the unique 31-factor design with this resolution. We applied both a projection approach and an enumeration based on a particular 7-factor projection. In the projection approach, we studied all possible projections of the 31-factor design onto 26–30 factors, 100,000 randomly chosen projections onto 9–25 factors, and all possible projections onto 7 and 8 factors.

The enumeration approach we used employed the algorithm of Schoen et al. (2010), which

Table 9: Numbers of non-isomorphic 32-run n -factor designs obtained from the lexicographically smallest 7-factor projection of the unique 31-factor design with generalized resolution 3.75.

n	Designs	n	Designs
8	1,331	20	182,095
9	108,589	21	95,401
10	354,426	22	43,371
11	28,488	23	16,965
12	40,961	24	5,661
13	121,206	25	1,591
14	273,842	26	367
15	464,073	27	69
16	584,077	28	11
17	561,426	29	1
18	443,224	30	1
19	303,509	31	1

extends arrays that are in lexicographically minimal form column by column. The algorithm rejects extensions that are not in lexicographically minimal form. This way, the algorithm guarantees that only non-isomorphic designs are enumerated. To guarantee that a 7-factor design can be extended to the 31-factor design, we should start from its lexicographically smallest projection onto 7 factors. To this end, we reduced all possible projections onto 7 factors to their lexicographically minimal form, selected the one that was lexicographically smallest and extended that design. During the column-by-column extension, we removed all designs with J_4 -characteristics larger than 8, both because these designs are less attractive and because this limited the computational burden. Table 9 shows the numbers of non-isomorphic designs obtained by the extension procedure. These numbers small enough to allow for a detailed exploration of the designs.

6.2 Results and comparisons with benchmark designs

In the supplementary materials, we list the best designs with generalized resolution 3.75 found by the enumeration and projection approaches and we show the designs' GA6 classifiers, their $A_3, \dots, A_6$ values and their PIC_5 values. Table 10 shows the frequencies of J_3 -

Table 10: Selected designs with 32 runs, generalized resolution 3.75 and 7–31 factors. For all designs in the table, $F_3(32, 24, 16) = F_4(32, 24, 16) = (0, 0, 0)$.

ID	$F_3(8)$	$F_4(8)$	PIC ₅	ID	$F_3(8)$	$F_4(8)$	PIC ₅
7.1.1 _p	13	18	0.91691	19.1.1 _e	520	2148	0.88784
8.1.1 _p	22	40	0.90937	19.2.1 _e	521	2134	0.88799
8.2.1 _p	23	37	0.91022	20.1.1 _e	615	2679	0.88748
9.1.1 _e	35	80	0.89912	21.1.1 _e	720	3306	0.88724
9.2.2 _p	37	66	0.90591	22.1.1 _{ep}	837	4041	0.88686
10.1.1 _e	51	121	0.90304	23.1.1 _{ep}	966	4894	0.88652
11.1.1 _e	75	180	0.90093	24.1.1 _{ep}	1108	5870	0.88621
12.1.1 _p	107	277	0.89553	25.1.1 _{ep}	1263	6984	0.88596
13.1.1 _p	144	396	0.89310	26.1.1 _{ep}	1430	8250	0.88583
14.1.1 _e	187	545	0.89239	27.1.1 _{ep}	1612	9684	0.88565
14.1.1 _p	185	555	0.89234	28.1.1 _{ep}	1807	11297	0.88558
15.1.1 _e	237	744	0.89117	29.1.1 _{ep}	2016	13104	0.88555
16.1.1 _e	294	1006	0.88985	30.1.1 _{ep}	2240	15120	0.88555
17.1.1 _e	360	1314	0.88923	31.1.1 _{ep}	2480	17360	0.88555
18.1.1 _e	436	1684	0.88856				

and J_4 -characteristics of 8 and the PIC₅ values for a selection of the designs. The designs in Table 10 have the smallest G -aberration among all collected designs, or the best PIC₅ value. All designs have the maximum possible rank of 32 for the interaction model matrix, a PEC₅ value equal to 1, and a PIC₅ value larger than 0.88. So, on average, the estimation of interaction models based on five-factor projections is precise. The subscripts e and p in the table indicate whether the designs were obtained by means of the enumeration or by means of the projection approach, respectively.

The 7-factor design in Table 10, labeled 7.1.1, is the best 7-factor design we discovered using our projection and enumeration approaches. An exploration of the complete catalog for 7-factor designs, however, shows that the design with the best D -efficiency for the interaction model has a PIC₅ value of 0.9556, $F_3(8) = 5$ and an $F_4(8) = 14$. This design, which cannot be obtained from the 31-factor Paley design and was reported by Eendebak and Schoen (2017), outperforms design 7.1.1.

The designs with 8–16 factors in Table 10 involve a smaller amount of complete aliasing of two-factor interactions and a better rank of the interaction model matrix than the minimum

G -aberration strength-3 designs recommended by Schoen and Mee (2012). Their designs with 10–16 factors have 1, 3, 6, 10, 14, 21 and 28 J_4 -characteristics of 32, respectively, while their 8- and 9-factor designs have 12 and 24 J_4 -characteristics of 16, respectively and no J_4 -characteristics equalling 32. Our new designs do not involve any J_4 -characteristic larger than 8. So, $F_4(32, 16) = (0, 0)$ for these designs. In addition, the minimum G -aberration designs with 8 and 11–15 factors in Schoen and Mee (2012) do not have the maximum possible rank for the interaction model matrix. In conclusion, these strength-3 designs are less suitable for identifying active two-factor interactions than the strength-2 designs reported in Table 10.

The 32-run 31-factor Plackett-Burman design is a regular resolution-3 design. Using the first n columns of this 31-factor design results in designs with a generalized resolution of 3 only, and a PIC_5 value of at most 0.53. We therefore do not recommend using the 32-run Plackett-Burman design as a starting point for creating 32-run designs.

Finally, the generalized resolution of all designs for 17–31 factors in Table 10, which equals 3.75, is better than the generalized resolution of at most 3.5 for the designs of Xu and Wong (2007) and Lin et al. (2012).

7 Designs with 36 runs

The numbers of non-isomorphic 36-run designs for 3, 4, and 5 factors are 5, 15 and 1,242, respectively (Evangelaras et al., 2007), while there are 3,352,529 non-isomorphic 6-factor designs (Eendebak, 2016). Complete catalogs for larger numbers of factors do not exist. In this section, we identify and characterize all top-3 designs with 4–18 factors in terms of the GA4 classifier. Our list of designs includes the minimum G -aberration designs with up to 18 factors.

We used the 12 Hadamard matrices considered by Evangelaras et al. (2006) and Angelopoulos and Koukouvinos (2007), which include the 36-run Plackett-Burman design, together with the Williamson-type Hadamard matrix from Sloane (2015b) to determine an upper limit for the J_3 characteristics in minimum G -aberration designs. The 13 Hadamard

Table 11: Numbers of non-isomorphic 36-run designs with a generalized resolution of $3 \frac{8}{9}$

n	$F_4(36, 28, 20, 12, 4)$	
	$(0,0,0,a,b)$	General
4	2	4
5	26	38
6	3,810	6,105
7	2,213,023	4,303,133
8	128,165,961	391,128,970
9	123,305,732	1,620,091,182
10	1,033,893	
11	8,848	
12	75	
13	36	
14	14	
15	4	
16	2	
17	2	
18	1	
19	0	

matrices give rise to 20 classes of 35-factor Hadamard designs that differ according to the GA4 classifier. The Plackett-Burman design is the best design in terms of this classifier, with $F_3(12, 4) = (1190, 5355)$ and $F_4(12, 4) = (9520, 42840)$. So, a 36-run minimum G -aberration design must have J_3 -characteristics of at most 12, and, therefore, a generalized resolution of at least $3 \frac{2}{3}$.

7.1 Enumerating designs with resolution $3 \frac{8}{9}$

We obtained designs with 4–18 factors using a procedure similar to the one we used for 28-run designs with 3–14 factors. More specifically, we enumerated designs with J_3 -characteristics of 4 only, and so a generalized resolution of $3 \frac{8}{9}$. However, the numbers of non-isomorphic designs are so large that we had to impose more stringent restrictions on the enumeration than for the 28-run case. The rightmost column of Table 11 shows the numbers of non-isomorphic designs with a generalized resolution of $3 \frac{8}{9}$ for up to 9 factors. It is computationally infeasible to extend all 1,620,091,182 9-factor designs. For this reason, in our enumeration, we

Table 12: Selected designs with 36 runs and 7 – 18 factors. For all designs in the table, $F_3(36, 28, 20, 12) = (0, 0, 0, 0)$ and $F_4(36, 28, 20) = (0, 0, 0)$.

ID	A_3	$F_3(4)$	A_4	$F_4(12, 4)$		PIC_5
7.1.1	0.43	35	0.43	0	35	0.96354
8.1.2	0.69	56	1.75	9	61	0.94922
9.1.1	1.04	84	3.63	21	105	0.94501
10.1.1	1.48	120	6.44	39	171	0.94245
11.1.1	2.04	165	13.16	92	238	0.92425
11.3.1	2.04	165	13.56	96	234	0.92466
12.1.1	2.72	220	24.38	185	310	0.92044
13.1.1	3.53	286	36.88	284	431	0.91767
14.1.1	4.49	364	51.86	400	601	0.91739
15.1.1	5.62	455	70.78	546	819	0.91734
16.1.1	6.91	560	94.37	728	1092	0.91734
17.1.1	8.40	680	123.41	952	1428	0.91734
18.1.1	10.07	816	158.67	1224	1836	0.91734

focus on the subclass of designs with J_4 -characteristics of at most 12. There were 123,305,732 such designs with 9 factors, and we extended these designs column by column, each time limiting our attention to designs with generalized resolution $3^{8/9}$ and J_4 -characteristics of at most 12. This extension procedure resulted in a single 18-factor design. Any extension of this design with an extra column has one J_3 characteristic of 28. As this exceeds the upper bound of 12 for J_3 established previously, these extensions do not minimize the G -aberration.

Our focus on designs with J_3 characteristics of 4 ensures that we are able to identify designs with the best possible F_3 vector and A_3 value. The restriction to designs with J_4 -characteristics of at most 12 ensures that our designs have the best possible F_4 vector. However, there may be designs with better A_4 values. So, our restricted enumeration procedure guarantees that the minimum G -aberration designs are included in our list, while minimum G_2 -aberration designs may not be included.

7.2 Results and comparisons with benchmark designs

For 4–18 factors, our complete listing includes 77 groups of designs that are best, second best or third best in terms of the GA4 classifier and which differ according to the GA6 classifier.

All these designs (other than the 4-factor design) permit estimation of the interaction model involving any subset of five factors. Hence, $\text{PEC}_5 = 1$ for each of the designs. From each of the 77 groups of designs, we retained the one with the best PIC_5 value.

In the supplementary materials, we show the designs' GA6 classifiers, their $A_3, \dots, A_6$ values and their PIC_5 values. Table 12 shows some of these properties for selected designs. For each number of factors, the table includes the designs with the best PIC_5 values we found among the top-3 designs and the designs with the best PIC_5 values among those that are optimal in terms of the GA4 classifier. All listed designs have the maximum possible rank of the interaction model matrix.

The minimum G -aberration 7-factor design in Table 12 was identified by Eendebak and Schoen (2017) as the orthogonal design that permits estimation of the 7-factor interaction model with a D -efficiency of 0.9022. Bulutoglu and Ryan (2015) identified minimum G_2 -aberration designs with up to 9 factors. Our minimum G -aberration designs have the same A_3, A_4, A_5 and A_6 values. For 8 and 9 factors, multiple designs exist with these values, and we identified the unique minimum G -aberration 8- and 9-factor design.

For 10 factors and more, Bulutoglu and Ryan (2015) conducted a heuristic search among all designs with a generalized resolution of $3 \frac{8}{9}$. Our search was also restricted to designs with this generalized resolution. However, our approach ensures that we identified minimum G -aberration designs. The designs with 10–16 factors obtained by Bulutoglu and Ryan (2015) all have J_4 -characteristics of 20 so that there are pairs of two-factor interaction contrasts with an absolute correlation of $5/9$. We were able to find designs without J_4 -characteristics of 20. Moreover, our 10- and 11-factor designs have smaller A_4 values. However, despite the absence of J_4 -characteristics of 20, our designs involving 12–16 factors have poorer A_4 values than those of Bulutoglu and Ryan (2015). For 10–16 factor designs, the differences in A_4 values between the designs of Bulutoglu and Ryan (2015) and ours equal 0.89, 0.30, -1.68 , -2.97 , -4.05 , -5.34 , -7.11 , respectively. For 17 and 18 factors, our results allow us to confirm that the designs obtained by Bulutoglu and Ryan (2015) are minimum G -aberration

designs.

Designs constructed by taking the first 7–18 factors from the 36-run Plackett-Burman design are inferior to the designs listed in Table 12. They all have a generalized resolution of $3^{2/3}$, as opposed to $3^{8/9}$ for the listed designs. Also, the PIC_5 values are 1.4–5.6% smaller than those of the designs listed here.

For 19–35 factors, minimum G -aberration designs with 36 runs are unknown. We established that minimum G -aberration designs with 36 runs must have J_3 -characteristics of at most 12. We restricted our enumeration to designs with J_3 -characteristics of at most 4 and J_4 -characteristics of at most 12 for computational reasons. There exist no such designs with more than 18 factors. However, it is possible that there are 36-run designs in more than 18 factors with a generalized resolution of $3^{8/9}$ and J_4 -characteristics larger than 12.

As for the 28-run designs with more than 14 factors, we could search for good 36-run designs with 19–34 factors by studying projections from the 35-factor designs with J_3 -characteristics of at most 12. However, unlike in the 28-run case, the available knowledge about Hadamard matrices of order 36 is far from complete, and Hadamard matrices from which 35-factor designs with J_3 -characteristics of at most 4 can be constructed may exist. Therefore, any search based on the few known Hadamard matrices of order 36 is likely to be suboptimal. We therefore do not pursue this issue.

8 Discussion

In this paper, we describe the results of a complete search for the best 24-run designs with up to 23 factors in terms of the GA4 classifier (a simplified version of the CFV on which the G -aberration criterion is based) and the mixed A4 classifier (which is related to the G_2 -aberration criterion), and for the best 28-run designs with up to 14 factors. We also found the best 36-run designs in terms of the GA4 classifier with up to 18 factors. For 28-run designs involving 15–26 factors, we searched for the good projections from 12 attractive 27-factor designs. For 32-run designs with up to 31 factors, we performed a heuristic search

involving a partial enumeration as well as projections from the unique 31-factor design with a generalized resolution of 3.75. Our work resulted in a substantial number of designs that outperform those from the literature.

We discussed a partial enumeration procedure for orthogonal designs with run sizes 28, 32 and 36 and a predetermined generalized resolution of $3^{6/7}$, 3.75 and $3^{8/9}$, respectively. For the cases of 28 and 36 runs, we started from the unique three-factor design with a J_3 characteristic of 4. For the 32-run case, we started with a specific 7-factor projection of the 31-factor design with a generalized resolution of 3.75. Each time a factor column has been added, we retained the resulting design only if it did not involve any J_3 -characteristics greater than 4 (for 28-run and 36-run designs) or 8 (for 32-run designs) and if it was in lexicographically minimal form. For the 32- and 36-run designs, we imposed an additional constraint on the J_4 -characteristics to limit the computational burden. For the 28- and 36-run cases, this procedure allowed us to find optimal designs with up to 14 and 18 factors, respectively. These maximum numbers of factors were hitherto unknown. For the 32-run case, we knew a priori that the partial enumeration procedure would be able to generate desirable designs with up to 30 factors and with generalized resolution 3.75. All our 32-run designs have the attractive feature that the J_3 - and J_4 -characteristics are at most 8, so that all main effect and two-factor interaction contrast vectors have low correlations. Our 32-run designs with 17–30 factors all have a higher generalized resolution than the designs available in the literature, but we cannot guarantee that the designs are minimum G -aberration designs.

Much has been said in this paper about incomplete searches for optimal designs. We pointed out that a substantial portion of the literature on designs derived from Hadamard matrices ignores the fact that several non-isomorphic orthogonal designs can be constructed from a given Hadamard matrix. At the same time, there exist N -run designs with fewer than $N-1$ factors that cannot be obtained from $(N-1)$ -factor designs by projection. However, our own search among all possible 24-run designs was also incomplete because, for computational

reasons, we only evaluated all 24-run designs in terms of the GA4 and mixed A4 classifiers, and not in terms of the PEC_5 and PIC_5 values, for instance. Also, our searches for good 28-run designs with 15–26 factors and 32-run designs with 8–30 factors involved a projection approach starting with 12 good 27-factor designs and a single good 31-factor design. In the only other published study of the complete catalog of 24-run designs, Bulutoglu and Ryan (2015) consider the G_2 -aberration criterion, and, unlike us, do not evaluate the designs in terms of model-based criteria. For 28-run designs with 15–26 factors, their approach, like ours, was heuristic. For 36-run designs with 10–18 factors, their approach was heuristic, while ours was conclusive for the G -aberration criterion.

While we think that the criteria we used are sensible, other criteria may lead to different designs being optimal. We encourage future research concerning designs' capacity to fit models with all main effects and subsets of g two-factor interactions. For $g = 1, 2, \dots, 0.5n(n-1)$, an information capacity vector Li and Nachtsheim (2000) containing average D -efficiencies can be constructed, and designs can be compared based on this vector. However, this will be computationally very demanding.

Supplementary material

Supplementary sections.pdf: Details on the construction of different orthogonal designs from a given Hadamard matrix and a discussion of 24-run orthogonal designs from the literature.

Supplementary tables.zip: Tables with $A_3, \dots, A_6$ values, $F_3, \dots, F_6$ vectors, PEC_5 and PIC_5 values of all top 3 designs in terms of the GA4 or mixed A4 classifiers.

Designs.zip: ASCII files with all top 3 designs in terms of the GA4 and mixed A4 classifiers.

Acknowledgements

This research was financially supported by the Flemish Foundation for Scientific Research FWO. The authors are grateful to Dursun Bulutoglu and Kenneth Ryan for providing details concerning their results.

References

- Angelopoulos, P., Evangalaras, H., Koukouvinos, C., and Lappas, E. (2007). An effective step-down algorithm for the construction and the identification of nonisomorphic orthogonal arrays. *Metrika*, 66:139–149.
- Angelopoulos, P. and Koukouvinos, C. (2007). Maximum estimation capacity projection designs from Hadamard matrices with 32, 36 and 40 runs. *Statistics and Probability Letters*, 77:220–229.
- Belcher-Novosad, S. and Ingram, D. (2003). Identifying minimum G -aberration designs from Hadamard matrices of order 28. *Journal of the Arkansas Academy of Sciences*, 57:202–205.
- Bulutoglu, D. A. and Kaziska, D. M. (2009). A counterexample to Beder’s conjectures about Hadamard matrices. *Journal of Statistical Planning and Inference*, 139:3381–3383.
- Bulutoglu, D. A. and Margot, F. (2008). Classification of orthogonal arrays by integer programming. *Journal of Statistical Planning and Inference*, 138:654–666.
- Bulutoglu, D. A. and Ryan, K. J. (2015). Algorithms for finding generalized minimum aberration designs. *Journal of Complexity*, 31:577–589.
- Chen, J., Sun, D. X., and Wu, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *International Statistical Review*, 61:131–145.
- 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). A library of orthogonal arrays. <http://pietereendebak.nl/oapage/>.
- Eendebak, P. T. and Schoen, E. D. (2017). Two-level designs to estimate all main effects and two-factor interactions. *Technometrics*, to appear.
- Evangelaras, H., Georgiou, S., and Koukouvinos, C. (2004). Evaluation of inequivalent projections of Hadamard matrices of order 24. *Metrika*, 59:51–73.
- Evangelaras, H., Koukouvinos, C., and Lappas, E. (2007). Further contributions to non-isomorphic two level orthogonal arrays. *Journal of Statistical Planning and Inference*, 137:2080–2086.
- Evangelaras, H., Koukouvinos, C., and Mylona, K. (2006). Projection properties of Hadamard matrices of order 36 obtained from Paley’s constructions. *Metrika*, 64:351–359.
- Hedayat, A. S., Sloane, N. J. A., and Stufken, J. (1999). *Orthogonal arrays: Theory and Applications*. Springer, New York NY, USA.
- Ingram, D. and Tang, B. (2005). Minimum G aberration design construction and design tables for 24 runs. *Journal of Quality Technology*, 37:101–114.
- Ito, N., Leon, J. S., and Longyear, J. Q. (1981). Classification of 3-(24,12,5) designs and 24-dimensional Hadamard matrices. *Journal of Combinatorial Theory*, A 31:66–93.
- Kharagani, H. and Tayfeh-Rezaie, B. (2010). On the classification of Hadamard matrices of order 32. *Journal of Combinatorial Designs*, 18:328–336.
- Kharagani, H. and Tayfeh-Rezaie, B. (2013). Hadamard matrices order 32. *Journal of Combinatorial Designs*, 21:212–221.

- Kimura, H. (1989). New Hadamard matrix of order 24. *Graphs and Combinatorics*, 5:235–242.
- Kimura, H. (1994a). Characterization of Hadamard matrices of order 28 with Hall sets. *Discrete Mathematics*, 128:257–268.
- Kimura, H. (1994b). Classification of Hadamard matrices of order 28. *Discrete Mathematics*, 133:171–180.
- Kimura, H. and Ohmori, O. (1986). Construction of Hadamard matrices of order 28. *Graphs and Combinatorics*, 2:247–257.
- Li, W. and Nachtsheim, C. J. (2000). Model-robust factorial designs. *Technometrics*, 42:345–352.
- Li, Y., Deng, L. Y., and Tang, B. (2004). Design catalog based on minimum G -aberration. *Journal of Statistical Planning and Inference*, 124:219–230.
- Lin, C. D., Sitter, R. R., and Tang, B. (2012). Creating catalogues of two-level nonregular fractional factorial designs based on the criteria of generalized aberration. *Journal of Statistical Planning and Inference*, 142:445–456.
- 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.
- Plackett, R. L. and Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, 33:305–325.

- Schoen, E. D., Eendebak, P. T., and Nguyen, M. V. M. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs*, 18:123–140.
- Schoen, E. D. and Mee, R. W. (2012). Two-level designs of strength 3 and up to 48 runs. *Journal of the Royal Statistical Society Series C*, 61:163–174.
- Sloane, N. J. A. (2015a). A library of Hadamard matrices. [http://www. neil-sloane.com/Hadamard/](http://www.neilsloane.com/Hadamard/).
- Sloane, N. J. A. (2015b). A library of orthogonal arrays. [http://www. neilsloane. com/oadir/](http://www.neilsloane.com/oadir/).
- Sloane, N. J. A. (2015c). The on-line encyclopedia of integer sequences. <http://oeis.org>.
- Sun, D. X., Li, W., and Ye, K. Q. (2008). Algorithmic construction of catalogs of non-isomorphic two-level orthogonal designs for economic run sizes. *Statistics and Applications*, 6:141–155.
- Tang, B. and Deng, L. Y. (1999). Minimum G_2 -aberration for nonregular fractional factorial designs. *Annals of Statistics*, 27:1914–1926.
- Vijayan, K. (1976). Hadamard matrices and submatrices. *Journal of the Australian Mathematical Society*, 22:469–475.
- Wu, C. F. J. and Hamada, M. S. (2009). *Experiments: planning, analysis, and parameter design optimization*. 2nd edition, Wiley, New York, NY, USA.
- Xu, H. and Wong, A. (2007). Two-level nonregular designs from quaternary linear codes. *Statistica Sinica*, 17:1191–1213.
- Yamamoto, S., Fujii, Y., Hyodo, Y., and Yumiba, H. (1995). Classification of orthogonal 24-run 2^5 factorial designs derivable from saturated two-symbol orthogonal arrays of strength 2, size 24 and index 6. *SUT Journal of Mathematics*, 31:39–53.

Yumiba, H., Hyodo, Y., and Yamamoto, S. (1997). Classification of two-symbol orthogonal arrays of size 24, strength 2, 6 constraints and index 6 derivable from saturated orthogonal arrays having 23 constraints. *SUT Journal of Mathematics*, 33:47–63.