Saturated Second-Order Two-Level Designs: An Empirical Approach

Randall D. Tobias
Senior Research Statistician
SAS Institute, Inc.
Box 8000, Cary, North Carolina, 27511.

Abstract

Computer methods are used to explore saturated designs which provide for optimal estimation of main effects and interactions between two-level factors. A series of designs is thus discovered, related to a known series but better for $k > 6$. Also, a relationship is discovered between two different classes designs which should be fruitful for future research.

Keywords: Fractional factorial design, screening design, optimal design

1 Introduction

In recent years, the application of statistical methods of experimentation to problems in industrial quality control has led to a resurgence of interest in large factorial designs. Almost exclusively, orthogonal arrays are employed, most often of the 2^k variety. For collecting information on main effects alone, designs based on Hadamaard matrices are relatively abundant, being available for any number of runs which is a multiple of 4. However, efficient second-order designs, which give information on two-factor interactions as well, are few and far between. Regular resolution V fractions are available, but this series provides, for example, a design for 7 factors in 64 runs, while there are less than half that number (29) of parameters in the model. When a resolution V design is too large, one can use a design of resolution IV: this will provide information on main effects which is unconfounded with interactions, but the interactions will be confounded with one another.

The approach taken in this paper is to explore second-order designs with a minimal number of runs. Thus, for k factors we are interested in designs of size $1+k+k(k-1)/2$. We look for designs which provide optimal estimation of main effects as well as *all* two-factor interactions. Being saturated, there will be no independent estimate of the nominal level of noise against which to compare observed effects. But in practise most of the interactions will be insignificant, and some method of obtaining a pooled estimate of error (either *ad hoc* or rigorously, as in Box and Meyer (1986)) should be effective for analysis.

Rechtschaffner (1967) gave a general method for obtaining a saturated fraction for any number of factors k. The designs in this series are constructed in terms of the following subsets of all 2^k possible combinations of k +'s and −'s:

$$2^k_i \quad = \quad \{x \in 2^k : x \text{ contains } i \text{ +'s}\}.$$

Then Rechtschaffner's designs can be concisely written as

$$R_k \quad = \quad 2^k_1 \cup 2^k_{k-2} \cup 2^k_k. \tag{1}$$

Srivastava and Chopra (1971) showed that for $4 \leq k \leq 6$, R_k is *balanced* in the sense that the variance matrix of the estimates is invariant under any permutation of the factors; and that within the class of balanced designs it is in fact A-optimal. However, it's not known how good the designs R_k are in general, ie. for $k > 6$. Besides this, there has been a fair amount of work on small designs for *response surface* experiments (where quadratic terms as well as linear terms are included), but little of it is directly applicable to the construction of two-level designs in particular. For example, Draper and Lin (1990) have proposed selecting columns of a Plackett-Burman design for use as the factorial portion of a central composite design, but they do not study how well these designs fare by themselves as second-order two-level designs.

The primary tool for the present research is the computer. By applying an optimal design search procedure to the problem and exploring the structure of the results, a series of saturated second-order two-level designs is discovered, related to R_k but better for $k > 6$. The designs in this series are fully efficient up to 7 factors, and moderately efficient up to at least 12 factors. Also, an interesting relationship is turned up between optimal second-order designs and certain balanced incomplete block designs.

Section 2 discusses the general issue of the rôle of computers in design research. Section 3 exhibits the use of computational facilities to generate optimal designs and explore their structure. In section 4 we use these techniques to discover the new class of saturated second-order designs, and in section 5 the relationship between factorial designs and block designs is explored. While these designs are interesting and useful in themselves, it is important that they are discovered *via* optimal design search techniques, rather than through traditional combinatorics.

2 Combinatorial Design *versus* "Optimal" Design

The practise of theoretical design has not traditionally been one of those areas of applied mathematics where well-posed questions lead directly to well-defined answers. Instead, combinatorial arrangements are typically first proposed as designs on more or less obscure and intuitive grounds; various precisely defined notions of design optimality might be employed, but almost exclusively to *confirm* that a design is good, rather than to *generate* the design in the first place. But fortuitous combinatorial arrangements are not always available: when they are not, classical design theory leaves the practitioner out in the cold.

On the other hand, there are several algorithms for searching for designs optimal according to a precise efficiency criterion: the DETMAX algorithm of Mitchell (1974a) is standard; Cook and Nachtsheim (1980) compare several. Typically, they are used to select from a **given** set of candidate points a set of points of a **given** size which optimize a **given** criterion—for example, the determinant of the information matrix $X'X$ with respect to a **given** linear model. While such algorithms are often implemented in packages for experimental design, they don't seem to be in common use, and all those **given**'s in the last sentence indicate why this might be the case. *Algorithmically optimal* designs are good only for precisely defined situations: there is no guarantee they will be good even for other optimality criteria. (In fact, while search algorithms invariably find efficient designs, they may fail to find *the* precisely optimal one even for the given criterion.) By contrast, classical designs, with their high degree of symmetry, tend to be efficient for a variety of reasons.

Thus, algorithmic optimality has by no means made classical combinatorial design theory obsolete, though it does provide answers in many situations for which there are no standard designs. Part of the purpose of this paper is to point out that optimal design search programs can be used in searching for combinatorial designs—that is, they can be used as *generators* of arrangements which are likely to have a high degree of combinatorial structure. This point of view results in an *empirical* approach to theoretical design, based on searching for structure in the results of an optimal design search. The term *experimental mathematics* has been coined to refer to similarly computer-dependent studies in fractal geometry; and indeed the search for structure in combinatorial arrangements is much like the search for rules of non-linear dynamics which account for the beauty of the Mandelbrot set (see Peitgen and Richter (1986)).

Use of an optimal search program to explore the general structure of a class of designs is reported in Mitchell (1974b), where small first-order two-level designs are studied. Mitchell used the DETMAX algorithm to generate optimal first-order designs for up to 9 factors, and found that the best design can apparently always be constructed

Table 1: Optimal Saturated Second-order Design for 4 Factors

A	B	C	D
$-$	$-$	$-$	$-$
$-$	$-$	$-$	$+$
$-$	$-$	$+$	$-$
$-$	$+$	$-$	$+$
$-$	$+$	$+$	$-$
$-$	$+$	$+$	$+$
$+$	$-$	$-$	$+$
$+$	$-$	$+$	$-$
$+$	$-$	$+$	$+$
$+$	$+$	$-$	$-$
$+$	$+$	$+$	$+$

from a near-by orthogonal design by adding or deleting 1 or 2 points. The Plackett-Burman/Hadamard matrix designs provide orthogonal first order designs for all practical design sizes $N = 4t$. This approach is not very useful in the case of a second-order model, however, where orthogonal designs are few and far between.

3 Exploring Design Structure

The smallest non-trivial saturated second-order design is for 4 factors in 11 runs. (For 2 factors the full 2^2 is required to estimate the parameters. For 3 factors, there are 7 parameters in the model, so that 7 of the 8 different 2^3 points are required: clearly, any 7 will do.) In searching for combinatorial structure, an obvious first step is to look for some pattern in the actual combinations of $+$'s and $-$'s of the optimal design. Table 1 shows the 11 runs selected by an optimal design search procedure (see the appendix for details on computer facilities.)

No patterns are immediately apparent. However, the efficiency of the design is invariant to changes in the *signs* of the factors, and this may well mask any pattern. Another option is to examine the $X'X$ matrix itself, since it is upon this that the efficiency of the design depends. Table 2 shows the $X'X$ matrix for this design.

The first row and column of $X'X$ correspond to the intercept. If factors C and D are multiplied by -1, the information matrix for the main effects will be of the form $aI + bJ$, where I is the identity and J the all-ones matrix—an auspicious combination! Once

Table 2: $X'X$ matrix for O_4

		A	B	C	D	AB	AC	AD	BC	BD	CD
	11	-1	-1	1	1	-1	1	1	1	1	-1
A	-1	11	-1	1	1	-1	1	1	-3	-3	3
B	-1	-1	11	1	1	-1	-3	-3	1	1	3
C	1	1	1	11	-1	-3	-1	3	-1	3	1
D	1	1	1	-1	11	-3	3	-1	3	-1	1
AB	-1	-1	-1	-3	-3	11	1	1	1	1	3
AC	1	1	-3	-1	3	1	11	-1	-1	3	1
AD	1	1	-3	3	-1	1	-1	11	3	-1	1
BC	1	-3	1	-1	3	1	-1	3	11	-1	1
BD	1	-3	1	3	-1	1	3	-1	-1	11	1
CD	-1	3	3	1	1	3	1	1	1	1	11

this is done, a pattern does emerge: Table 3 shows the transformed runs of the design. It is composed of all combinations with one, two, or four +'s. Symbolically,

$$O_4 \quad = \quad 2^4_1 \cup 2^4_2 \cup 2^4_4. \tag{2}$$

For 5 factors the optimal design is known, the 2^{5-1}_V fraction. Can it also be articulated in terms of the subsets 2^5_i? Yes: constructing the design as $E = A \times B \times C \times D$, where the RHS ranges over all 2^4 possible combinations, it is clear that the 2^{5-1}_V fraction is comprised of all those runs in the full 2^5 candidate set with an odd number of +'s. That is,

$$O_5 \quad = \quad 2^{5-1}_V \quad = \quad 2^5_1 \cup 2^5_3 \cup 2^5_5. \tag{3}$$

Finally, for 6 factors an optimal design procedure is used again, and examination of the $X'X$ matrix is also again required in order to scale the factor levels appropriately. When this is done, the optimal saturated second-order design for 6 factors has the form

$$O_6 \quad = \quad 2^6_1 \cup 2^6_4 \cup 2^6_6. \tag{4}$$

Thus, techniques which can be used to explore the combinatorial structure of a given arrangement include examining the actual points as well as the information matrix, possibly under permutations of the factors and/or their respective levels. It's also handy to be able easily to compute and organize the design by any functions which may appear important (in this case, the number of +'s). Using these techniques, we have "found" the balanced combinatorial structure of Rechtschaffner's series of designs R_k (1) for k=4, 5, and 6.

Table 3: Optimal Design for 4 Factors, Transformed

A	B	$-C$	$-D$
+	−	−	−
−	+	−	−
−	−	+	−
−	−	−	+
−	−	+	+
−	+	−	+
−	+	+	−
+	−	−	+
+	−	+	−
+	+	−	−
+	+	+	+

Table 4: Optimal Design for 7 Factors

		A	B	C	D	E	F	G
	29	5	5	1	1	1	1	1
A	5	29	5	1	1	1	1	1
B	5	5	29	1	1	1	1	1
C	1	1	1	29	5	5	5	5
D	1	1	1	5	29	5	5	5
E	1	1	1	5	5	29	5	5
F	1	1	1	5	5	5	29	5
G	1	1	1	5	5	5	5	29

4 A Series of Efficient Saturated Second-order Designs

In the rest of the paper we will continue to refer to Rechtschaffner's design for k factors as R_k, and the optimal design for k factors as O_k. So far, we have had $R_k = O_k$. For $k = 7$ factors, however, an optimal design search finds a design which is better than R_k. After appropriate permutations of the factor levels, the part of the $X'X$ matrix of this design corresponding to the intercept and main effects displays a *partially balanced* structure, shown in Table 4.

It isn't possible to make this $X'X$ matrix invariant to permutations of the factors, so the design isn't balanced in the sense of Srivastava and Chopra (1971) and in particular it can't be written as a union of 2_i^7 terms. It does appear that the $k+1$ runs $2_1^k \cup 2_k^k$

are still in the design; but instead of 2^k_{k-2}, the "left-over" part now takes the form

$$O_7 - \left(2^7_1 \cup 2^7_7\right) \quad = \quad \left\{2^7_5 \; - \; (++) \oplus 2^5_3\right\} \;\cup\; \left\{(++) \oplus 2^5_2\right\}, \tag{5}$$

where $(++) \oplus 2^5_3$, for example, means the runs for which the first two factors are both + and the rest are chosen from 2^5_3; $\{2^7_5 \; - \; (++) \oplus 2^5_3\}$ means all runs in 2^7_5 except those for which the first two factors are both +.

Note that by exchanging +'s and −'s we have $2^5_2 = -2^5_3$. But 2^5_3 was the left-over part for the 5 factor design. This observation inspires a guess at the general structure.

$$D_k \quad = \quad 2^k_1 \cup 2^k_k \cup A_k \tag{6}$$

where the left-over part A_k is *recursively* defined as 2^k_2 for $k = 2, 3$, and for $k > 3$,

$$A_k \quad = \quad \left\{2^k_{k-2} \; - \; (++) \oplus 2^{k-2}_{k-4}\right\} \;\cup\; \{(++) \oplus -A_{k-2}\}. \tag{7}$$

From (5) it's clear that $O_7 = D_7$. Furthermore, if $A_{k-2} = 2^{k-2}_2 = -2^{k-2}_{k-4}$ then $A_k = 2^k_{k-2}$, so that D_k matches R_k for $4 \leq k \leq 6$ also. D_k is also of the right size, as the next result shows.

Theorem 1 *For all $k \geq 3$, $|D_k| \; = \; 1 \; + \; k(k+1)/2$.*

Proof:

First note that

$$\begin{aligned} |2^k_1| &= k \\ |2^k_k| &= 1 \\ |2^k_2| &= k(k-1)/2 \qquad (8) \\ &= |2^k_{k-2}|. \qquad (9) \end{aligned}$$

Then $|D_k| = 1 + k + |A_k|$ and the theorem will be proved if we can show that

$$|A_k| \quad = \quad k(k-1)/2. \tag{10}$$

Now, by (8), (10) clearly holds for $2 \leq k \leq 6$, where $A_k = 2^k_2$ or $A_k = 2^k_{k-2}$. Then

$$\begin{aligned} |A_k| &= |2^k_{k-2}| - |2^{k-2}_{k-4}| + |A_{k-2}| \\ &= \tfrac{k(k-1)}{2} - \tfrac{(k-2)(k-3)}{2} + |A_{k-2}| \qquad \text{by (9),} \\ &= \tfrac{k(k-1)}{2} - \tfrac{(k-2)(k-3)}{2} + \tfrac{(k-2)(k-3)}{2} \qquad \text{by induction,} \\ &= \tfrac{k(k-1)}{2}. \end{aligned}$$

Table 5: Efficiencies of D_k Relative to Optimal Design

Number of Factors	Number of Runs	Efficiency Factors D	A	G
4	11	100%	100%	100%
5	16	100%	100%	100%
6	22	100%	100%	100%
7	29	100%	100%	100%
8	37	92%	94%	119%
9	46	84%	81%	103%
10	56	76%	75%	111%
11	67	60%	47%	67%
12	79	61%	59%	115%

From the theorem, $|D_3| = 7$, so that by our remarks at the beginning of Section 3 D_3 is also optimal. Thus, the series of designs D_k recursively defined by (6) and (7) gives optimal saturated second-order designs for all $3 \leq k \leq 7$. How about for $k > 7$? Tables 5 and 6 show the efficiency of D_k relative to O_k and R_k, respectively, for $4 \leq k \leq 12$.

Table 5 shows that the series of designs D_k is not absolutely optimal for $k > 7$, but it also shows that an absolutely optimal design is not available for these situations: D_k has worse D-efficiency but better G-efficiency than the D-optimal design in all cases except for 11 factors. On the other hand, Table 6 shows that D_k is better than R_k for $k > 7$, in terms of all three efficiency measures.

The design $D_7 = O_7$ is not balanced in the sense of Srivastava and Chopra (1971): the $X'X$ matrix is not invariant to arbitrary permutations of the factors. However, it is *partially* balanced in the sense of Kuwada (1988): the factors can be partitioned so that the $X'X$ matrix is invariant to permutations within each partition. Unfortunately, examination of O_8 failed to reveal any such structure, except that it contained $2^k_1 \cup 2^k_k$.

5 Factorial Designs and Block Designs

O_{11} appears to be somewhat unique: it is the one design in Table 5 where the G-efficiency for O_k is (much) better than the G-efficiency of D_k, and also the one point where the ratio of D-efficiencies is not monotone. It has a very interesting structure; the runs and factors in the "left-over" part $O_{11} - (2^{11}_1 \cup 2^{11}_{11})$ are associated with the blocks and treatments, respectively, of a certain balanced incomplete block design (BIBD) for

Table 6: Efficiencies of D_k Relative to Rechtschaffner's Designs

Number of Factors	Number of Runs	Efficiency Factors		
		D	A	G
4	11	100%	100%	100%
5	16	100%	100%	100%
6	22	100%	100%	100%
7	29	108%	111%	104%
8	37	112%	115%	102%
9	46	120%	124%	105%
10	56	125%	127%	103%
11	67	132%	133%	105%
12	79	136%	135%	103%

11 treatments in 55 blocks of size 6. In particular, the runs of the left-over part of the factorial design are given by $2N-1$, where N is the block$\times$treatment incidence matrix for the BIBD. That is, if treatment j occurs in block i, then factor j is set to 1 for run i; otherwise, it is set to -1. Note that 2^k_i is also associated in this way with the collection of all combinations of k objects taken i at a time, a trivial BIBD. Thus, Rechschaffner's designs can also be derived from BIBD's in this manner.

Moreover, the condition of balance in the factorial design translates directly to a balance condition on the associated block design, in the following way: the factorial runs are balanced with resolution $t+1$ if and only if in the associated block design all j-plets of treatments occur equally often together in blocks, for $1 \leq j \leq t$. Such block designs are called *t-designs* or *tactical configurations* (Raghavarao, 1971), and in fact it's not difficult to see that any resolution $t+1$ balanced two-level fractional factorial design is equivalent to a unique t-design (not necessarily with equal block sizes). In particular, balanced two-level factorial designs of resolution V are equivalent to 4-designs. For example, there exists a 4-design for 11 treatments in 66 blocks of size 6; the 78-run factorial design obtained from this by appending $2^{11}_1 \cup 2^{11}_{11}$ to the transformed block$\times$treatment incidence matrix, as above, appears to be optimal.

The relationship between t-designs and balanced fractional factorial designs has been noted before (Chakravarti, 1961), but since 4-designs usually require many blocks, the resulting factorial designs are usually too large to be of much use. On the other hand, the BIBD which is associated with O_{11} is *not* a 4-design. It is, however, *nearly* 4-balanced: each j-plet of treatments occurs together in blocks either λ_j or λ_j+1 times, for $3 \leq j \leq 5$. This near t-plet balance promises to be useful for constructing efficient factorial designs, and will be the subject of further research.

Table 7: Simulated Data for D_7

A	B	C	D	E	F	G	Y	A	B	C	D	E	F	G	Y
-	-	-	-	-	-	+	4.96	+	-	+	-	+	+	+	15.96
-	-	-	-	-	+	-	2.90	+	-	+	+	-	+	+	15.60
-	-	-	-	+	-	-	2.76	+	-	+	+	+	-	+	15.30
-	-	-	+	-	-	-	2.78	+	-	+	+	+	+	-	5.85
-	-	+	-	-	-	-	3.07	+	+	+	+	-	-	-	15.73
-	+	-	-	-	-	-	5.50	+	+	+	-	+	-	-	15.18
+	-	-	-	-	-	-	3.48	+	+	+	-	-	+	-	15.62
+	+	+	+	+	+	+	25.79	+	+	+	-	-	-	+	23.97
-	-	+	+	+	+	+	5.57	+	+	-	+	+	-	-	15.49
-	+	-	+	+	+	+	7.18	+	+	-	+	-	+	-	16.47
-	+	+	-	+	+	+	5.56	+	+	-	+	-	-	+	26.84
-	+	+	+	-	+	+	7.45	+	+	-	-	-	+	+	23.87
-	+	+	+	+	-	+	6.36	+	+	-	-	+	-	+	23.44
-	+	+	+	+	+	-	6.70	+	+	-	-	+	+	-	13.62
+	-	-	+	+	+	+	15.03								

6 Example

Table 7 shows the optimal design D_7 along with simulated response values. The response was generated as

$$Y = 10 + 5A + 3B + 3G + 2AB + 2AG + \epsilon$$

where $\epsilon \sim N(0, 1)$.

Table 8 gives the largest of the estimated effects for a least-squares fit of the second-order model to this data. The true effects stand out clearly, and since the design allows for detection of any main effect or interaction, the experimenter may be confident that the significant factors have been isolated.

7 Discussion

In table 5 and in the discussion above, we have referred to the result O_k of the optimal design search for a minimal second-order design for k factors as "optimal," without qualification. This is somewhat optimistic. Specifically, the reference design used as

Table 8: Estimated Regression Coefficients

Term	Estimate
Intercept	10.06
A	4.89
B	3.11
G	2.82
AG	2.20
AB	2.08
D	0.42
BD	0.38
BG	-0.29
CD	-0.24
$\vdots$	$\vdots$

the optimal design for k factors in table 5 is the best design of 200 tries, using a partially random initial design (see Galil and Kiefer (1980)) and the "Modified Federov" algorithm of Cook and Nachtsheim (1980); the complete search took hours for the larger k on a very fast workstation (HP 720). And even this was not enough: the efficiencies of the resulting designs improved noticeably when the points $2^k_1 \cup 2^k_k$ were included by default. (The recurrence of these $k+1$ points in optimal second-order designs has been dubbed the "haystack effect": 2^k_k is the tip of the stack and 2^k_1 its base.*) Even so, the best design for 8 factors was found in only 1 of the 200 tries. Thus, it is highly likely that O_k is in fact sub-optimal for $k \geq 9$ (except for $k = 11$) although we can have confidence that it is fairly close to the optimum.

We have shown how computer experimental design facilities can be used to search for designs of the more traditional combinatorial type. By way of example, we have used standard experimental design and data manipulation tools to discover construction rules for a fairly efficient series of saturated second-order factorial designs. We have also discovered a relationship between balanced factorial designs and balanced block designs which should be fruitful for further research.

*Neil Sloane, personal communication.

8 References

Box, G.E.P. and Meyer, R.D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics* 28, pp. 11-18.

Chakravarti, I.M. (1961), "On Some Methods of Construction of Partially Balanced Arrays," *Annals of Mathematical Statistics*, 32, pp. 1181-1185.

Cook, R.D. and Nachtsheim, C.J. (1980). "A Comparison of Algorithms for Constructing Exact D-optimal Designs". *Technometrics* 22, pp. 315-324.

Draper, N.R., and Lin, D.K.J. (1990). "Small Response Surface Designs," *Technometrics*, 32:2, pp. 187-194.

Dykstra, O.Jr. (1971). "The Augmentation of Experimental Data to Maximize $|X'X|$". *Technometrics* 13, pp. 682-688.

Federov, V.V. (1972). *Theory of Optimal Experiments,* translated and edited by W.J. Studden and E.M. Klimko. Academic Press, New York.

Galil, Z. and Kiefer, J. (1980). "Time- and Space-saving Computer Methods, Related to Mitchell's DETMAX, for Finding D-Optimum Designs." *Technometrics* 22, pp. 301-313.

Kuwada, M. (1988). "A-optimal Partially Balanced Fractional $2^{m_1+m_2}$ Factorial Designs of Resolution V, with $4 \leq m_1 + m_2 \leq 6$." *Journal of Statistical Planning and Inference* 18, pp. 177-193.

Mitchell, T.J. (1974a). "An Algorithm for the Construction of D-optimal Experimental Designs". *Technometrics* 20, pp. 203-210.

Mitchell, T.J. (1974b). "Computer Construction of 'D-optimal' First-Order Designs". *Technometrics* 20, pp. 211-220.

Peitgen, H.O. and Richter, P.H. (1986) *The Beauty of Fractals.* Springer-Verlag, New York.

Raghavarao, D. (1971). *Constructions and Combinatorial Problems in Design of Experiments.* Wiley, New York.

Rechtschaffner, R.L. (1967). "Saturated Fractions of 2^n and 3^n Factorial Designs". *Technometrics* 9, pp. 569-575;

Srivastava, J.N. and Chopra, D.V. (1971). "Balanced Optimal 2^m Fractional Factorial Designs of Resolution V, $m \leq 6$." *Technometrics* 13, pp. 257-269.

A Computational Facilities

The commercial SAS® software package provides a variety of modules for statistical analysis. For generating optimal saturated second-order designs, the two relevant procedures are FACTEX and OPTEX, provided with SAS/QC® software. FACTEX is designed to construct regular q^k fractions from input model specifications, but it can also be used to create the full 2^k candidate set; the OPTEX procedure is used to select an optimal subset. For example, Table 9 shows a program using the two procedures to find a D-optimal saturated second-order design for 5 factors; the output is a listing of relative efficiencies of the designs found in several tries. Because of a certain amount of randomness in the starting design for the search, the procedure makes 10 tries by default. In 7 of these 10, the orthogonal 2_V^{5-1} design was found.

The efficiency factors for the various designs were computed using a ± 1 coding for the two-level factors. The design matrix X is then composed of a single column of 1's corresponding to the general mean effect, k columns for the main effects, and $k(k-1)/2$ columns for the cross-product effects. Let p be the number of independent variables in the linear model and N_D the number of points in the target design: for the saturated second-order designs which we consider, $p = N_D = 1 + k + k(k-1)/2$. Finally, let σ_M be the maximum standard error for prediction over the candidate points. Then the definitions for the efficiency factors are shown in Table 10. The D- and A-efficiencies are the relative number of runs (expressed as percents) required by a hypothetical orthogonal design to achieve the same $|X'X|$ and trace $\left((X'X)^{-1}\right)$, respectively. (See Mitchell (1974b).) The numerator in the G-efficiency is a lower bound on σ_M for the optimal design.

The SAS/IML® matrix programming language was also used to examine the information matrices of various designs under a variety of permutations of the factors and the factor levels, as well as to explore the balance properties of block designs associated with some of the second-order designs.

Table 9: SAS Program to Find a D-optimal 5 Factor Design

`PROC FACTEX;`	
`   FACTORS A B C D E;`	*Declare the factors.*
`   OUTPUT OUT=FULL;`	*Output the* 2^5 *to the dataset FULL.*
`PROC OPTEX;`	
`   MODEL A|B|C|D|E@2;`	*Declare a second-order model.*
`   GENERATE N=SATURATED;`	*Ask for a saturated design.*
`RUN;`	

```
                                                                   Prediction
 Design                                                              Standard
 Number     D-efficiency     A-efficiency     G-efficiency             Error
 ---------------------------------------------------------------------------
      1         100.0000         100.0000         100.0000            1.0000
      2         100.0000         100.0000         100.0000            1.0000
      3         100.0000         100.0000         100.0000            1.0000
      4         100.0000         100.0000         100.0000            1.0000
      5         100.0000         100.0000         100.0000            1.0000
      6         100.0000         100.0000         100.0000            1.0000
      7         100.0000         100.0000         100.0000            1.0000
      8          71.7594          46.5183          43.7595            1.4662
      9          71.7594          46.5183          43.7595            1.4662
     10          71.7594          45.7304          42.9547            1.4788
```

Table 10: Efficiency Factors

D	$100 \times \frac{\|X'X\|^{1/p}}{N_D}$
A	$100 \times \frac{p}{\mathrm{trace}(N_D \times (X'X)^{-1})}$
G	$100 \times \frac{\sqrt{p/N_D}}{\sigma_M}$

SAS, SAS/QC, and SAS/IML are registered trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.