SOME CONTRIBUTIONS TO DESIGN THEORY AND APPLICATIONS

A Thesis
Presented to
The Academic Faculty

by

Abhyuday Mandal

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

School of Industrial and Systems Engineering
Georgia Institute of Technology
August 2005

SOME CONTRIBUTIONS TO DESIGN THEORY AND APPLICATIONS

Approved by:

Dr. C. F. Jeff Wu, Advisor
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Jye-Chyi Lu
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Roshan Joseph Vengazhiyil
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Shabbir Ahmed
School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Cheng Zhu
Woodruff School of Mechanical Engineering and Coulter Department of Biomedical Engineering
Georgia Institute of Technology

Date Approved: 4 May 2005

To my mother,

Chhaya Mandal,

whose inspiration made me what I am today.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my appreciation to all who have influenced, stimulated, expedited, and warmly supported my work in various ways.

First and foremost, it is my immense pleasure to express my deep and sincere gratitude to my advisor, Professor C. F. Jeff Wu, for his guidance, assistance, encouragement, and hearty support in all phases of my doctoral program. He took care of me in every possible way even when he was in hospital. He has not only been my academic advisor, but also a great mentor for my graduate life.

I am thankful to Professor Rahul Mukerjee for his guidance and encouragement on my research work. His kind words and active support have inspired me to do research, starting from my undergraduate studies in Indian Statistical Institute.

I would like to thank Dr. Kjell Johnson, Dr. Derek Bingham and Dr. Kerby Shedden for having extended their support to my research. I am also thankful to Dr. Jye-chyi Lu, Dr. Roshan Joseph Vengazhiyil, Dr. Shabbir Ahmed and Dr. Cheng Zhu for serving on my dissertation committee and for their valuable comments and suggestions. I thank all my lab members who shared time, space and knowledge with me at Georgia Tech. They made my graduate days wonderful and enjoyable.

Last, but by no means the least, my heartful appreciation and gratitude goes to my family, especially my parents, for their support and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The thesis focuses on the development of statistical methodology in experimental design with applications in global optimization. It has four chapters. In the first chapter, a new criterion for design efficiency under model uncertainty is proposed and its properties studied. The second chapter is devoted to identification of optimal designs to distinguish between competing models using Hellinger distance. In the third chapter, aliasing patterns of the mixed-level fractional designs in the form of cross arrays are discussed. Design of experiment ideas are used in the last chapter to introduce a new global optimization technique.

Design of experiments is an important tool in many scientific investigations. A theoretically fundamental and practically important question is how to choose the best design among the available competing designs. The first two chapters of the thesis focus on this issue. Much efforts have been devoted to this problem in the literature. Nevertheless, most existing methods and results are limited to regular designs or two-level designs. There are only a few general results for nonregular and mixed-level designs. To tackle this problem, a new criterion of design efficiency under model uncertainty is studied with reference to both regular and nonregular fractions of general factorials. Cheng, Deng and Tang (2002) proposed a criterion for nonregular fractions of two-level factorials. Our research extends their work to general factorials including asymmetrical ones. The findings are reported in the first chapter. The criterion is expressed in terms of the departure of the design from being an orthogonal array of strength three or four. This approach is in the spirit of the criterion of estimation capacity. The findings are seen to be in agreement with those according to the generalized minimum aberration and minimum moment aberration criteria.

In the next chapter, the research is done on identification of optimal designs using Bayesian methods. This work is linked with response surface methodology, where the first step is to perform factor screening, followed by response surface exploration using

different experimental plans. Recently, Cheng and Wu (2001) proposed a new approach that aims to achieve both goals using one design. This methodology can lead to spurious identification of effects. Instead we propose new design and analysis methodology that aims to overcome these difficulties. The approach is Bayesian in nature and attempts to more directly incorporate the standard assumptions of industrial experiments into the design and analysis. In addition we use a Bayesian design criterion based on the priors for the analysis approach. This creates an integrated design and analysis framework. From an analysis standpoint, the aim is to identify a subset of predictors that best explain the data. From a design perspective, the goal is to identify the set of treatments that best facilitates the aim of the analysis. Thus we aim to identify the optimal design. To distinguish between competing models, the HD criterion is used, which is based on the pairwise Hellinger distance between predictive densities.

Two- and three-level symmetric factorial and fractional factorial designs are widely used in industrial experimentations and are discussed in detail in design of experiments textbooks. Mixed-level fractional factorial designs are also commonly used in industries but its aliasing relations have not been studied in full rigor. These designs take the form of a product array. In robust parameter designs they are called *cross arrays* (or *inner-outer array* in Taguchi's terminology). Aliasing patterns of mixed-level factorial designs are discussed in the third chapter. A rigorous study of mixed-level cross arrays gives a deeper insight into the estimation properties of the design.

In the last chapter, design of experiment ideas are used to introduce a new global optimization technique called $SELC$ (Sequential Elimination of Level Combinations), which is motivated by genetic algorithms but can find the optimum faster. Genetic Algorithm is a widely used optimization technique when searching for global optimum. Although popular, it is one of the slower optimization techniques. The SELC algorithm overcomes this limitation. It is a nice blend of design of experiment ideas and genetic algorithms, and it outperforms genetic algorithm in many cases. The two key features of the SELC algorithm, namely, forbidden array and weighted mutation, enhance the performance of its search procedure. Weighted mutation is the driving force of SELC. The process is

also enriched by a Bayesian method for identifying the important main effects and two-factor interactions, which is needed for the weighted mutation. This Bayesian method is similar to the one discussed in the second chapter. SELC starts with an orthogonal design that helps identifying important effects. Bayesian variable selection gives better insight into the problem. The follow-up runs are very flexible and data-driven; the weighted mutation uses sequential learning. This SELC method is useful in many real-life examples, ranging from computer experiments to compound identification. Illustration is given with the optimization of three functions, one of which is from Shekel's family. A real example on compound optimization from pharmaceutical industry is also given. Scientific knowledge about the problem is incorporated in choosing the starting design and constructing the forbidden array. A follow-up experiment demonstrates the success of the SELC method in identifying a rich set of compounds.

CHAPTER I

DESIGN EFFICIENCY UNDER MODEL UNCERTAINTY FOR NONREGULAR FRACTIONS OF GENERAL FACTORIALS

1.1 Introduction

Recently, Cheng, Deng and Tang (2002), hereafter abbreviated CDT, reported results on design efficiency, under model uncertainty, for nonregular fractions of two-level factorials. Their criterion concerns models that include the general mean, all main effects and a selection of two-factor interactions (2fi's) and, in the absence of prior knowledge on which 2fi's are active, it considers the average performance of a design over all possible models with the same number of 2fi's. As discussed by these authors, this approach is in the spirit of the criterion of estimation capacity introduced by Sun (1993), and studied by Cheng, Steinberg and Sun (1999), Cheng and Mukerjee (1998, 2001) and Mukerjee, Chan and Fang (2000) for regular fractions.

The present article aims at extending the work of CDT on design efficiency to general factorials including the asymmetrical ones. This calls for a substantial modification of their mathematical techniques since, unlike in the two-level case, each factorial effect may no longer be represented by a single treatment contrast. A Kronecker calculus for factorial arrangements facilitates the formulation of the model matrices as well as the derivation of the key results. The main results are presented in the next Section where we also indicate the connection with the departure of the design from being an orthogonal array of various strengths. This, in turn, entails a link with the generalized minimum aberration (GMA) criterion (Tang and Deng (1999); Xu and Wu (2001)). The present criterion is applied to 18-run nonregular fractions of 2×3^3 and 2×3^4 factorials. The findings are seen to be in agreement with those according to the GMA criterion and the minimum moment aberration

(MMA) criterion (Xu (2003)). Proofs appear in the appendix.

1.2 Main results

Suppose there are m factors $F_1, \ldots, F_m$ at $s_1, \ldots, s_m (\geq 2)$ levels respectively. For $1 \leq j \leq m$, the levels of F_j are coded as $0, 1, \ldots, s_j - 1$. Consider a possibly nonregular fraction or design consisting of the treatment combinations $a_{i1}a_{i2}\ldots a_{im}, 1 \leq i \leq N$, where $a_{ij} \in \{0, 1, \ldots, s_j - 1\}$ for every i, j, i.e.

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ & & \vdots & \\ a_{N1} & a_{N2} & \ldots & a_{Nm} \end{pmatrix} \tag{1}$$

Throughout, N is fixed and it is supposed that these N treatment combinations, when written as rows, form an orthogonal array (OA) of strength two.

We assume the absence of interactions involving three or more factors. Note that altogether there are $W(= m(m-1)/2)$ 2fi's. For $1 \leq w \leq W$, let $H(w)$ be the collection of all sets of w 2fi's. For any $h \in H(w)$, let $M(h)$ be the model consisting of only the general mean, all main effects and the w 2fi's in h, and $X(h)$ be the model matrix under $M(h)$. The matrix $X(h)$ consists of blocks of columns that correspond to the general mean and the factorial effects in $M(h)$. The blocks of columns associated with the 2fi's are related to those associated with the main effects via Kronecker products. A detailed expression for $X(h)$ appears in (A.5) in the appendix. As usual, it is assumed that the observational errors are homoscedastic and uncorrelated.

Under $M(h)$, the D-criterion aims at maximizing $det\{X(h)^T X(h)\}$. If one wishes to include w 2fi's in the model, but has no prior knowledge on which w should be included, then it makes sense to consider the average of $det\{X(h)^T X(h)\}$ over all $h \in H(w)$. This is the D_w-criterion of CDT. However, it is difficult to handle this criterion algebraically. On the other hand, minimization of $tr[\{X(h)^T X(h)\}^2]$ is a good surrogate for the maximization of $det\{X(h)^T X(h)\}$. This happens because $tr\{X(h)^T X(h)\}$ is the same for all designs under consideration; cf. (A.5), (A.7)$-$(A.10) and (A.18) in the appendix. Consequently, a large

$det\{X(h)^TX(h)\}$ is typically accompanied by a small $tr[\{X(h)^TX(h)\}^2]$ since both occur when the eigenvalues of $X(h)^TX(h)$ are close to one another (the same argument shows that minimization of $tr[\{X(h)^TX(h)\}^2]$ would be a good surrogate also if one worked with the A- or E-criteria.). Hence following CDT, we consider the design criterion

$$E_w \;=\; \binom{W}{w}^{-1} \sum_{h\in H(w)} tr[\{X(h)^TX(h)\}^2], \tag{2}$$

and aim at studying designs that keep E_w small for every w, especially for smaller values of w which are more relevant under effect sparsity.

Lemma 1, presented below and proved in the appendix, gives an expression for E_w which is useful both algebraically and numerically. Some more notation will help. For any distinct α,β,γ $(1 \le \alpha,\beta,\gamma \le m)$, let $n_{jkl}^{(\alpha,\beta,\gamma)}$ be the number of times the factors F_α, F_β and F_γ appear at levels j,k and l respectively among the N treatment combinations in the design, and define

$$\phi(jkl) \;=\; s_\alpha s_\beta s_\gamma \sum\sum\sum \left(n_{jkl}^{(\alpha,\beta,\gamma)}\right)^2, \tag{3}$$

where the triple sum is over $0 \le j \le s_\alpha - 1$, $0 \le k \le s_\beta - 1$, $0 \le l \le s_\gamma - 1$. Similarly, for any distinct $\alpha,\beta,\gamma,\delta$ $(1 \le \alpha,\beta,\gamma,\delta \le m)$, define the quantities $n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}$, and hence $\phi(ijkl)$, exactly along the lines of (3). Let $\Delta(3)$ be the set of all ordered triplets jkl, where $1 \le j < k < l \le m$, and $\Delta(4)$ be the set of all ordered four-tuples $jklu$, where $1 \le j < k < l < u \le m$. Finally, in (2.4) below and the rest of this paper, a "constant" may depend on $w, N, m, s_1, \ldots, s_m$ but is the same for all designs.

Lemma 1 *For $1 \le w \le W$, with*

$$E_w^* = \sum_{jkl\in\Delta(3)} \left(6 + \frac{2(w-1)}{W-1}(s_j + s_k + s_l - 3m + 3)\right)\phi(jkl) + \frac{6(w-1)}{W-1} \sum_{jklu\in\Delta(4)} \phi(jklu), \tag{4}$$

$$E_w = constant + (w/W)E_w^*,\ 1 \le w \le W. \tag{5}$$

In view of (5), hereafter we consider the quantities E_w^*. By (4), for $3 \leq w \leq W$,

$$E_w^* = E_1^* + (w-1)(E_2^* - E_1^*), \tag{6}$$

a fact which is useful for computational purposes. Lemma 1 also helps in expressing E_w^*, and hence E_w, in terms of measures of the departure of the design from being represented by an OA of strength three or four. To that effect, some more notation is introduced.

For $1 \leq j \leq m$, let

$$V_j(0) = s_j^{-1} 1_j 1_j^T, \qquad V_j(1) = I_j - s_j^{-1} 1_j 1_j^T, \tag{7}$$

where 1_j is the $s_j \times 1$ vector with all elements unity and I_j is the identity matrix of order s_j. For any binary m-tuple $x = x_1 \ldots x_m$, define the matrix

$$V(x) = V_1(x_1) \otimes V_2(x_2) \otimes \ldots \otimes V_m(x_m), \tag{8}$$

where $\otimes$ denotes the Kronecker product. Let $\nu = \prod_{j=1}^m s_j$, and n be the $\nu \times 1$ vector whose elements represent the replication numbers of the ν treatment combinations in the design, arranged in the lexicographic order. Note that

$$\sum_{x \in \Omega} V(x) \;\; = \;\; I_{s_1} \otimes I_{s_2} \otimes \ldots \otimes I_{s_m}$$

where Ω is the set of all binary m-tuples. Hence

$$n^T n \;\; = \;\; n^T \Big(\sum_{x \in \Omega} V(x) \Big) n = \sum_{x \in \Omega} n^T \Big(V(x) \Big) n,$$

which gives an ANOVA decomposition of n, pretending that n is an observational vector rather than frequencies of level combinations in a design. Let

$$\Omega(g) = \Big\{ x : x \in \Omega, x \text{ has exactly } g \text{ coordinates that equal to } 1 \Big\},$$

then $\Omega = \Omega(0) \cup \Omega(1) \cup \ldots \cup \Omega(m)$, and

$$n^T n = \sum_{u=0}^{m} \sum_{x \in \Omega(g)} n^T V(x) n = \sum_{u=0}^{m} B_g,$$

where, for $1 \leq g \leq m$,

$$B_g = \sum_{x \in \Omega(g)} n^T V(x) n. \tag{9}$$

Clearly, $B_g \geq 0$ for every g, as the matrices $V(x)$ are nonnegative definite. Since the treatment combinations in the design form an OA of strength two, by (7) − (9), $B_1 = B_2 = 0$. Similarly, it can be seen that the design is represented by an OA of strength three if and only if $B_3 = 0$, and an OA of strength four if and only if, in addition, $B_4 = 0$. Hence, as argued by Fang, Ma and Mukerjee (2002) (see also Tang (2001)), B_3 is a natural measure of the departure of the design from being represented by an OA of strength three, whereas B_4 measures the additional departure of the design from an OA of strength four. It can also be seen that

$$B_g = \nu^{-1} N^2 A_g,\ 1 \leq g \leq m, \tag{10}$$

where $(A_1, \ldots, A_m)$ is the generalized wordlength pattern (GWP) of the design (Tang and Deng (1999); Xu and Wu (2001)). Theorem 1 below expresses E_w^* in terms of B_3, B_4 and the related quantities $B(jkl)$, where for $jkl \in \Delta(3)$, $B(jkl) = n^T V(x(jkl))n$, with $x(jkl)$ being the binary m-tuple that has 1 in the jth, kth and lth positions and 0 elsewhere. Again, $B(jkl)$ is nonnegative and equals zero if and only if the projection of the design onto the three factors F_j, F_k and F_l is an OA of strength three. The proof of Theorem 1 appears in the appendix.

Theorem 1 *For* $1 \leq w \leq W$,

$$\begin{aligned} E_w^* &= \textit{constant} \\ &\quad +6\nu \left[B_3 + \frac{w-1}{W-1} \left(B_4 - 2B_3 + \frac{1}{3} \sum_{jkl \in \Delta(3)} (s_j + s_k + s_l) B(jkl) \right) \right]. \end{aligned} \tag{11}$$

Remark 1. By (9),

$$B_3 = \sum_{jkl \in \Delta(3)} B(jkl). \tag{12}$$

Hence if $s_1 = \ldots = s_m (= s$, say) then (2.10) simplifies to

$$E_w^* = \text{constant} + 6\nu \left[\left(1 + \frac{w-1}{W-1}(s-2) \right) B_3 + \frac{w-1}{W-1} B_4 \right]. \tag{13}$$

The coefficient of B_3 in (13) is much larger than that of B_4, especially for relatively smaller values of w. Hence a design that sequentially minimizes $B_3, B_4, \ldots$ (recall that $B_1 = B_2 = 0$ for any design considered) should perform well under the criterion considered here. Therefore, strengthening the findings of CDT, from (9) it follows that for general symmetrical factorials a GMA design should have an edge over others under the present criterion as well.

Remark 2. For two-level factorials, by (9) and (13),

$$E_w^* = \text{constant} + 6N^2[A_3 + \{(w-1)/(W-1)\}A_4],$$

which, in conjunction with (5), is in agreement with CDT.

Remark 3. For asymmetrical factorials, by (9) and (11),

$$E_1^* = \text{constant} + 6N^2 A_3.$$

While the link between $E_w^*, w \geq 2$, and the GWP is less obvious, the numerical study in the next Section suggests that the GMA criterion tends to be in agreement with the present one.

Remark 4. Interestingly, $B(jkl)$ actually occurs in (11) only for $w \geq 2$ and not for $w = 1$. If two or more 2fi's are included in the model, then any two of them can potentially involve a common factor. Such common factors contribute to the term involving $B(jkl)$ in (11). The same happens with the coefficient of $\phi(jkl)$ in (4). Equation (A.21) in the appendix and the discussion preceding it make this explicit.

1.3 A numerical study

Table 7C.2 of Wu and Hamada (2000) shows an OA(18,$2^1 3^7$) of strength two, with 18 rows and 8 columns, where the first column has two symbols and the remaining columns have three symbols each. Consideration of the first column together with any three other columns of this array yields a nonregular fraction of a 2×3^3 factorial in 18 treatment combinations. Any such design, given by the first, jth, kth and lth columns, is denoted by $1jkl$ ($2 \leq j < k < l \leq 8$). For any of the 35 possible designs so obtained, a simple counting of degrees of freedom reveals that $X(h)^T X(h)$ is singular whenever the model involves five

or more 2fi's. Hence, we consider E_w^* only for $1 \leq w \leq 4$. It is seen that the collection of these 35 designs can be partitioned into six classes, as shown in Table 1, such that all designs in the same class have the same E_w^* for every w and also the same GWP. Table 2 shows $E_w^*, 1 \leq w \leq 4$, and the GWP against these six classes. Equations (4) and (5) facilitate these computations. From Table 2 , it is clear that, for every w, the ranking of designs according to E_w^* is precisely the same as that according to the GMA criterion. In fact, it can be seen that this ranking is also the same as that under the MMA criterion, with natural weights, as based on the first five moments.

The phenomenon of identical ranking of designs according to the E_w^* and the criteria of GMA and MMA continues to hold if one instead considers designs for a 2×3^4 factorial that arise in a similar manner from the OA(18,$2^1 3^7$) mentioned above. The details are omitted here. This suggests that even for asymmetrical factorials the latter two criteria are good surrogates for the present criterion which has a direct statistical meaning.

Table 1: Equivalent classes of designs for a 2×3^3 factorial arising from an OA(18,$2^1 3^7$)

Class	Designs
1	1248, 1258, 1367, 1458
2	1236, 1237, 1267
3	1234, 1235, 1246, 1247, 1256, 1257
4	1238, 1268, 1278
5	1345, 1346, 1347, 1348, 1356, 1357, 1358, 1368, 1378, 1456, 1457, 1467, 1468, 1478, 1567, 1568, 1578, 1678
6	1245

1.4 References

Cheng, C.S., Deng, L.Y. and Tang, B. (2002), "Generalized Minimum Aberration and Design Efficiency for Nonregular Fractional Factorial Designs," *Statist. Sinica* **12**, 991-1000.

Cheng, C.S. and Mukerjee, R. (1998), "Regular Fractional Factorial Designs with Minimum Aberration and Maximum Estimation Capacity," *Ann. Statist.* **26**, 2289-2300.

Table 2: Values of E_w^* and GWP for the six classes of designs

Class	E_1^*	E_2^*	E_3^*	E_4^*	GWP
1	8748.0	9525.6	10303.2	11080.8	(0, 0, 1/2, 3/2)
2	9720.0	10497.6	11275.2	12052.8	(0, 0, 1, 1)
3	10044.0	10735.2	11426.4	12117.6	(0, 0, 7/6, 5/6)
4	11016.0	11707.2	12398.4	13089.6	(0, 0, 5/3, 1/3)
5	11340.0	11944.8	12549.6	13154.4	(0, 0, 11/6, 1/6)
6	11664.0	12441.6	13219.2	13996.8	(0, 0, 2, 0)

Cheng, C.S. and Mukerjee, R. (2001), "Blocked Regular Fractional Factorial Designs with Maximum Estimation Capacity," *Ann. Statist.* **29**, 530-548.

Cheng, C.S., Steinberg, D.M. and Sun, D.X. (1999), "Minimum Aberration And Model Robustness For Two-Level Fractional Factorial Designs," *J. Roy. Statist. Soc. B* **61**, 85-93.

Cheng, S-W, and Ye, K (2004), "Geometric Isomorphism And Minimum Aberration For Factorial Designs With Quantitative Factors," *Ann. Statist.*, to appear.

Fang, K.T., Ma, C.X. and Mukerjee, R. (2002), "Uniformity In Fractional Factorials," In: *Monte Carlo and Quasi- Monte Carlo Methods 2000* (eds. K.T. Fang et al.), pp. 232-241, Springer-Verlag, Berlin.

Mukerjee, R. (1999), "On The Optimality Of Orthogonal Array Plus One Run Plans," *Ann. Statist.* **27**, 82-93.

Mukerjee, R., Chan, L.Y. and Fang, K.T. (2000), "Regular Fractions Of Mixed Factorials With Maximum Estimation Capacity," *Statist. Sinica* **10**, 1117-1132.

Sun, D.X. (1993), *Estimation Capacity and Related Topics in Experimental Designs.* Ph.D. thesis, University of Waterloo.

Tang, B. (2001), "Theory Of J-Characteristics For Fractional Factorial Designs And Projection Justification Of Minimum G_2-Aberration," *Biometrika* **88**, 401-407.

Tang, B. and Deng, L.Y. (1999), "Minimum G_2-Aberration For Nonregular Fractional Factorial Designs," *Ann. Statist.* **27**, 1914-1926.

Wu, C.F.J. and Hamada, M. (2000), *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley, New York.

Xu, H. (2003), "Minimum Moment Aberration For Nonregular Designs And Supersaturated Designs," *Statist. Sinica* **13**, 691-708.

Xu, H. and Wu, C.F.J. (2001), "Generalized Minimum Aberration For Asymmetrical Fractional Factorial Designs," *Ann. Statist.* **29**, 1066-1077.

CHAPTER II

BAYESIAN FACTOR SCREENING AND RESPONSE SURFACE DESIGNS

2.1 Introduction

The statistical design and analysis of experiments is an important tool in the investigation of processes. By varying factors of interest over level settings and performing experiment trials, insight into which models best explain the observed responses is gained. The related choices of the best experiment design and analysis approach are among the most fundamental issues facing an experimenter.

An important application of the design and analysis of experiments is response surface methodology (Wu and Hamada (2000)) (hereafter denoted RSM), which typically aims to "optimize" a process using a fitted model. RSM can be broadly described as consisting of two-stages: (i) screening; and (ii) response surface exploration. In the first stage, a relatively small first-order design (e.g., a design aimed at fitting a model with only linear terms - a first-order model) such as a 2^{n-k} fractional factorial design or a 2-level orthogonal array is performed. Using the responses from the first stage design, the important factors are identified and unimportant factors are screened out. Next, a second-order design (e.g., a central composite design) is performed using only the important factors. The responses from the second-order design are used to fit a linear model with linear and quadratic main effects and linear-by-linear interactions (i.e., a second-order model).

Recently, Cheng and Wu (2001) (hereafter denoted CW) proposed an approach for RSM that aimed to do both screening and response surface exploration with a single design, thereby saving experimentation time and run size. The approach is based on a two-stage analysis that employs screening, projection and response surface exploration. The analysis procedure can be summarized in the following steps: (i) perform factor screening by fitting a

linear model containing only linear main effects (a first-order model); and (ii) using the same data, fit a second-order model based on only the significant factors from step (i). In essence, the first stage screens out unimportant effects, then recycles the data to fit a second-order model on a projected design space. While simplifying analysis, this methodology implies some unstated, but important assumptions with respect to the underlying structure of the model. For instance, interactions are entertained only on the projected design space, thus one is assuming that all factors in significant interaction effects also have significant main effects. The assumption that a two-factor interaction can only be active if both main effects are active is called *strong heredity* (Chipman (1996)). In the event that strong heredity does not hold, the two-stage analysis can miss interactions, and may misspecify the response surface. Indeed, we demonstrate in the next section that this can lead to detection of spurious effects.

In this chapter, we propose a new design and analysis approach to overcome some of the limitations of the two-stage analysis approach. For analysis, we adapt the Bayesian variable selection approach of Chipman, Hamada and Wu (1997) to the screening/RSM situation. This approach will result in a more comprehensive search of the model space explaining the data. In addition we use a Bayesian design criterion (Binghan and Chipman (2002)), based on the priors for the analysis approach. This creates an integrated design and analysis framework.

The chapter is outlined as follows. In Section 2, we give a motivating example to illustrate the necessity of Bayesian framework. In the next two sections, the analysis and identification of optimal designs are discussed. Section 3 is devoted to the development of Bayesian framework and Section 4 contains the discussion on optimal designs based on the Hellinger distance (HD) criterion. Simulation studies are presented in Section 5. Summary and concluding remarks are given in Section 6.

2.2 Notation and Motivating Examples

In this section we explain why CW analysis have some limitations. Next, several examples are presented to illustrate some of challenges faced by the two-step analysis approach. In

the next section, we propose new methodology to address these difficulties.

We now investigate some of the implications of two-stage approach for factor screening and response surface exploration. Suppose that there are q factors in the experiment, and the true model relating the response to the design factors is

$$y = X_1\theta_1 + X_2\theta_2 + X_3\theta_3 + \epsilon \tag{14}$$

where θ_1, θ_2 and θ_3 are the vectors of coefficients for the main effects, quadratic effects and linear-by linear interactions, respectively, X_1, X_2 and X_3 are the corresponding model matrices, and $\epsilon \sim N(0, \sigma^2)$. CW follow the usual *effect sparsity* assumption that most effects are not active, and thus most of the coefficients are negligible. The first step of the two-stage procedure is to fit a first-order model (main effects only) and estimate θ_1 with the usual least-squares estimate of,

$$\hat{\theta_1} = (X_1'X_1)^{-1}X_1'y. \tag{15}$$

Being cognizant that quadratic and interaction terms are potentially important, the estimate of θ_1 has a bias of

$$B_1 = -(X_1'X_1)^{-1}X_1'(X_2\theta_2 + X_3\theta_3). \tag{16}$$

The amount of this bias will depend on the values of θ_2 and θ_3 as well as on X_1, X_2 and X_3 (i.e., the design).

The second stage of the procedure *projects* the original q factors onto a smaller factor space, consisting of only the q' factors identified as active in the first step. A second-order model is then fit to these q' factors only, using the same experiment data. The model matrices in (14) can be partitioned into those columns that do and do not correspond to terms in the second-order model that fit to the q' factors from the first step. That is, let $X_1 = (X_1^{(1)} : X_1^{(2)})$, $X_2 = (X_2^{(1)} : X_2^{(2)})$, and $X_3 = (X_3^{(1)} : X_3^{(2)})$, where $X_i^{(1)}$ are the model matrices for the second-order model based on the q' active factors and $X_i^{(2)}$ are the remaining columns from the corresponding model matrices. Thus, the second step of the two-stage procedure is to fit the second-order model

$$y = X_1^{(1)}\theta_1^{(1)} + X_2^{(1)}\theta_2^{(1)} + X_3^{(1)}\theta_3^{(1)} + \epsilon \tag{17}$$

which leads to the estimate of

$$\widehat{\theta_1^{(1)}} = \left(X^{(1)\prime}X^{(1)}\right)^{-1} X^{(1)\prime}y \tag{18}$$

where $X^{(1)} = \left(X_1^{(1)} : X_2^{(1)} : X_3^{(1)}\right)$. In this case, the bias associated with the second stage regression estimates are

$$B_2 = -\left(X^{(1)\prime}X^{(1)}\right)^{-1} X^{(1)\prime}X^{(2)}\theta^{(2)}. \tag{19}$$

as the true model is $y = X^{(1)}\theta^{(1)} + X^{(2)}\theta^{(2)} + \epsilon$. The amount of this bias will depend on values of $\theta^{(2)}$ as well as $X^{(1)}$ and $X^{(2)}$.

Based on equations (16), (19) and the two-stage procedure itself, we make the following observations:

(i) Interactions containing one or more factors without a significant parent cannot be identified. Only significant main effects survive stage 1, and terms surviving stage 1 are used to build effects in stage 2.

(ii) Quadratic and interaction effects can bias estimates in the first step, since X_2 and X_3 appear in the first-stage bias (3).

(iii) Interactions containing one or more factors without a significant parent can bias estimates in both steps. Bias arising in step 1 was already mentioned in (ii) above. Bias can occur in step 2 because such interactions will be contained in $X^{(2)}$ and thus appear in the second-stage bias (6).

The above arguments are slightly simplified, since they assume that the partition of data into $X^{(1)}$ and $X^{(2)}$ is a fixed choice. The two-stage nature of the CW algorithm implies that this partition is random. Thus the second-stage bias will actually involve an expectation over the choice of $X^{(1)}$ and $X^{(2)}$, instead of being conditional on this choice, as in (6). However, limitations (i) - (iii) still apply, since they can occur conditional on some stage 1 choices.

The reason for two stages in the CW strategy is to avoid simultaneously considering all possible subsets of effects. All subsets is computationally more intensive, and the scarcity

of runs implies that not all effects are simultaneously estimable. However, as pointed out above, this assumption comes with a cost: the possibility that exclusion of terms in the first stage can lead to exclusion of important terms at the second stage.

Of course, the severity of the bias in observations (ii) and (iii) will depend on the design and the magnitude of the unidentified effects. Indeed, one might conjecture that a design based on an orthogonal array might ease these concerns. Unfortunately, this is not always the case. In the remainder of this section, we present a few examples to demonstrate how the two-stage procedure can lead to a mis-specified model in some very simple cases.

Example 1

Consider the example in presented in CW, and first presented by Taguchi (1987). The experiment is a 27-run experiment (Taguchi (1987, p.423)) to study the PVC insulation for electric wire. There are 8 factors ($A - H$), each with 3 levels, and the design is given in Table 3. The aim of the endeavor is to identify the quadratic model that best explains the data in light of standard assumtions such as effect sparsity, effect hierarch and effect heredity. We will delay the analysis of Taguchi's data and proeceed with simulated data for illustrative purposes.

In what follows (and throughout this chapter), we follow CW and use the *linear-quadratic system* for coding linear and quadratic effects (Wu and Hamada (2000)) in order to reduce correlation among a factor's linear and quadratic components. The linear-quadratic coding is expressed as follows :

$$\begin{array}{cccc} factor & & linear & quadratic \\ 0 & & -1 & 1 \\ & \longrightarrow & & \\ 1 & & 0 & -2 \\ 2 & & 1 & 1. \end{array}$$

Although the true model is unknown, we can use the design in Table 3 to illustrate some difficulties with the two-stage approach. Suppose the true model for the response is

$$y_1 = \theta_A x_A + \theta_B x_B + \theta_{BE} x_{BE} + \epsilon,$$

Table 3: Design matrix and response data

y = PVC insulation data, y_1 = model 1, y_2 = model 2

run	A	B	C	D	E	F	G	H	J	y	y1	y2
1	0	0	0	0	0	0	0	0	0	5	-17	-13
2	0	0	0	0	1	1	1	1	1	2	-22	-10
3	0	0	0	0	2	2	2	2	2	8	-30	9
4	0	1	1	1	0	0	0	2	2	-15	-9	4
5	0	1	1	1	1	1	1	0	0	-6	-9	-11
6	0	1	1	1	2	2	2	1	1	-10	-10	0
7	0	2	2	2	0	0	0	1	1	-28	-4	12
8	0	2	2	2	1	1	1	2	2	-19	2	20
9	0	2	2	2	2	2	2	0	0	-23	7	-10
10	1	0	1	2	0	1	2	0	1	-13	-9	-10
11	1	0	1	2	1	2	0	1	2	-17	-13	3
12	1	0	1	2	2	0	1	2	0	-7	-19	-31
13	1	1	2	0	0	1	2	2	0	-23	-1	-1
14	1	1	2	0	1	2	0	0	1	-31	1	1
15	1	1	2	0	2	0	1	1	2	-23	2	18
16	1	2	0	1	0	1	2	1	2	-34	7	13
17	1	2	0	1	1	2	0	2	0	-37	11	-1
18	1	2	0	1	2	0	1	0	1	-29	19	9
19	2	0	2	1	0	2	1	0	2	-27	7	-4
20	2	0	2	1	1	0	2	1	0	-27	-3	-21
21	2	0	2	1	2	1	0	2	1	-30	-9	-10
22	2	1	0	2	0	2	1	2	1	-35	11	0
23	2	1	0	2	1	0	2	0	2	-35	11	14
24	2	1	0	2	2	1	0	1	0	-38	10	-20
25	2	2	1	0	0	2	1	1	0	-39	16	6
26	2	2	1	0	1	0	2	2	1	-40	20	12
27	2	2	1	0	2	1	0	0	2	-41	29	30

with $(\theta_A, \theta_B, \theta_{BE}) = (10, 12, 6)$ (all other coefficients in the linear model are zero) and $\epsilon \sim N(0, 1)$ error. Table 3 contains data simulated from this model under column y_1.

Table 4 shows the ANOVA table for the fit of the first-order model. Notice that not only are factors A and B identified as significant, so is factor H. Next, a second-order response surface is fit to the three active factors (A, B and H) (see Table 5), and no additional significant effects are identified. So, in this fairly simple case, the presence of a weak heredity interaction results in a spurious main effect being found and failure to identify the interaction.

Table 4: Least square estimates, t-statistics and p-values

Coefficients	Estimate	Std. Error	t value	Pr($>|t|$)
Intercept	1.85	1.57	1.18	0.2540
A	9.33	1.92	4.86	0.0001
B	9.44	1.92	4.92	0.0001
C	-0.78	1.92	-0.41	0.6906
D	-2.00	1.92	-1.04	0.3124
E	-1.00	1.92	-0.52	0.6094
F	-1.89	1.92	-0.98	0.3393
G	-3.28	1.92	-1.71	0.1062
H	-4.39	1.92	-2.29	0.0355
J	-1.28	1.92	-0.67	0.5149

Table 5: Least square estimates, t-statistics and p-values

Coefficients	Estimate	Std. Error	t value	Pr($>|t|$)
Intercept	1.85	1.46	1.27	0.2220
A	9.33	1.79	5.21	0.0001
B	9.44	1.79	5.28	0.0001
H	-4.39	1.79	-2.45	0.0253
A^2	1.04	1.03	1.00	0.3295
B^2	0.59	1.03	0.57	0.5737
C^2	1.87	1.03	1.81	0.0879
AB	1.00	2.19	0.46	0.6539
AH	3.75	2.19	1.71	0.1052
BH	2.75	2.19	1.26	0.2265

Example 2

Let us see another example. The data are given in Table 3 under column y_2. The true model is

$$y_2 = \theta_B x_B + \theta_D x_D + \theta_{BJ} x_{BJ} + \epsilon.$$

The first-order fit identifies factors B, D and J as significant and in the second-order fit, along with those factors, D^2 and DJ come out to be significant. As we see, the weak heredity interaction BJ gives spurious main effect J, quadratic effect D^2 and interaction DJ.

2.3 Design and Model Selection

Now that we have seen the existing approach can lead to incorrect identification of the model, we need to find a method to eliminate this problem. The Bayesian approach of Bingham and Chipman (2002) for design and analysis of screening experiments is taken here, with adaptations to the three-level response surface problem. Analysis is described first, since the design criterion is based on priors used for modelling.

2.3.1 Analysis

The model selection problem amounts to identifying a subset of predictors as active, and in this setting there are typically more parameters to estimate than unique treatments. The possible models will be labeled as $M_1, M_2, \ldots, M_K$. Priors for M_i and (β_i, σ) will be discussed later.

Here we propose stochastic variable selection, based on Gibbs sampler. We start with a given design and the corresponding responses. For the linear regression with normal errors,

$$y = X\theta + \sigma\epsilon, \qquad \epsilon \sim N(0,1) \tag{20}$$

where θ contains linear and quadratic main effects and linear-by-linear interaction effects. The Bayesian framework of Chipman, Hamada, and Wu (1997) approaches model selection as follows: Importance of effects is captured via an unobserved vector δ of zeros and ones where $\delta_i = I\{\theta_i \neq 0\}$. A normal mixture prior is used for the coefficients β :

$$f(\beta_i|\delta_i) = \begin{cases} N(0, \tau_i^2) & \text{if } \delta_i = 0 \\ N(0, (c_i\tau_i)^2) & \text{if } \delta_i = 1 \end{cases} \tag{21}$$

When $\delta_i = 0$, β_i has a high mass around zero and thereby, is not likely to have a large effect. On the other hand, when $\delta_i = 1$ a large value of c_i ensures that the variable is likely to have a large influence.

Not all models are are equally likely. Based on the assumptions of *effect sparsity, effect hierarchy* and *effect inheritance*, we can distinguish between the "likely" and "unlikely" models. Note that the commonly used independence prior, which implies that the importance of one factor is independent of that of another, is not very attractive as there are

quadratic main and linear-by-liner interaction effects. Instead, we have used hierarchical priors, motivated by Chipman (1996). Consider a simple example with three main effects, A, B and C, each having three levels. It is logical to think that the importance of the interaction effect AB will depend on the importance of main factors A and B only. Also, the quadratic effect of level A will less likely to be important if the linear effect of A is not important. This belief can be expressed in the prior for $\delta = (\delta_A, \delta_B, \delta_C, \delta_{A^2}, \delta_{B^2}, \delta_{C^2}, \delta_{AB}, \delta_{AC}, \delta_{BC})$ as follows:

$$\begin{aligned}
P(\delta) &= P(\delta_A, \delta_B, \delta_C, \delta_{A^2}, \delta_{B^2}, \delta_{C^2}, \delta_{AB}, \delta_{AC}, \delta_{BC}) \\
&= P(\delta_A, \delta_B, \delta_C) P(\delta_{A^2}, \delta_{B^2}, \delta_{C^2} | \delta_A, \delta_B, \delta_C) P(\delta_{AB}, \delta_{AC}, \delta_{BC} | \delta_A, \delta_B, \delta_C) \\
&= P(\delta_A) P(\delta_B) P(\delta_C) P(\delta_{A^2} | \delta_A, \delta_B, \delta_C) P(\delta_{B^2} | \delta_A, \delta_B, \delta_C) P(\delta_{C^2} | \delta_A, \delta_B, \delta_C) \\
&\qquad P(\delta_{AB} | \delta_A, \delta_B, \delta_C) P(\delta_{AC} | \delta_A, \delta_B, \delta_C) P(\delta_{BC} | \delta_A, \delta_B, \delta_C) \\
&= P(\delta_A) P(\delta_B) P(\delta_C) P(\delta_{A^2} | \delta_A) P(\delta_{B^2} | \delta_B) P(\delta_{C^2} | \delta_C) \\
&\qquad P(\delta_{AB} | \delta_A, \delta_B) P(\delta_{AC} | \delta_A, \delta_C) P(\delta_{BC} | \delta_B, \delta_C)
\end{aligned}$$

The first equality comes from the *conditional independence principle* which assumes that the higher order terms are independent when conditioned on the first-order terms. Also it is assumed that first-order terms are independent. The *inheritance principle* assumes that the importance of a higher order term depends only on its lower order parents. The nature of the exact dependence, which is followed in all our analysis, is given next.

$$P(\delta_A = 1) = p \tag{22}$$

$$P(\delta_{A^2} = 1 | \delta_A) = \begin{cases} 0.1p & \text{if} \quad \delta_A = 0 \\ p & \text{if} \quad \delta_A = 1. \end{cases} \tag{23}$$

$$P(\delta_{AB}=1|\delta_A,\delta_B)=\begin{cases}0.1p & \text{if} \quad \delta_A+\delta_B=0\\ 0.5p & \text{if} \quad \delta_A+\delta_B=1\\ p & \text{if} \quad \delta_A+\delta_B=2.\end{cases} \tag{24}$$

In our analysis, we choose $p=0.25$. We delay the discussion of the choice of p until the sequel.

A prior must also be specified for σ. Following George and McCulloch (1993), we take

$$\sigma^2 \sim IG(\nu/2, \nu\lambda/2)$$

where IG denotes the inverted gamma distribution. It can be shown that, $\nu\lambda/\sigma^2 \sim \chi^2_\nu$.

Following George and McCulloch (1993), we took

$$\tau_j = \frac{\Delta y}{3\Delta X_j}$$

where Δy represents a "small" change in y, and ΔX_j represents a large change in X_j. In our examples, $\Delta X_j = \max(X_j) - \min(X_j)$ and $\Delta y = \sqrt{\text{Var}(y)}/5$ is used. For priors of σ, $\nu = 5$ and $\lambda = \text{Var}(y)/25$ is used.

A more general case of (23)−(24) is considered in the appendix, along with calculations that show that the expected number of main effects under (23)−(24) is

$$E(\#\text{ effects}) = pq + pq(q-1)\left\{.05 + .4p + .05p^2\right\} + pq\left\{.9p + .1\right\}. \tag{25}$$

Here q is the number of factors considered in the experiment. The first term in (25) is the expected number of main effects, second and third terms give that of 2fi's and quadratic effects, respectively.

The choice of p is now made so that expected number of active effects under the prior matches that of the experimenter's prior belief. For a specified number of effects expected to be active, (25) can easily be solved for p. This is a particularly attractive feature of the methodology since it explicitly builds in the experimenter's prior belief about the size of the model. In most situations, it is easier for an experimenter to express belief about the number of anticipated effects rather than a probability associated with an effect.

2.3.2 An illustration of analysis

We begin with a brief description of the Bayesian framework. Let y denote the data, $f(y|\theta)$, their likelihood given the vector of parameters θ, and $\pi(\theta)$ the prior distribution for θ. Then the posterior distribution for θ is

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta} \tag{26}$$

Using the posterior distribution, inference about θ can then be made.

The evaluation of the posterior for the vector θ in (26) for variable selection can be conveniently implemented by using Gibbs sampling (Geman and Geman (1984)). Gibbs sampling is a simple Markov Chain Monte Carlo (MCMC) technique for drawing samples from a posterior distribution. In general, an MCMC constructs a Markov chain $\theta^1, \theta^2, \ldots$ whose limiting distribution is the posterior (Smith and Roberts (1993)). In Gibbs sampling, full conditional distributions are employed, namely $f(\theta_i|\theta_{(-i)}, y)$, where

$$\theta_{(-i)} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_{2p+3}).$$

Starting with an arbitrarily chosen value θ^0, θ^1 is obtained by the following sequence of random draws:

$$\begin{aligned}
\theta_1^1 &\sim f(\theta_1|\theta_{(-1)}^0, y), \\
\theta_2^1 &\sim f(\theta_2|\theta_1^1, \theta_3^0, \ldots, \theta_{2p+3}^0, y), \\
\theta_3^1 &\sim f(\theta_3|\theta_1^1, \theta_2^1, \theta_4^0, \ldots, \theta_{2p+3}^0, y), \\
&\vdots \\
\theta_{2p+3}^1 &\sim f(\theta_{2p+3}|\theta_1^1, \theta_2^1, \ldots, \theta_{2p+2}^1, y).
\end{aligned} \tag{27}$$

In practice, the Markov chain is run "long enough" until it converges. Cowles and Carlin (1996) provided an overview of convergence and mixing diagnostics for MCMC.

Although the description of Gibbs sampling indicated the use of full conditionals of each of the $2p + 3$ parameters, parameters can be grouped if the full conditionals have a simple form. This occurs for the variable selection problem as formulated above, namely, the joint

distribution for $f(\theta, y)$ can be expressed as

$$\begin{aligned} f(\theta, y) &= f(y|\beta, \sigma^2, \delta)\pi(\beta, \sigma^2, \delta) \\ &= f(y|\beta, \sigma^2, \delta)\pi(\beta|\sigma^2, \delta)\pi(\sigma^2|\delta)\pi(\delta) \\ &= f(y|\beta, \sigma^2)\pi(\beta|\sigma^2, \delta)\pi(\sigma^2)\pi(\delta). \end{aligned} \tag{28}$$

The third line of (28) follows from its second line by assuming that the distribution of y depends only on β and σ^2 and that the prior for σ^2 does not depend on δ. From (28), the full conditionals can be derived. For example,

$$f(\beta|\delta, \sigma^2, y) \propto f(y|\beta, \sigma^2)\pi(\beta|\sigma^2, \delta), \tag{29}$$

which simplifies to a multivariate normal density. The other full conditionals can be determined similarly so that the Gibbs sampling algorithm can be shown to consist of a multivariate normal draw for $(\beta|\sigma^2, \delta)$, an inverse gamma draw for $(\sigma^2|\beta, \delta)$, and $p + 1$ Bernoulli draws for $(\delta_i|\beta, \sigma^2, \{\delta_j\}_{j\neq i})$.

It can be shown that,

$$\begin{aligned} f(\beta|\sigma^2, \delta, y) &\sim MN(\sigma^{-2}A_\delta X^T y, A_\delta), \\ f(\sigma^2|\beta, \delta, y) &\sim IG(a, b), \\ f(\delta|\beta, \sigma^2, y) &\propto \pi(\beta|\sigma^2, \delta)\pi(\delta) \end{aligned}$$

where $A_\delta = \sigma^2(X^T X + D_\delta^{-1})^{-1}$ and D_δ is a diagonal matrix $diag\{(c_1^{\delta_1}\tau_1)^2, \ldots, c_{p+1}^{\delta_{p+1}}\tau_{p+1}^2)\}$, $a = \frac{1}{2}(N + p + 1 + \nu)$ and $b = \frac{1}{2}[\nu\lambda + (y - X\beta)^T(y - X\beta) + \beta^T D_\delta^{-1}\beta]$.

After a burn-in of 1000 runs, the results of 10000 iterations of the Gibbs sampler are considered and the proportion of times a factor is identified as significant is taken to be the posterior probability of δ_i's. Finally, a factor is declared significant if its posterior probability is estimated to be more than a specified number c (say, 90%).

Let us come back to our motivating examples. For both of the models (response y_1 and y_2), this analysis was applied and found that this method captures the "true" model of nature. For the first example, $\Pr(\delta_i = 1) = 1.0$ for δ_i corresponding to A, B, and BE, while it is less than 0.09 for other effects. For the second example, similar results are obtained.

2.3.3 The design criterion

The analysis setting that motivates our methodology was introduced in the last Section. From an anaysis standpoint, the aim is to identify a subset of predictors that best explain the data. From a design perspective, the goal is to identify the set treatments that best facilitates the aim of the analysis. Thus we aim to identify the "optimal" design with q three-level factors in n trials for estimating the parameters of the linear model in (20).

Common choices of experiment plans are the regular fraction factorial designs that are most often ranked by the minimum aberration criterion (Fries and Hunter (1980)), which sequentially minimizes the elements of the word-length pattern. There are two underlying assumptions which motivate the use of the minimum aberration criterion:

A1. *Effect Sparsity*: The number of important effects is relatively small.

A2. *Effect Hierarchy*: Lower order effects are more likely to be important than higher order effect and effects of the same order are equally important.

In order to explore the complex aliasing structure between main effects and two-factor interactions, non-regular fractional factorials such as Plackett-Burman (1946) designs are also used frequently. Consider the most "notorious" case, the 12 run Plackett-Burmann design with 11 factors. For each factor X, its main effect is partially aliased with the 45 two-factor interactions not involving X, thereby making it difficult to disentangle or interpret the significance of interactions. Draper (1985) commented that unless (i) the interactions are small or negligible or (ii) there is relatively few "important" factors, the results from Plackett-Burmann designs may be confusing. In Taguchi and Wu (1979, p. 35), it is stated that "... no interactions are calculated even if they exist ... these interactions are treated as errors, so it is advantageous to have the effects of these interactions uniformly distributed in all (design matrix) columns." But this is not true that estimated main effects are not affected by interactions. In particular, as noted by Hamada and Wu (1992), ignoring interactions can lead to (i) important effects being missed, (ii) spurious effects being detected, and (iii) estimated effects having reversed signs resulting incorrectly recommended factor levels. We will come back to this issue in subsequent examples.

Recently, methods have been proposed for analyzing complex aliasing designs that entertain models with both main effects and two-factor interactions (Hamada and Wu, 1992; Chipman, Hamada and Wu, 1997). Hamada and Wu (1992) proposed an iterative guided stepwise regression strategy for analyzing the data from such designs that allows entertainment of interactions. While providing a feasible alternative to an all-subset regression, their strategy explore a small part of the entire model space. More comprehensive searches are provided by Chipman, Hamada and Wu (1997). They took a Bayesian Approach which combines the Stochastic Variable Search Selection (SSVS) algorithm of George and McCulloch (1993) with priors for related predictors given by Chipman (1996). A suitable class of hierarchical prior distributions focuses the search on a reasonable class of models. These iterative approaches rely on an additional assumption to help sort through the complex aliasing structure:

A3. *Effect Inheritance*: An interaction is more likely to be important if one or more of its parent factors are also important.

In light of this additional assumption, Hamada and Wu (1992) viewed the complex aliasing of non-regular fractional factorials as an advantage because non-regular fractional factorials give the opportunity to identify promising interactions as well as main effects. The belief that estimation of some models is more important than others is not easy to incorporate into criteria such as minimum aberration and estimation capacity in some practical applications but the Bayesian approach is ideal for these kind of situations.

The measurement of the ability to discriminate among models is accomplished via a Bayesian approach, which has similarities to the model discrimination criterion (MD) (Meyer, Steinberg and Box (1996)) which is based on the Kullback-Leibler information. Bingham and Chipman (2002) propose a different criterion, based on the Hellinger distance between predictive densities, to help distinguish between competing models. The Hellinger distance is preferable to the Kullback-Leibler information in this setting because it requires half the computational expense and is bounded. Also, the Hellinger distance has appealing properties that allows experimenters to use it as a basis for choosing an appropriate run

size.

Bingham and Chipman (2002) considered only two level factors. Their idea can be generalized to three level factors where linear and quadratic effect of the main factors and linear-by-linear interactions can be considered. This has similarities to the factor screening and response surface exploration of Cheng and Wu (2001). Standard practice in response surface methodology performs factor screening and response surface exploration sequentially, using different designs. In this chapter, we show how Baysian approach can overcome this limitation. We find the optimal design based on Cheng and Wu's (2001) approach and also analyse that with our Baysian approach. We see that the optimal design need not be same as that of ours and following our approach, we get the same result when the actual model follows strong heridity, but unlike their approach, it successfully identifies the models with weak heredity.

The HD criterion is based on the pairwise Hellinger distance between predictive densities, and is written as:

$$HD = \sum_{i<j} P(M_i)P(M_j)H(f_i, f_j), \tag{30}$$

where $P(M_i)$, and $P(M_j)$ are the prior probabilities for models M_i and M_j, f_i, and f_j are the predicitve densities for response vector y under models M_i and M_j, respectively. $H(f_i, f_j)$ is the Hellinger distance between the two models

$$H(f_i, f_j) = \int (f_i^{1/2} - f_j^{1/2})^2 dY = 2 - 2\int (f_i\ f_j)^{1/2} dY. \tag{31}$$

The design enters this criterion through f_i and f_j, since the predictive densities are evaluated at the values of the factors specified in the design. The motivation for (31) is that it is easiest to discriminate between models if they make different predictions. It therefore makes sense to consider designs that maximimize (31). The weighting of the Hellinger distances in (30) serves to put priority on distinguishing the most probable models.

2.3.4 Priors for Design

The HD criterion (30) uses predictive densities, which implies a Bayesian formulation of the problem. Section 3.1 outlines the priors used in analysis, both for models (the M_i) and

for parameters conditional on a model (β_i, σ). Similar priors are used in the design criterion (31), (30), with slight modifications to the prior on β.

The slight difference is the prior on $\beta_i|\delta_i$. In (21), a mixture of two normals is used. For purposes of identifying the design, we use a related prior, in which inactive effects have a prior that is degenerate at $\beta_i = 0$. The prior on active terms remains the same.

The coefficient vector β_i and the associated matrix of regressors X_i are indexed by i. Let r_i be the number of columns in X_i (i.e., the number of effects in model M_i plus one one additional column for the intercept). Thus $\pi(\beta_i|\sigma^2) \sim MVN(0, \sigma^2\Gamma_i)$, where

$$\Gamma_i = \gamma^2 \begin{pmatrix} c & 0 \\ 0 & I_{r_i-1} \end{pmatrix}. \tag{32}$$

We choose $c = 1,000,000$ so that the prior on the intercept is has mean 0 and large variance. In all calculations presented we take $\gamma = 2$. MSB suggest this is a reasonably uninformative value in the context of starting designs. More careful choice of γ is required for follow-up designs, which are not considered in this article.

This prior formulation means the Hellinger distance (31) can be written as

$$H(f_i, f_j) = 2 - \frac{2}{\left|\frac{1}{2}\left(\Sigma_i^{-1/2}\Sigma_j^{1/2} + \Sigma_i^{1/2}\Sigma_j^{-1/2}\right)\right|^{1/2}} \tag{33}$$

where

$$\Sigma_i = (I_n + X_i'\Gamma_i X_i). \tag{34}$$

Some intuition about the distance measure can be gained by considering the situation when there are only two competing models and only one trial to be conducted. In this instance, the Hellinger distance will be greatest for a design where there is little uncertainty about one model and large uncertainty about the other. The criterion amounts to choosing trials where the average relative uncertainty between models is largest.

2.4 Searching for optimal designs

The basic strategy of Bingham and Chipman will be employed. We review this strategy, and describe modifications to deal with the increased computational complexity of three-level designs.

The search for promising designs involves two challenges:

- Evaluation of the HD criterion.
- An effective search algorithm for HD-optimal designs.

In many cases, to evaluate HD, we cannot calculate (30) for all pairs of models because the model space is too large. Instead, we attempt to evaluate the largest terms in (30), by discarding models that have small prior probability $P(M_i)$. By replacing an average over all pairs of models with an average over the most probable models, the evaluation of HD becomes tractable. The most probable models are identified by sampling from the prior. This approach was taken in Bingham and Chipman, and is used here without modification.

Bingham and Chipman utilize an exchange algorithm for design optimization. Optimization is over experimental settings consisting of all possible combinations of three-level factors. A modified version of this algorithm is as follows:

1. Begin with a random n-run design, with design points sampled with replacement from the 3^q full factorial design.
2. Repeat:
   (a) Identify the index k of the run whose removal will least decrease HD.
   (b) Sample one run at a time, without replacement, from the set of all possible three-level runs. This repeats until either all runs have been sampled, or one run leads to an increase in HD when run k is replaced with the new run.
   (c) Replace run k with the new run identified in step 2 (b).

Step 2 will repeat until there is little or no improvement in HD over several iterations.

Bingham and Chipman's algorithm evaluated all possible runs in step 2b, instead of stopping as soon as an improvement was observed. The large number of candidate runs (3^q) makes this infeasible for most problems. There is no guarantee that this approach will converge to the HD-optimal design, therefore multiple random designs are used as start points to generate a variety of promising designs. Another interesting start point would be a conventional 3-level design.

2.5 Examples and a simulation experiment

The performance of Bayesian approach as well as CW approach are studied via simulations with two initial designs both having 18 runs. The first design is an Orthogonal Array and the other is the HD-optimal design. There are seven factors under study. The models are generated from the priors (22), (23) and (24) with $p = 0.344$ which gives the expected number of priors to be 5. Here coefficients are set to constant values. Analysis is performed with CW as well as Bayesian methods and there a factor is declared significant if its posterior probability is greater than a fixed preassigned constant, c. The results are tabulated for three different values of c (0.9, 0.75 and 0.5).

Table 6 summarizes the results for 1000 simulations. The columns under CW gives the average performance of CW method whereas the columns under Bayes give the results of Bayesian analysis for different c's. Let us illustrate the meanings of the rows through an example. Suppose the data are generated from the model A, B, C and AD and the model identified is A, C and BC. Then the first row ("Missed") calculates the proportion of missed effects, which is 2/4 as there are 4 factors in the model among which 2 (B and AD) are missing in the model identified. The second row ("Captured") calculates the complement i.e. proportion of the true model captured by the analysis which is again 2/4. The third row ("False +ve") gives the proportion of factors identified by the model but not in the true model. In this example, this number is 1/4 (the numerator is 1 as there is only one such effect, namely BC). The last roe ("Correct") counts the proportion of times the exact "true" model is identified. The numbers reported in Table 2.5 are the values averaged over 1000 simulations.

It is easy to see that the Bayesian analysis performs better. As expected, with an increase of cut-off value c, the rate of identification of false effects decreases, but at the same time the chances of missing an important effect increases as well as the rate at which the correct model is identified is also decreases. However, under all circumstances, the Bayesian approach outperforms the CW approach. For the 18-run OA, CW approach identifies the correct model only 8% of the time whereas the Bayesian approach identifies it more than 34% of the cases, with probability of wrongly identifying significant effects

Table 6: Simulation studies (in %) for different designs

Performance Criteria	OA				HD-optimal design			
	CW	Bayes			CW	Bayes		
		$c = .9$	$c = .75$	$c = .5$		$c = .9$	$c = .75$	$c = .5$
Missed	65	36	28	21	52	47	39	28
Captured	35	64	72	79	48	53	61	72
False +ve	4	0.0	0.3	1	1.6	0.1	0.8	4
Correct	8	20	28	34	10	21	27	35

uniformly smaller.

The 18-run HD-optimal design is given in the appendix. It does not perform significantly better than the 18-run OA. Although Orthogonal Arrays are widely used in practice, they are not available for every possible run size. For example, if we have resource to have 17 runs of an experiment, there are no fractional factorials or Orthogonal Arrays available. However, one can get a 17-run HD optimal design and use it for running the experiment.

2.6 Summary and Conclusions

Instead of conducting two experiments in the response surface methodology, CW approach aims to use only one design to identify the correct model. However, this method can lead to misleading conclusions. A Bayesian analysis approach is suggested here, which identifies the correct model using only one experiment, but with much higher accuracy. Orthogonal arrays are widely used but they are limited in numbers. The HD-optimal design criterion is introduced here which can be used to obtain optimal designs for any arbitrary run size.

2.7 References

Bingham, D.R. & Chipman, H. A. (2002), "Optimal Designs for Model Selection," *Technometrics*, tentatively accepted.

Bingham, D.R. & Li, W. (2002), "A Class of Optimal Robust Parameter Designs," *Journal of Quality Technology*, 34, 244–259.

Bingham, D.R. & Sitter, R.R. (2003), "Fractional Factorial Split-Plot Designs for Robust Parameter Experiments," *Technometrics*, 45, 80–89.

Cheng, C.S. and Mukerjee, R. (1998), "Regular Fractional Factorial Designs with Minimum Aberration and Maximum Estimation Capacity," *Annals of Statistics*, 26, 2289-2300.

Cheng, C.S., Steinberg, D.M. and Sun, D.X. (1999), "Minimum Aberration and Model Robustness for Two-Level Fractional Factorial Designs," *JRSS B*, 61, 85–93.

Cheng, Shao-Wei and Wu, C. F. J. (2001), "Factor Screening and Response Surface Exploration," *Statistica Sinica*, 11, 3, 553–604.

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors," *Canadian Journal of Statistics*, 24, 17–36.

Chipman, H., Hamada, M. and Wu, C. F. J. (1997), "A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing," *Technometrics*, 39, 372–381.

Cowles, M. K. and Carlin, B. P. (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A Comparative Review," *Journal of the American Statistical Association*, 91, 883–904.

Draper, N. R. (1985), "Plackett and Burman Designs," *Encyclopedia of Statistical Sciences* Jhon Wiely & Sons, New York, NY, 754–758.

Franklin, M. F. and Bailey, R. A. (1977), "Selection of Defining Contrasts and Confounded Effects in Two-Level Experiments," *Applied Statistics*, 26, 321–326

Fries, A. and Hunter, W. G. (1980), "Minimum Aberration 2^{k-p} Designs," *Technometrics*, 22, 601–608.

Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

George, E. I. and McCulloch, R. E. (1993), "Variable selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

Hall, Jr., M. (1961), "Hadamard Matrices of Order 16," *Research Summary*, 1, 21–26 Pasadena, CA., Jet Propulsion Laboratory.

Hamada, M. and Wu, C. F. J. (1992), "Analysis of Designed Experiments with Complex Aliasing," *Journal of Quality Technology*, 24, 130–137.

Meyer, R. D., Steinberg, D. M. and Box, G. E. P (1996), "Follow-up Designs to Resolve Confounding in Fractional Factorials," *Technometrics*, 38, 303–313.

Plackett, R. L. and Burman, J. P. (1946), "The Design of Optimum Multifactorial Experiments," *Biometrika*, 33, 305–325 and 328–332.

Sun, D. X. (1997), *Estimation Capacity and Related Topics in Experimental Design*, PhD. Thesis, University of Waterloo.

Taguchi, G. (1987), "System of Experimental Design" *Unipub/Kraus International Publications*, White Plains, New York.

Taguchi, G. and Wu, Y. (1979), "Introduction to Off-line Quality Control," *Central Japan Quality Control Association*, Nagoya, Japan.

Wu, C. F. J and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, John Wiley & Sons.

Wu, C. F. J and Zhu, Y. (2003), "Optimal selection of single arrays for parameter design experiments," *Statistica Sinica*, 4, 1179–1199.

CHAPTER III

ALIASING RELATIONS OF MIXED FACTORIALS IN THE FORM OF PRODUCT ARRAYS

3.1 Introduction

Two- and three-level factorial and fractional factorial designs are widely used in industrial experimentations and are discussed in detail in design of experiments textbooks (Box, Hunter and Hunter (1978); Cochran and Cox (1950)). The literature on symmetric designs is already voluminous. For example, the theory of regular fractions for symmetric factorials is given by Dey and Mukerjee (1999). Wu and Hamada (2000) devote a full chapter in their applied design of experiment textbook, on analysis techniques for mixed-level factorial plans. Although these are important designs, their aliasing patterns have not been studied explicitly.

Mixed-levels designs typically occur when there are both qualitative and quantitative factors in the experiment, and the qualitative factors have more than two levels and the quantitative factors have two levels. Consider an experiment by Hale-Bennett and Lin (1997) and reported in Wu and Hamada (2000) that was performed to improve a painting process of charcoal grill parts. A mixed-level 36-run design (Table 16) was used to study six factors: three of them (A, B, C) were at two levels and the other three (D, E, F) were at three levels. It is a $2^{3-1} \times 3^{3-1}$ design which consists of $4 \times 9 = 36$ runs and is a "product" of a 4-run 2^{3-1} and a 9-run 3^{3-1} design. Now it is not evident that the factorial effect ABD^2E is same as that of $ABDE^2$. If all the factors of a factorial effect are at two-levels, AB for example, a modulo 2 operation should be performed. Similarly, modulo 3 operations are used when all of them are have three level, as in the case for DE^2. But what about $ABDE^2$? It is not obvious whether modulo 2 or modulo 3 operations should be done in calculating the aliasing relationship of a mixed-level factorial effect. In fact, there is no

simple answer to this question, as will be clear from the discussions of Section 2.

In Section 2, we develop the general theory for $s_1^{n_1} \times s_2^{n_2}$ factorial designs and illustrate it in the context of a $2^3 \times 3^3$ design. In Section 3, we discuss $s_1^{n_1-k_1} \times s_2^{n_2-k_2}$ factorial designs and discuss the paint experiment as an example of $2^{3-1} \times 3^{3-1}$ designs.

3.2 $s_1^{n_1} \times s_2^{n_2}$ factorial designs

An experiment involving n_1 factors each at s_1 levels and n_2 factors each at s_2 levels is an $s_1^{n_1} \times s_2^{n_2}$ asymmetrical factorial experiment. Suppose the levels of the s_i-level factor are coded as s_i elements of Galois field $GF(s_i)$ where s_i is a prime or prime power. With levels as $0, 1, \ldots s_i - 1$, a typical treatment combination, i.e., a combination of the levels of the $n_1 + n_2 = n$ factors will be represented by an ordered n-tuple $i_1 \ldots i_{n_1} j_1 \ldots j_{n_2}$ where $i_k \in \{0, 1, \ldots, s_1 - 1\}, 1 \leq k \leq n_1$ and $j_k \in \{0, 1, \ldots, s_2 - 1\}, 1 \leq k \leq n_2$. Clearly, altogether there are $s_1^{n_1} s_2^{n_2}$ treatment combinations.

In what follows, (a, b) and $(a', b')'$ will be used interchangeably for the sake of notational simplicity where a and b are column vectors of dimension n_1 and n_2, respectively.

A treatment contrast L is said to belong to the pencil (a, b) if it is of the form

$$L = \sum_{i=0}^{s_1-1} \sum_{j=0}^{s_2-1} l(i,j) \{ \sum_{(x,y) \in V_{i,j}(a,b)} \tau(x,y) \}, \tag{1}$$

where

$$V_{i,j}(a,b) = \{(x,y) = (x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2},)' : a'x = \alpha_i, b'y = \beta_j\},$$

$0 \leq i \leq s_1 - 1$, $0 \leq j \leq s_2 - 1$; the effect of a treatment combination represented by (x, y) will be denoted by $\tau(x, y)$ and $l(i, j)$'s are real numbers, not all zeros, satisfying

$$\sum_{i=0}^{s_1-1} l(i,j) = \sum_{j=0}^{s_2-1} l(i,j) = 0. \tag{2}$$

In other words, a treatment contrast L belongs to (a, b) if for all (x, y) belonging to the same $V_{i,j}(a, b)$, the coefficient of $\tau(x, y)$ in L is also the same.

In general, consider any two pencils (a, b) and (a^*, b^*). These two pencils are distinct if a is distinct from a^* and b is distinct from b^*, in the sense of symmetric factorial designs.

Recall that, in symmetric fractions, pencils with proportional entries are considered as identical.

For a $s_1^{n_1} \times s_2^{n_2}$ product array, for the $s_i^{n_i}$ factorial part, there are $(s_i^{n_i}-1)/(s_i-1)$ distinct pencils which involve only s_i-level factors. The distinct pencils involving both s_1- and s_2-level factors are given by the products of those two sets of pencils involving only s_1- or only s_2-level factors. A simple counting of degrees of freedom justifies this formulation. Recall that the total number of factorial effects involving only s_i-level factors is $(s_i^{n_i}-1)/(s_i-1)$, each with (s_i-1) d.f.. As the interactions involve both s_1- and s_2-level factors are given by products of the other two sets of pencils, i.e., there are $(s_1^{n_1}-1)/(s_1-1) \times (s_2^{n_2}-1)/(s_2-2)$ pencils of this kind, each with $(s_1-1)(s_2-1)$ d.f.. Thus, the above description accounts for

$$\frac{s_1^{n_1}-1}{s_1-1}(s_1-1)+\frac{s_2^{n_2}-1}{s_2-1}(s_2-1)+\frac{s_1^{n_1}-1}{s_1-1}\frac{s_2^{n_2}-1}{s_2-1}(s_1-1)(s_2-1)=s_1^{n_1}s_2^{n_2}-1$$

d.f. which agrees with the fact that there are $s_1^{n_1}s_2^{n_2}$ in all.

Following Bose (1947), in the Appendix we give the definition for treatment contrasts belonging to factorial effects for the general case of an $s_1 \times \ldots \times s_n$ factorials. The next two results link pencils with factorial effects.

Result 1 *(a) Treatment contrasts belonging to distinct pencils are orthogonal to each other. (b) Let (a,b) be a pencil such that $a_i \neq 0$ if $i \in \{i_1,\ldots,i_g\}$, and $=0$ otherwise, $b_j \neq 0$ if $j \in \{i_1,\ldots,i_h\}$, and $=0$ otherwise, where $1 \leq i_1 < \ldots < i_g \leq n_1$, $1 \leq j_1 < \ldots < i_h \leq n_2$ and $1 \leq g \leq n_1$, $1 \leq h \leq n_2$. Then any treatment contrast belonging to (a,b) also belongs to the factorial effect $F_{i_1}\ldots F_{i_g}F'_{j_1}\ldots F'_{j_h}$.*

Example

Let us consider the $2^3 \times 3^3$ full factorial design with two-level factors A, B, C and three-level factors D, E, F. The levels of A, B, C are denoted by 0 and 1, and those of D, E, F are denoted by 0, 1 and 2. Then a typical treatment combination, i.e., the combination of the levels of six factors will be represented by $x=(a,b,c,d,e,f)$, where $a,b,c \in \{0,1\}$ and $d,e,f \in \{0,1,2\}$. For example, the factorial effect $ABDE^2$ is denoted by $(a,b,c,d,e,f) \equiv (1,1,0,1,2,0)$. Clearly there are $2^3 \times 3^3 = 216$ possible treatment

combinations.

The pencils involving only the two-level factors or only the three-level factors can be described as usual. Thus the pencil AB is given by the contrasts between the two sets of treatment combinations for which $a + b = 0$ or 1 mod 2. More explicitly, these two sets are $\{x : x = (a, b, c, d, e, f), a + b = 0 \text{ mod } 2\}$ and $\{x : x = (a, b, c, d, e, f), a + b = 1 \text{ mod } 2\}$. Clearly, there are 108 treatment combinations in each of these sets, e.g., the first set consists of the treatment combinations $(0, 0, c, d, e, f)$ and $(1, 1, c, d, e, f)$, where $c \in \{0, 1\}$ and $d, e, f \in \{0, 1, 2\}$, leading to 54+54=108 treatment combinations in all. In a similar manner, the pencil DEF^2, involving exclusively the three-level factors, is given by contrasts among three sets of treatment combinations for which $d + e + 2f = 0, 1$ or 2 mod 3. As before, there are $8 \times 9 = 72$ treatment combinations in each of these sets. It is clear that any pencil involving A, B, C will have 1 d.f.while any pencil involving only D, E, F will have 2 d.f..

Now consider the interactions that involve both two- and three-level factors. Recall that there are 7 pencils A, B, C, AB, AC, BC and ABC involving only the two-level factors. Similarly there are 13 distinct pencils D, E, F, DE, DE^2, ..., DE^2F^2 involving only the three-level factors. The pencils representing interactions that involve both two- and three-level factors are given by the products of these two sets of pencils, i.e., there are $7 \times 13 = 91$ pencils of this kind, namely, AD, AE, ..., ADE^2F^2, BD, BE, ..., BDE^2F^2,..., $ABCD$, $ABCE$, ..., $ABCDE^2F^2$. Each of these 91 pencils carries 2 d.f.. Clearly, for example, AD and AD^2 mean the same thing in this formulation (so we write only AD). Similarly $ABDE^2 = ABD^2E = AB(DE^2)^2$. Taking care of the 7 pencils involving only the two-level factors and the 13 pencils involving only the three-level factors, the above description accounts for $7 \times 1 + 13 \times 2 + 91 \times 2 = 215$ d.f., which agrees with the fact that there are $2^3 \times 3^3 = 216$ treatment combinations in all.

How does one actually define contrasts belonging to pencils as considered in the last paragraph ? Consider, for example, the pencil $ABDE^2$. For $i = 0, 1$ and $j = 0, 1, 2$, define $V_{i,j} = V_{i,j}(110, 120) = \{x : x = (a, b, c, d, e, f), a + b = i \text{ mod } 2, d + 2e = j \text{ mod } 3\}$. Note that $a + b = i$ mod 2 corresponds to AB, and $d + 2e = j$ mod 3 corresponds to DE^2.

Clearly, each of the six sets $V_{i,j}$ has cardinality $4 \times 9 = 36$. Let $T(i,j)$ be the total of the treatment effects for the treatment combinations in $V_{i,j}$. Then a typical contrast belonging to the pencil $ABDE^2$ will be of the form $\sum_i \sum_j l(i,j)T(i,j)$, where the scalars $l(i,j)$, not all zeros, must satisfy $l(0,j)+l(1,j) = 0$ for every j and $l(i,0)+l(i,1)+l(i,2) = 0$ for every i. Thus there will be two such independent treatment contrasts, namely, $L_1 = T(0,0) - T(1,0) - T(0,2) + T(1,2)$ and $L_2 = T(0,0) - T(1,0) - 2T(0,1) + 2T(1,1) + T(0,2) - T(1,2)$.

3.3 $s_1^{n_1-k_1} \times s_2^{n_2-k_2}$ fractional factorial designs

A regular fraction of an s^n symmetrical factorial, where s (≥ 2) is a prime or prime power, is specified by any k ($1 \leq k < n$) linearly independent pencils, say $b^{(1)}, \ldots, b^{(k)}$, and consists of treatment combinations z satisfying $Bz = c$, where B is a $k \times n$ matrix with rows $(b^{(i)})', 1 \leq i \leq k$, and c is a fixed $k \times 1$ vector over $GF(s)$. The specific choice of c is inconsequential. Hence, without loss of generality, it is assumed that $c = 0$, the $k \times 1$ null vector over $GF(s)$. Then a regular fractional factorial plan is given by, say,

$$d(B) = \{z : Bz = 0\}.$$

In the same line, for a $s_1^{n_1} \times s_2^{n_2}$ design, a regular fractional factorial plan, $s_1^{n_1-k_1} \times s_2^{n_2-k_2}$ is given by

$$d(B) = \{z : Bz = 0\} = \{(x,y) : B_1 x = 0, B_2 y = 0\}$$

where

$$B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}.$$

Note that $d(B_i)$ gives a regular $s_i^{n_i-k_i}$ fractional factorial plan. For a symmetric fractional factorial, a pencil is called a defining pencil if it belongs to the row space of B. Equivalently, a defining pencil of a $s_1^{n_1-k_1} \times s_2^{n_2-k_2}$ design is of the form (b_1, b_2) where b_i is a defining pencil of $s_i^{n_i-k_i}$.

Consider now any defining pencil (a,b). Then $a' = \lambda' B_1$ and $b' = \xi' B_2$ for suitable λ and ξ with entries from $GF(s_1)$ and $GF(s_2)$, respectively. Now it is not difficult to see that

$d(B) \subset V_{0,0}(a,b)$. Recalling the definition of a treatment contrast, the following result is evident.

Result 2 *No treatment contrast belonging to any defining pencil is estimable in* $d(B)$.

Two pencils are aliases of each other if their difference belongs to the row space of B. Let $\mathcal{C}$ be the set of distinct pencils which are not defining pencils. Then we get the following lemma.

Lemma 1 *Let the pencils* $(a,b),(a^*,b^*) \in \mathcal{C}$ *be aliases of each other and*

$$L = \sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}(a,b)} \tau(x,y)\},$$

and

$$L^* = \sum_{k=0}^{s_1-1}\sum_{l=0}^{s_2-1} l^*(k,l)\{\sum_{(x,y)\in V_{k,l}(a^*,b^*)} \tau(x,y)\},$$

be the treatment contrasts belonging to (a,b) *and* (a^*,b^*)*, respectively. Then the parts of* L *and* L^**, which involve only the treatment combinations included in* $d(B)$*, are identical.*

Let $L(B)$ be the part of L that involves only the treatment combination involved in the fraction $d(B)$ and is often called the relevant part of L. Then the relevant parts of corresponding contrasts belonging to pencils that are aliases of each other, are identical. Let $V_{i,j}((a,b),B) = V_{i,j}(a,b) \cap d(B)$. Then for any pencil $(a,b) \in \mathcal{C}$ and for its alias set $\mathcal{A}$, we get the following theorem.

Theorem 1 *For* $(a,b) \in \mathcal{A}$, consider the corresponding treatment contrast

$$L = \sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}(a,b)} \tau(x,y)\},$$

Then

$$\sum_{(a,b)}\Big[\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}(a,b)} \tau(x,y)\}\Big]$$
$$= s_1^{k_1}s_2^{k_2}\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}((a,b),B)} \tau(x,y)\} \qquad (3)$$

where $\sum_{(a,b)}$ denote the sum over all $(a,b) \in \mathcal{A}$.

To prove this theorem, we need the following lemma.

Lemma 2 *Consider any pencil $(a,b) \in \mathcal{C}$ and let $\mathcal{A}$ denote its alias set. Let*

$$\phi_{i,j}((a,b)(x,y)) = \begin{cases} 1 & \text{if } a'x = \alpha_i, b'y = \beta_j, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then for every treatment combination (x,y) and every (i,j), $0 \le i \le s_1 - 1, 0 \le j \le s_2 - 1$,

$$\sum_{(a,b)} \phi_{i,j}((a,b),(x,y)) =$$

$$\begin{cases} s_1^{k_1} s_2^{k_2} & \text{if } (x,y) \in V_{i,j}((a,b),B), \\ 0 & \text{if } (x,y) \in d(B) - V_{i,j}((a,b),B), \\ s_1^{k_1-1} s_2^{k_2-1} & \text{if } (x,y) \notin d(B). \end{cases} \tag{5}$$

Proof. A pencil in $\mathcal{A}$ is of the form (p,q) where $p = a + B_1'\lambda$ and $q = b + B_2'\xi$ where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{k_1})', \lambda_i \in GF(s_i)$ and $\xi = (\xi_1, \xi_2, \ldots, \xi_{k_2})', \xi_j \in GF(s_2)$. For fixed (x,y) and (i,j),

$$\begin{aligned} &\sum_{(a,b)} \phi_{i,j}((a,b),(x,y)) = \\ \#\Big\{ \quad &\lambda = (\lambda_1, \ldots, \lambda_{k_1})' : x + \lambda' B_1 x = \alpha_i; \\ &\xi = (\xi_1, \ldots, \xi_{k_2})' : b'y + \xi' B_2 y = \beta_j, \\ &\lambda_i \in GF(s_i), \xi_j \in GF(s_2) \forall i,j \ \Big\} \end{aligned}$$

where # denotes the cardinality of a set.

(i) If $(x,y) \in V_{i,j}(a,b)$ then $a'x + \lambda' B_1 x = \alpha_i$ for all $k_1 \times 1$ vectors over $GF(s_1)$ and $b'y + \xi' B_2 y = \beta_j$ for all $k_2 \times 1$ vectors over $GF(s_2)$. Hence the RHS of (5) is $s_1^{k_1} s_2^{k_2}$.

(ii) If $(x,y) \in d(B) - V_{i,j}((a,b),B)$, then $B_1 x = 0$, $B_2 = 0$. Also, $a'x \ne \alpha_i$ and/or $b'y \ne \beta_j$. Then $\sum_{(a,b)} \phi_{i,j}((a,b),(x,y)) = \#\Big\{(\lambda,\xi) : a'x = \alpha_i, b'y = \beta_j\Big\} = 0$.

(iii) If $(x,y) \notin d(B)$, then $B_1 x \ne 0$, $B_2 y \ne 0$. Trivially $a'x + \lambda' B_1 x = \alpha_i$ iff $(B_1 x)'\lambda = \alpha_i - a'x$. Since $B_1 x \ne 0$, exactly as in the proof of Lemma 2.1, one can freely choose $(\lambda_2, \ldots, \lambda_{k_1 - 1})$ in $s_1^{k_1 - 1}$ ways to satisfy the above equation. Similarly $b'y + \xi' B_2 y = \beta_j$ gives $s_2^{k_2 - 1}$ choices of ξ_l's. Combining the values of λ_k's and ξ_l's, the result follows.

Proof of Theorem 1. Let Ω denote the set of all $s_1^{n_1}s_2^{n_2}$ treatment combinations. Using Lemma 3.2 and the indicator variable $\phi_{i,j}((a,b)(x,y))$ in (4),

$$\begin{aligned}
&\sum_{(a,b)}\Big[\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}(a,b)}\tau(x,y)\}\Big]\\
=\;&\sum_{(a,b)}\Big[\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in\Omega}\phi_{i,j}((a,b)(x,y))\tau(x,y)\}\Big]\\
=\;&\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\Big\{\sum_{(x,y)\in\Omega}\Big[\sum_{(a,b)}\phi_{i,j}((a,b)(x,y))\Big]\tau(x,y)\Big\}\\
=\;&s_1^{k_1}s_2^{k_2}\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1} l(i,j)\{\sum_{(x,y)\in V_{i,j}((a,b),B)}\tau(x,y)\},
\end{aligned}$$

since $\sum l(i,j)=0$.

The RHS of (3) is a contrast involving only the treatment combinations included in $d(B)$. Therefore the RHS and hence the LHS of (3) will be estimable using the plan $d(B)$. In other words, while pencils belonging to the same alias set, are confounded with one another (Lemma A.3), the sum of corresponding contrasts, belonging to such pencils is estimable in $d(B)$. Thus any treatment contrast belonging to a pencil (a,b) which is not a defining pencil is estimable in $d(B)$ if and only if corresponding contrasts belonging to all other pencils that are aliased with (a,b) are ignorable.

We say that a pencil is estimable in $d(B)$ if so is every treatment contrast belonging to it. Similarly, if every treatment contrast belonging to a pencil is ignorable, then the pencil itself is called ignorable. Hence the following result is immediate.

Result 3 *A pencil b, which is not a defining pencil, is estimable in $d(B)$ if and only if all other pencils that are aliased with b are ignorable.*

Example

Now consider the fractional factorial design used for the Paint experiment. This kind of fraction treats the two- and three-level factors separately, leading to a product array. It is easy to see that these two-level factors form a 2^{3-1} design with $C=AB$. The 2^{3-1} design is used for the first four rows and repeated for the next eight groups of four rows. The three-level factors form a 3^{3-1} design with $F=DE$. Each of the nine combinations of the 3^{3-1} design appears in four consecutive entries in Table 16. The 36-run design in Table 16

consists of the 4×9 combinations of the 2^{3-1} design and the 3^{3-1} design and is called a $2^{3-1} \times 3^{3-1}$ design.

The defining relation of the $2^{3-1} \times 3^{3-1}$ design can be obtained from those of its two component designs: $\mathbf{I} = ABC$ and $\mathbf{I} = DEF^2$. So we decide to include the treatment combinations $x = (a, b, c, d, e, f)$ satisfying $a + b + c = 0 \mod 2$ and $d + e + 2f = 0 \mod 3$. There are 4 such choices of (a, b, c) and 9 such choices of (d, e, f). Combining these, we will have $4 \times 9 = 36$ treatment combinations in our plan which will be in the from of product array. The alias sets will again be of three types :

Type I (involving only two-level factors arising from $I = ABC$): These are $A = BC$, $B = AC$, $C = AB$, each carrying 1 d.f..

Type II (involving only three-level factors arising from $I = DEF^2$): there will be 4 such alias sets, each carrying 2 d.f.. these are $D = DE^2F = EF^2$; $E = DF^2 = DE^2F^2$; $F = DE = DEF$ and $DE^2 = DF = EF$.

Type III (involving the "mixed" pencils discussed earlier): These are obtained by combining each type I alias set with each type II alias set, e.g., a typical alias set of type III will be $AD = ADE^2F = AEF^2 = BCD = BCDE^2 = BCEF^2$. There will be $3 \times 4 = 12$ such alias sets each carrying 2 d.f..

Thus the three types of alias sets mentioned above will together carry $3 \times 1 + 4 \times 2 + 12 \times 2 = 35$ d.f.. This again agrees with the fact that there are 36 treatment combinations in the chosen fraction.

For any pencil in a type III alias set, it is not hard to see that each set $V_{i,j}$ corresponding to that pencil will contain six of the treatment combinations included in our fraction. To see this, consider the pencil $BCDE^2F$. A treatment combination $x = (a, b, c, d, e, f)$ in our fraction will then belong to the corresponding $V_{i,j}$ if it satisfies $b + c = i \mod 2$ and $d + 2e + f = j \mod 3$, in addition to satisfying $a + b + c = 0 \mod 2$ and $d + e + 2f = 0 \mod 3$ needed for inclusion in the fraction. Now the first and third of the equations just mentioned yield two solutions for (a, b, c) while the second and fourth of these equations yield three solutions for (d, e, f). Combining these, we get six solutions altogether.

3.4 Summary

The designs discussed here are called *cross arrays* (*inner-outer array* in Taguchi's terminology). They are commonly used in robust parameter designs. Although the cross arrays have been used for mixed-level designs (Shoemaker et al., 1991), their aliasing relations have not been studied rigorously. In this chapter, we have discussed the effect aliasing for fractional factorial designs of mixed-level designs. The results obtained for mixed-level factorial and fractional designs are similar to that of symmetric factorials. Here only $s_1^{n_1} \times s_2^{n_2}$ factorials are discussed, although with heavier notation, and without any significant conceptual change, it is possible to obtain general results for $s_1^{n_1} \times s_2^{n_2} \times \ldots s_m^{n_m}$ factorials. One drawback of the above approach is that the cross arrays may become too large. The rigorous study of mixed-level cross arrays gives a deeper insight on the estimation properties of the design and paves the way for further research in extending the minimum aberration and estimation capacity criteria for such designs.

3.5 References

Bose, R. C. (1947), "Mathematical theory of the symmetrical factorial design," *Sankhyā*. 8, 107–166.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, John Wiley, New York.

Cochran, W. G. and Cox, G. M. (1950), *Experimental Designs*, John Wiley, New York.

Dey, A. and Mukerjee, R. (1999), *Fractional Factorial Plans*, John Wiley, New York.

Hale-Bennett, C. and Lin, D. K. J. (1997), "From SPC to DOE: a Case Study at Meco, Inc.," *Quality Engineering*, 9, 489–502.

Shoemaker, A. C., Tsui, K. L. and Wu, C. F. J. (1991), "Economical Experimentation Methods for Robust Design," *Technometrics*, 33, 415–427.

Wu, C.F.J. and Hamada, M. (2000), *Experiments: Planning, Analysis and Parameter Design Optimization*, John Wiley, New York.

CHAPTER IV

SELC : SEQUENTIAL ELIMINATION OF LEVEL COMBINATIONS BY MEANS OF MODIFIED GENETIC ALGORITHMS

4.1 Introduction

In many scientific problems, the goal is to select an optimal candidate from a large pool of potential candidates. Genetic Algorithms (GAs) are a popular optimization technique when searching for global optimums in a large candidate or design space. A modification of GAs, called SELC, is proposed in this paper, which outperforms classical GAs in several practical situations. Here we present two scenarios where the SELC method can be useful. The first example is in the context of computer experiments, while the second example arises in pharmaceutical industries.

In the last fifteen years many phenomena that could only be studied using physical experiments can now be studied by computer experiments. In a computer experiment, a deterministic output, $y(\mathbf{x})$, is computed for each set of input variables, $\mathbf{x}$, using numerical methods that are implemented by (complex) computer codes (Santner et al., 2003). In such cases, the complex function can be thought of as a "black box" and the proposed SELC method can be used to obtain the optimal settings efficiently. In Section 5, we illustrate how the SELC method can be efficiently used for a "black box" type problem.

The SELC method also has potential applications in the pharmaceutical industry. Within the past thirty years, technologies have been developed to explore and synthesize vast numbers of chemical entities. This technology, known as combinatorial chemistry, has been widely applied in the pharmaceutical industry, and is gaining interest in other areas of chemical manufacturing (Leach and Gillet, 2003, Gasteiger and Engel, 2003). In general, combinatorial chemistry identifies molecules that can be easily joined together, and employs

robotics to physically make each molecular combination. Depending on the initial number of molecules, the number of combinations can be extremely large. For example, consider a core molecule onto which various reagents can be theoretically added to three locations. If one hundred reagents can be added at each location on the core, then one million products can potentially be synthesized. In the pharmaceutical industry, combinatorial chemistry has been used to enhance the diversity of compound libraries, to explore specific regions of chemical space (i.e. focused library design), and to optimize one or more pharmaceutical endpoints such as target efficacy, or ADMET (absorption, distribution, metabolism, excretion, toxicology) properties (Rouhi, 2003). While it is theoretically possible to make a large number of chemical combinations, it is generally not possible to follow-up on each newly synthesized entity. Instead of synthesizing all possible molecular combinations, combinatorial libraries are computationally created and evaluated using structure-based models. (For this purpose, specialized software uses "black box" type functions.) In addition, chemists look for reagent combinations that are known to produce undesirable compounds, and attempt to avoid these combinations during synthesis. Using these constraints, a subset of promising reagents is selected to generate a combinatorial library. By construction, the SELC is a natural fit for searching for optimal molecules in combinatorial chemistry.

These real-life scenarios can be thought of as large dimensional design of experiment problems where the challenge is to identify the optimal design settings. Statistical design and analysis of experiments is an effective and commonly used tool in scientific and engineering investigation to understand and/or improve a system. Identifying important factors and choosing factor levels are among the first and most fundamental issues facing an experimenter. But, when confronted with a large number of important factors, designing an experiment can be difficult. Classical experimental design relies heavily on algebraic properties such as orthogonality. However, orthogonality does not allow the flexibility to accommodate all kinds of promising follow-up runs, which, in turn, makes finding suitable designs for large-scale problems difficult, particularly when the factors have more than two levels.

The use of high-fidelity computer simulations of physical phenomena (Bates et al., 1996)

has stimulated new research into ways in which experimental design can be applied to such problems. One technique, motivated by design of experiments, was introduced by Wu, Mao, and Ma (1990) (hereafter abbreviated as WMM) known as Sequential Elimination of Levels (SEL). The idea of SEL is opposite to that of the "greedy algorithm": instead of focusing on factor levels that improve the response, SEL focuses on those levels that worsen the response. Based on this idea, SEL eliminates one level of each factor in each sequence of the experiment. However, this kind of marginal analysis does not perform well in the presence of interactions, which is generally the case for high dimensional response surfaces. In this paper, the idea of SEL is extended to accommodate situations where important interactions are present. But, to make this accommodation, we must abandon follow-up designs that are orthogonal. Instead, a modified version of Genetic Algorithms (GAs) will be used to determine subsequent design points.

GAs have most often been viewed from a biological perspective. The metaphors of natural selection, cross breeding, and mutation have been helpful in providing a framework to explain how and why GAs work. Thus, it makes sense that most practical applications of GAs are rooted in the context of optimization. In an attempt to understand how GAs function as optimizers, Reeves and Wright (1999) considered GAs as a form of sequential experimental design. Recently, GAs have been used quite successfully in solving statistical problems, particularly for finding near optimal designs (Hamada et al. 2001, Heredia-Langner et al. 2003, 2004).

This chapter is organized as follows. In Section 2, we review the idea of SEL and classical GAs. A new version of SEL, called SELC is proposed in Section 3. The Bayesian model selection, which can be used in this process, is discussed in the Appendix. Behavior of the SELC algorithm is discussed in Section 4. In Section 5, the proposed algorithm is applied to three functions, including one from Shekel's family, and the performance of this search methodology is investigated via simulations. In Section 6, SELC is used for identifying potentially good compounds for synthesization in a pharmaceutical industry. Some concluding remarks are given in Section 7.

4.2 Review : Sequential Elimination of Levels and Genetic Algorithms

SEL : WMM proposed their search method, based on orthogonal arrays, as follows :

1. Start with an appropriate orthogonal array.
2. For each factor, *eliminate* those level(s) with the worst mean value(s) of the performance measure computed from the current array.
3. Choose an orthogonal array (typically of a smaller size) for the *remaining* levels, and replace the array in step 1 with the new array.
4. Conduct another experiment on the *new* array.
5. Repeat steps $2 - 4$ if necessary.

In step 1, if the mean is replaced by another descriptive statistic z (for example, minimum), the method is called SEL(z).

The main drawback of SEL is that its method of search is too restrictive for many optimization problems. First, for experiments that contain important interactions, the SEL method is not optimal because it eliminates individual levels of each factor without considering interactions. Hence, SEL can blindly eliminate a factor level that is required for the optimal run of the experiment. Second, SEL requires that subsequent experiments follow an orthogonal array. As mentioned previously, our modification of the SEL will prevent it from using an orthogonal array. In addition, orthogonal arrays are not flexible enough to handle complex response surfaces. To overcome this problem, we have developed a modified GA to determine subsequent design points.

GAs : Before describing the novel approach to improve SEL, we shall briefly review GAs (Holland, 1975). GAs are stochastic optimization tools that work on "Darwinian" models of population biology and are capable of obtaining near-optimal solutions for multivariate functions without the usual mathematical requirements of strict continuity, differentiability, convexity or other properties. The algorithm attempts to mimic the natural evolution of a population by allowing solutions to reproduce, creating new solutions, and to compete

for survival. The idea of GAs is to get "better solutions" using "good solutions", and the algorithm process is as follows:

1. *Solution representation :* For problems that require real number solutions, a simple binary representation is used where unique binary integers are mapped onto some range of the real line. Each bit is called a *gene* and this binary representation is called *chromosome*.

   Once a representation is chosen, the GA proceeds as follows. A large initial population of random candidate solutions is generated; these are then continually transformed following steps 2 and 3.

2. *Select* the best and eliminate the worst solution on the basis of a fitness criterion (e.g., higher the better for a maximization problem) to generate the next population of candidate solutions.

3. *Reproduce* to transform the population into another set of solutions by applying the genetic operations of "crossover" and "mutation".

   (a) **Crossover :** A pair of binary integers (chromosomes) are split at a random position and the head of one is combined with the tail of other and vice-versa.

   (b) **Mutation :** The state (0 or 1) of a randomly chosen bit is changed. This helps the search avoid being trapped into local optima.

4. *Repeat* steps (2) and (3) until some convergence criterion is met or some fixed number of generations has passed.

This algorithm has been shown to converge by Holland (1992), who first proposed this procedure in its most abstract form and discussed it in relation to adaptive and nonlinear systems.

4.3 SELC : Sequential Elimination of Level Combinations

The main drawback of SEL is that its search is too restrictive. This method eliminates a level on the basis of marginal means which can be affected by the presence of interactions.

In order to overcome this drawback, we propose eliminating level combinations instead of just a single level. This modification is capable of capturing important interactions and provides more flexibility in the choice of follow-up design points. Our modification of SEL, Sequential Elimination of Level Combinations (SELC), incorporates the fundamentally new ideas of the forbidden array and weighted mutation. In Section 4, we shall see how these two novel concepts, motivated by the ideas of Design of Experiments, make the search algorithm much more efficient than classical GAs.

Recall that by the *effect hierarchy* principle (Wu and Hamada, 2000), two-factor interactions are more important than higher order interactions. In SELC, we employ this principle by allowing the algorithm to identify important interactions with respect to the optimization problem. Here we propose to eliminate those factor settings which have the same level combinations as that of the *worst* one for two factors. For larger dimensions, third or higher order tuples may need to be considered to narrow the search space. The worst observed runs are stored in the *forbidden array* as the search procedure continues. New experiments are conducted with runs suggested by the SELC algorithm, which uses the idea of GAs, and promising level settings for a new run are achieved by using better runs from the previous experiments. Before formally defining the SELC algorithm, we define the concepts of the *forbidden array* and *weighted mutation*, both required by the algorithm. We end this section with a constructed example to illustrate the SELC algorithm.

Forbidden array: In some situations, prior knowledge is available about certain factor level combinations that lead to undesirable results. Consider the introductory combinatorial chemistry example. In this setting, chemists can often identify runs (i.e. new molecules), based on their scientific knowledge and prior experience, which are not worth creating in the laboratory. These runs can be placed into the forbidden array before initializing the SELC algorithm.

In the absence of prior knowledge, the SELC is initialized with an orthogonal design. The data from this initial experiment are then used to suggest run(s) that are not optimal. These run(s) are then placed into the forbidden array.

In subsequent steps of the experiment, the worst run(s) are chosen with probability

governed by a "fitness" measure (i.e., value of y) and are stored in the forbidden array. Furthermore, we specify the *strength* and *order* of the forbidden array. The number of runs placed into the forbidden array at each sequence of the experiment defines the array's *strength*. More specifically, a forbidden array of strength s contains the level combinations of the s worst runs of the experiment at each stage of the iterations. In addition, the runs stored in the forbidden array define a set of level combinations that will be prohibited from subsequent runs of the experiment. The number of level combinations that are prohibited from subsequent experiments defines the *order* of the forbidden array. A forbidden array of order k implies that any combinations of k or more levels from any array in the forbidden array will be prohibited from being used in subsequent runs of the experiment. Thus, as the order decreases, the number of forbidden design points increases. Consequently, the forbidden array is the generating set of all runs which are forbidden by SELC.

For example, consider an experiment in which the goal is to maximize a response. Suppose the experiment has four factors, each at three levels (0, 1, and 2) and we choose a forbidden array with strength 1 and order 2. Further, suppose that the minimum value of $E(y)$ occurs when all factors are set to 0, and this design point is run during the experiment. When this run is placed into the forbidden array, it will prevent any design points with *two* or more factors set to level 0 (order=2). Note that only one member will be added to the forbidden array at each step (strength = 1).

Here the special case of $k = 1$ corresponds to the SEL method of WMM. Also, $s = 1$ corresponds to SEL(mini) of WMM. However, unlike the SEL-approach, the choice of worst run is probabilistic in SELC. In Section 6, we will illustrate how the choice of strength affects the performance of the search procedure.

After constructing the forbidden array, SELC starts searching for better level settings. The search procedure is motivated by GAs. The first step, as discussed in the review of GAs, is *solution representation*. Here the runs are viewed as chromosomes. For an m-level factor, the levels are denoted by $0, 1, \ldots, m-1$. For example, for a 3^4 experiment, the design points(chromosomes) would take the form $(0,0,0,0)$, $(0,0,0,1)$,..., $(2,2,2,2)$. *Unlike classical GAs, the chromosomes are not required to be binary arrays.* Next we identify, with

probability proportional to the "fitness", i.e. the value of y, the best runs to produce offspring of the next generation. After the good candidates are identified, they *reproduce* to generate potentially better candidates. In SELC, crossover is performed in the usual way, as explained in Section 2, but a modification is proposed for mutation.

Weighted mutation: In a generic GA, genes mutate with an equivalent specified probability. Hence, the mutation rate does not incorporate other information gathered from prior knowledge about the system. For the SELC, we propose the use of prior information for generating mutation probabilities. For instance, suppose we know that the factor, F, has a significant main effect and *no* significant two-factor interactions. Then, we will change the level of this factor to a new level, l, with probability p_l, where

$$p_l \propto \overline{y}(F = l). \tag{1}$$

Next, suppose that factors F_1 and F_2 have a significant interaction. Then, the mutation should have a joint probability on F_1 and F_2. That is, the mutation will occur if either F_1 or F_2 is randomly selected. Factor F_1 will be set to level l_1 and factor F_2 to level l_2 with probability q_{l_1,l_2}, where

$$q_{l_1 l_2} \propto \overline{y}(F_1 = l_1, F_2 = l_2). \tag{2}$$

If the selected factor does not have significant main effects or interactions, then its value is changed to any admissible levels with equal probability. Note that if the aim is to minimize $E(y)$, then the probabilities in (1) and (2) should be inversely proportional to $\bar{y}$.

A linear regression model can be used to identify the significant effects. But, a better, more time consuming approach is to consider a Bayesian variable selection strategy which is discussed in the Appendix. This method is used in the analysis illustrated at the end of this section.

Starting Design : The starting design is an orthogonal array, which allows us to efficiently estimate factor effects used in the process of weighted mutation. However, as the search proceeds, unlike SEL, the orthogonal structure of the design matrix will not be retained. Nonorthogonality is justified because the follow-up designs should be more flexible than the starting one, utilizing the information already at hand.

The SELC algorithm:

Initialize the design with an appropriate orthogonal array.

1. Conduct the experiment.
   - Stop when the stopping criterion is achieved. (See below).
2. Construct the *forbidden array* and choose its strength and order.
3. Generate b new *offspring*.
   - *Select* offspring for reproduction with probability proportional to their "fitness."
   - *Crossover* the offspring.
   - *Mutate* the positions using *weighted mutation*.
4. Check eligibility. An offspring is *eligible* if it is not prohibited by any of the members of the forbidden array. If an offspring is ineligible, then discard and generate another new offspring.
5. If $b = 1$ and more than one offspring were generated, then randomly select one offspring for the experiment.

Depending on the situation, the SELC method can be *fully* ($b = 1$) or *batch* ($b = b$) *sequential*. For fully sequential SELC, a new eligible offspring is generated in each iteration and the experiment is conducted. For batch sequential SELC, a new set of eligible offspring is generated in each iteration and the experiment is conducted. Depending on the application, either fully sequential or batch sequential may be more suitable. For example, in combinatorial chemistry, a batch sequential SELC is more appropriate.

Fully sequential SELC method is used in the illustrative example of this section as well as in Section 6. On the other hand, a batch sequential SELC method is used in Section 5. In the later case, a fixed number of offspring are generated before running the experiments to evaluate their performance.

Stopping Rules : The stopping rule is subjective and depends on progression of the algorithm and experimental constrants. As the runs are added, the experimenter can decide,

in a sequential manner, whether significant progress has been made towards optimization. Sometimes a target value or near optimum is predetermined for the experiment. Once the target is attained, the search can be stopped. But, typically, the number of experiments is limited by the resources at hand. This is often the case for the combinatorial chemistry example discussed in the Introduction. Examples of Section 5 illustrate a situation in which an experiment is limited by number of runs.

To illustrate the SELC method, consider a hypothetical experiment with 9 factors (denoted by A-I) each at 3 levels. In this example (and throughout this paper), we use the *linear-quadratic system* for coding linear and quadratic effects (Wu and Hamada, 2000) in order to eliminate correlation among a factor's linear and quadratic components. The linear-quadratic coding is expressed as follows :

$$\begin{array}{cccc} \textit{level} & & \textit{linear} & \textit{quadratic} \\ 0 & & -1 & 1 \\ & \longrightarrow & & \\ 1 & & 0 & -2 \\ 2 & & 1 & 1 \end{array}$$

The response is generated from the following model :

$$\begin{aligned} y = 2 + (&A + 2B - 3C + D + 2E - 2A^2 + 2B^2 + 1.5C^2 \\ &-3AC + 2.5AE - BF - 2CG + DGI)^2 + \epsilon, \end{aligned}$$

where ϵ is the standard normal error. In this analysis, we only consider the linear and quadratic effects and linear-by-linear interactions. Our aim is to find a setting for which the expected value of y is maximized.

The starting design for the SELC is an orthogonal array, 9 columns of an $OA(243, 3^{20}, 3)$. Without having a prior knowledge about the unfavourable runs, here we use a forbidden array with $s = 1$ and $k = 6$, and use a weighted mutation with the Bayesian variable selection strategy. After choosing the first member of the forbidden array, the search for better level settings is continued via crossover and weighted mutation. Upon computing the posterior probabilities of C and BC, we find that these are much larger than the posterior probabilities of the other effects. According to the weighted mutation scheme, if factor B

or C is randomly selected for mutation, we must evaluate the $q_{l_1 l_2}$'s in (2). The $q_{l_1 l_2}$'s are given below :

Factors	$C=1$	$C=2$	$C=3$
$B=1$	0.0526	0.0556	0.1212
$B=2$	0.0973	0.0524	0.0865
$B=3$	0.2933	0.1368	0.1043

After generating the new offspring, we check for eligibility and the the search continues. In this example, using the fully sequential version of SELC, the search was stopped after 400 runs. The maximum value of y was 679.68, which corresponds to the level setting of the third best design point. Note that the SELC algorithm found this near optimum design point by evaluating only 2.03% of all possible combinations.

4.4 A Justification Of Crossover And Weighted Mutation

Steps of crossover and weighted mutation may be better understood through the following analysis. Consider the problem of maximizing $K(\mathbf{x})$, $\mathbf{x} = (x_1, \ldots, x_p)$, over $a_i \leq x_i \leq b_i$. Instead of solving the p-dimensional maximization problem

$$\max\Big\{K(\mathbf{x}) : a_i \leq x_i \leq b_i, i = 1, \ldots, p\Big\}, \tag{3}$$

the following p one-dimensional maximization problems are considered,

$$\max\Big\{K_i(x_i) : a_i \leq x_i \leq b_i, i = 1, \ldots, p\Big\}, \tag{4}$$

where $K_i(x_i)$ is the ith marginal function of $K(\mathbf{x})$,

$$K_i(x_i) = \int K(\mathbf{x}) \prod_{j \neq i} dx_j \tag{5}$$

and the integral is taken over the intervals $[a_j, b_j], j \neq i$. If the x_i in (3) and (4) can take only a finite number of values (discrete x_i), the integral in (5) is replaced by a finite sum. Let x_i^* be a solution to the ith problem in (4). The combination $\mathbf{x}^* = (x_1^*, \ldots, x_p^*)$ may be proposed as an approximate solution to (3). A sufficient condition for $\mathbf{x}^*$ to be a solution of (3) is that $K(\mathbf{x})$ can be represented as

$$K(\mathbf{x}) \quad = \quad \psi\Big(K_1(x_1), \ldots, K_p(x_p)\Big) \tag{6}$$

and

$$\psi \text{ is nondecreasing in each } K_i.$$

A special case of (6), which is of particular interest to statisticians, is

$$K(\mathbf{x}) = \sum_{i=1}^{p} \alpha_i K_i(x_i) + \sum_{i=1}^{p} \sum_{j=1}^{p} \lambda_{ij} K_i(x_i) K_j(x_j). \tag{7}$$

If λ_{ij} is nonzero, then SEL will have difficulty finding the optimal solution. However, SELC is more flexible and is better suited to find the optimal solution.

While the SEL method emphasizes on orthogonal arrays, SELC does not. The basic nature of GAs does not allow us to retain the orthogonal structure of the design. Though orthogonal arrays are good for estimating the factorial effects, they are not available for every combination of factor levels and for every run size. GAs do not require orthogonality and hence are more flexible in exploring new design points. This flexibility enhances the chance of getting the best setting in relatively fewer runs. If the response surface is very smooth, then any standard design and analysis should find the optimal settings. However, for many problems the response surface is not smooth. For instance, if the surface is undulated with local maxima and minima, the SELC method can perform well. The random nature of the GA-type search explores the whole surface rapidly, while the weighted mutation uses prior knowledge about the surface to wisely direct the search.

The convergence of classical GAs was provided by Holland (1975) using the concept of schema. The SELC method makes a significant amount of modification to classical GAs and it is not obvious that the modifications proposed meet the requirements for convergence in Holland's paper. However, the simulation studies provided in the next section are quite convincing about the convergence.

4.5 Examples

We investigate the performance of SELC via several diverse simulations. Three different "ill-behaved" functions are considered and the effects of the fine tunings are illustrated through a variety of examples. For all these examples, we have the following settings. For crossover, after choosing one position randomly, parent chromosomes are split at that

position and the left fragment of the first parent chromosome is combined with the right fragment of the second parent chromosome to produce the first offspring. Then, mutation-locations are chosen randomly for each offspring and weighted mutation is performed as described in Section 3. For comparison, some simulations have been done with unweighted mutation which allows the level of the factor to be changed randomly to any other admissible level. For all simulations, the population size is 20 which corresponds to the batch size 20 (i.e. $b = 20$). For each of these examples, we assume there is no prior knowledge about undesirable runs. Hence, we initialize the forbidden array using information gathered from the initial orthogonal array.

Example 1 : Shekel 4 function (SQRIN)

The function

$$y(x_1, \ldots, x_4) = \sum_{i=1}^{m} \frac{1}{\sum_{j=1}^{4}(x_j - a_{ij})^2 + c_i}$$

is known as Shekel's function (Dixon and Szego, 1978), where the quantities $\{a_{ij}\}$ and $\{c_i\}$ are given in Table 7. The region of interest is $0 \leq x_j \leq 10$ and only integer values are considered. This function is one of the "black box" functions of computer experiments, discussed in the Introduction.

Table 7: Coefficients for Shekel's function ($m = 7$)

i	$a_{ij}, j = 1, \ldots, 4$				c_i
1	4.0	4.0	4.0	4.0	0.1
2	1.0	1.0	1.0	1.0	0.2
3	8.0	8.0	8.0	8.0	0.2
4	6.0	6.0	6.0	6.0	0.4
5	3.0	7.0	3.0	7.0	0.4
6	2.0	9.0	2.0	9.0	0.6
7	5.0	5.0	3.0	3.0	0.3

This set-up corresponds to an experiment with four factors each at 11 levels (i.e. the 11 integers). The starting design is an orthogonal array of 242 runs which is obtained by choosing 4 columns from the OA$(242, 11^{23})$ (Hedayat et al., 1999). In this example, unlike Section 3, Bayesian variable selection strategy was not used. In each step, Gibbs

sampling consumes significant amount of time which would make it extremely difficult to run thousands of simulations. Instead, regression analysis is used to identify the important factors (at 5% level of significance). Forbidden arrays of order 3 are considered because order 1 or 2 becomes too restrictive for this problem by forbidding too many runs (and also, the results are not satisfactory). The results are compared with those of a random search and with simple GA.

Table 8 summarizes the results. "Random Search" corresponds to a design where all runs are selected randomly. "Random Followup" stands for a design, where the search begins with the same starting design and follow-up runs are selected randomly. "Genetic Algo" stands for a classical GA where the runs are looked upon as chromosomes and crossovers and mutations are done in the usual way. Recall that GA corresponds to a special case of SELC with strength 0 and unweighted mutation. "SELC (No Forbiddance)" refers to weighted mutation only, because in this case, forbidden array is set to be empty (i.e., strength = 0). On the other hand, "SELC (Unweighted Mutation)" refers to forbiddance only. Here unweighted mutation is performed instead of weighted mutation. Finally, "SELC (Weighted Mutation)" refers to the SELC method proposed in Section 3.

The performance of the search algorithm is measured by its ability to find the global maximum. We also include its performance on finding second through fifth best values, because these five values stand apart from the others on the response surface.

In the first simulation, the search is stopped after 1000 runs, which is 6.83% of all possible 11^4 runs (Figure 4.5). As seen from Figure 4.5, GA performs better than random searches and SELC performs better than GA. The values for "SELC (No Forbiddance)" show the beneficial effect of weighted mutation (here strength of the forbidden array is 0) and the values for "SELC (Unweigthed Mutation)" show the beneficial effect of forbidden array. "SELC (No Forbiddance)" finds the maximum in 53% of the cases, as opposed to 48% of GA. On the other hand, "SELC (Unweigthed Mutation)" has success rate 55.5%. Finally, when the power of both forbidden array and weighted mutation are explored, SELC performs satisfactorily in 57.8% of the times. The most benefits are achieved by considering the weighted mutation. This effect is even more pronounced in the next example.

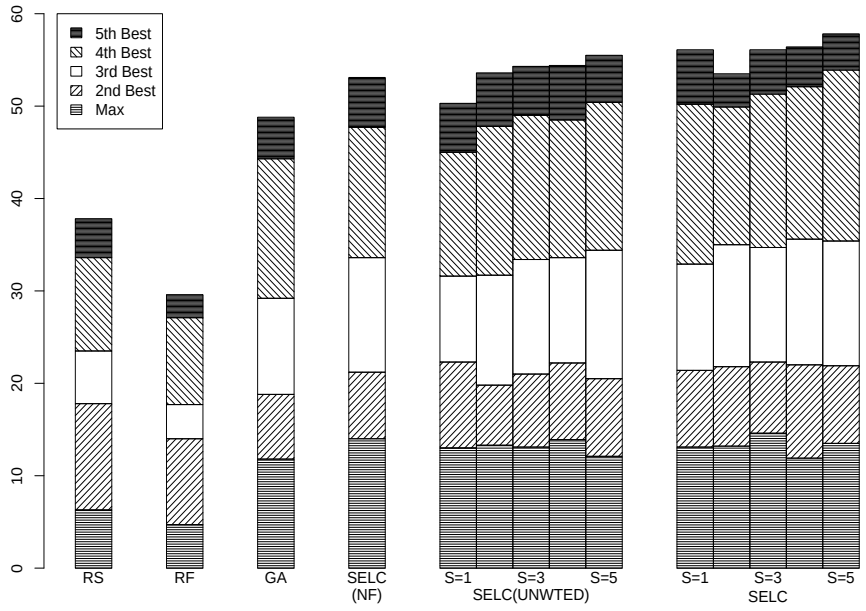


Figure 1: Example 1 : % of success in identifying global maximum for different methods (run size = 1000) [RS = Random Search, RF = Random Followup, GA = Genetic Algo, SELC(NF) = SELC(No Forbiddance), SELC(UNWTED) = SELC(Unweighted Mutation), SELC = SELC(Weighted Mutation), S = Strength]

As the strength of the forbidden array increases, the power of the search algorithm also increases. However, the strength cannot be increased arbitrarily, because it will then prohibit too many design points from being considered. It should also be noted that the improvement of the performance of SELC with the increment of the strength is not so prominent for the same function when smaller run sizes are considered. In the second case, the search is stopped after 700 runs, and the improvements are not as significant. For Shekel 4 function, evolutionary algorithms would take larger runs to reap the benefits.

Table 8: Example 1 : % of success in identifying global maximum for different methods based on 1000 simulations (run size = 1000 and 700).

Strength		Max	2nd best	3rd best	4th best	5th best	Total
1000 RUNS							
Random Search		6.3	11.5	5.7	10.1	4.2	37.8
Random Followup		4.7	9.3	3.7	9.4	2.5	29.6
Genetic Algo		11.8	7.0	10.4	15.1	4.5	48.4
SELC (No Forbiddance)		14.0	7.2	12.4	14.1	5.4	53.1
SELC (Unweighted Mutation)	1	13.0	9.3	9.3	13.4	5.3	50.3
	2	13.3	6.5	11.7	16.1	5.8	53.4
	3	13.1	7.9	12.4	15.6	5.3	54.3
	4	13.9	8.3	11.4	14.9	5.9	54.4
	5	12.1	8.4	13.9	16.0	5.1	55.5
SELC (Weighted Mutation)	1	13.1	8.3	11.5	17.3	5.9	56.1
	2	13.2	8.6	13.2	14.9	3.6	53.5
	3	14.6	7.7	12.4	16.6	4.8	56.1
	4	11.9	10.1	13.6	16.5	4.3	56.4
	5	13.5	8.4	13.5	18.5	3.9	57.8
700 RUNS							
Random Search		4.2	9.0	4.0	9.2	4.1	30.5
Random Followup		3.0	6.8	3.0	5.1	2.4	20.3
Genetic Algo		5.8	5.6	6.0	9.2	3.3	29.9
SELC (No Forbiddance)		5.4	4.7	7.2	11.3	4.8	33.4
SELC (Unweighted Mutation)	1	5.8	6.1	6.0	9.9	4.9	32.7
	2	6.4	4.3	4.6	10.1	5.7	31.1
	3	7.1	4.3	5.9	8.7	5.2	31.2
	4	7.6	4.1	6.0	11.5	4.7	33.9
	5	5.2	4.6	6.6	10.2	4.9	31.5
SELC (Weighted Mutation)	1	6.3	5.5	6.9	11.5	4.0	34.2
	2	6.6	4.9	7.2	10.6	3.1	32.4
	3	7.2	4.6	9.6	10.6	4.1	36.1
	4	5.9	5.9	7.0	10.7	3.3	32.8
	5	5.9	4.7	8.5	10.3	4.1	33.5

Example 2

Consider the function

$$y(x_1, \ldots, x_4) \quad = \quad 1 + \{\beta' x + (\gamma' x)^2 + \eta' x \times \tau' x\}^2,$$

where the parameters are given in Table 9. The region of interest is $0 \leq x_j \leq 10$ and only integer values are considered. This choice is motivated by discussions in Section 4, especially (4.5).

Table 9: Coefficients for the function in Example 2

β	γ	η	τ
1	-3	2	-5
-2	-4	-10	0
2	5	2	-5
-1	-6	4	0

As in Example 1, this set-up also corresponds to an experiment with four factors each at 11 levels. The simulations are done with two starting designs: orthogonal array of size 121 and 242, which are obtained by choosing four columns from $\text{OA}(121, 11^{12})$ and $\text{OA}(242, 11^{23})$ respectively (Hedayat et al., 1999). The results are summarized in Table 10 and also in Figure 4.5. The simulations are done for a total of 300, 500 and 1000 runs.

GA performs much better than random search. This example shows that forbiddance need not always enhance the performance. In fact, without weighted mutation, forbiddance alone (i.e., "SELC (Unweighted Mutation)") can perform worse than GA. This means that good runs are located in the "neighborhood" of bad runs and the response surface $y(x_1, \ldots, x_4)$ is very undulated. However, weighted mutation significantly improves the performance of SELC. The main advantage of using SELC is that it uses prior information to direct the GA, thus finding a near optimum more quickly. This effect is clearly demonstrated for smaller runs, namely, with total run size 300 and 500. If the search is continued long enough, this gap will be narrowed and SELC may not perform much better than GA. Consider the first case where the starting design is an orthogonal array of size 121. For 300 runs, GA finds the maximum in 15% of the cases whereas SELC (Weighted Mutation) finds

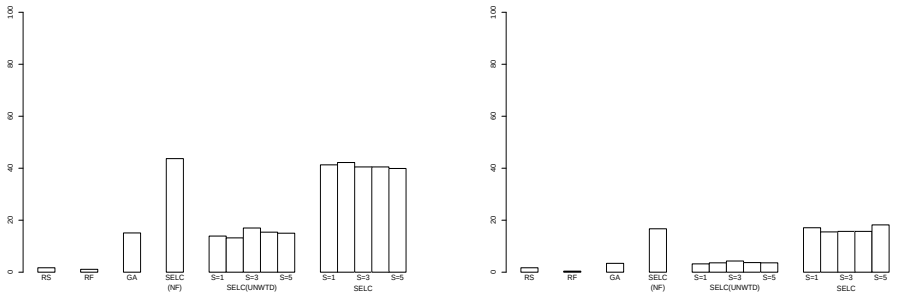
Table 10: Example 2 : % of success in identifying global maximum for different methods based on 1000 simulations

	Strength	121-Run Design			242-Run Design		
Total Run Size		300	500	1000	300	500	1000
Random Search		1.7	3.6	7.0	1.7	3.6	7.0
Random Followup		1.1	2.5	5.7	0.4	2.4	4.6
Genetic Algo		15.1	39.5	79.7	3.4	28.9	79.5
SELC (No Forbiddance)		43.7	76.7	97.8	16.7	68.3	97.5
SELC (Unweighted Mutation)	1	13.9	39.9	80.9	3.2	30.9	77.2
	2	13.2	38.9	83.4	3.6	30.1	79.6
	3	17.0	41.4	82.3	4.3	31.4	78.4
	4	15.4	40.2	81.1	3.7	28.3	76.9
	5	15.0	44.1	81.5	3.6	29.5	78.5
SELC (Weighted Mutation)	1	41.3	76.9	97.3	17.1	67.4	98.4
	2	42.2	76.5	97.3	15.5	65.4	96.8
	3	40.5	75.1	98.2	15.7	67.6	97.8
	4	40.5	75.9	98.0	15.7	69.6	98.0
	5	39.9	73.9	97.9	18.2	66.1	96.9

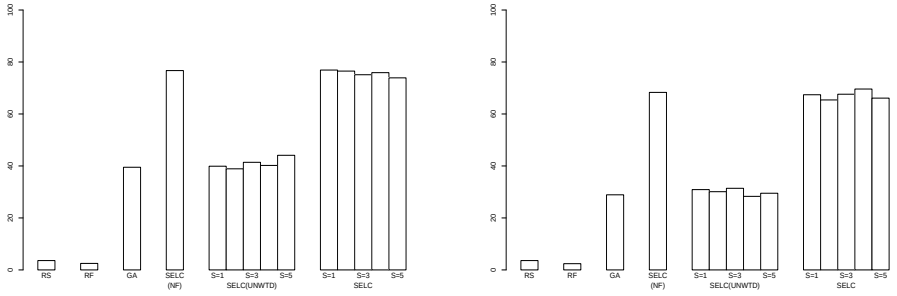
it in more than 40% of the cases. For 500 runs, the values are 40% and 75%, respectively. Finally for 1000 runs, the success rates are 80% and 97%, respectively. The ratio of the success rate decreases as the run size increases, which is not surprising because these kind of evolutionary algorithms eventually find the near optimal solution, if they are run long enough. However, SELC finds the optimal quickly.

For the second case, the starting design is an orthogonal array of size 242. Here, for a total run size 300, the evolutionary-type algorithms are not expected to perform well because only 58 follow-up runs are available. Even with these few follow-up runs, SELC (Weighted Mutation) finds the maximum in more than 15% of the cases. With larger run sizes, the performance of both GA and SELC improves, with SELC performing significantly better than GA.

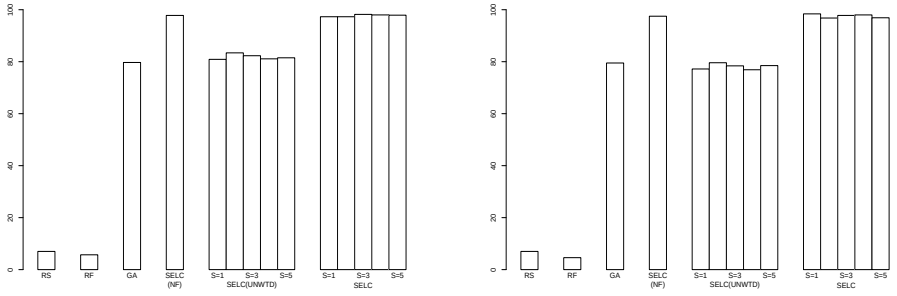
The overall pattern of the performance of SELC for both starting designs are similar. Also for 1000 runs, the effect of starting design diminishes and the success rates are very close for both cases. Note that, for this example, starting with a 121-run design, with only 300 evaluations (2.05% of all possible 11^4 runs), SELC finds the global maximum in more than 40% of the cases.



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 300



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 500



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 1000

Figure 2: Example 2 : % of success in identifying global maximum for different methods

Example 3

Levy and Montalvo (1985) provide the following function:

$$y(x_1,\ldots,x_n) = \sin^2\left\{\pi\left(\frac{x_i+2}{4}\right)\right\} + \sum_{i=1}^{n-1}\left(\frac{x_i-2}{4}\right)^2\left\{1+10\sin^2\left(\pi\left(\frac{x_i+2}{4}\right)+1\right)\right\} + \left(\frac{x_n-2}{4}\right)^2\left\{1+\sin^2\left(2\pi\left(x_n-1\right)\right)\right\}.$$

Here $n = 4$ and only integer values of x_i's ($0 \leq x_i \leq 10$) are considered. This again corresponds to an experiment with four factors each at 11 levels. Results are summarized in Table 11. Here the performance of SELC is quite similar to that of Example 2. Note that the analytic nature of the test function is quite different from that of the previous two examples. It is a standard test function in global optimization literature and is presented here to demonstrate the satisfactory performance of the SELC method over a variety of test functions.

Table 11: Example 3 : % of success in identifying global maximum for different methods based on 1000 simulations

Strength		121-Run Design			242-Run Design		
Total Run Size		300	500	1000	300	500	1000
Random Search		5.8	9.3	18.4	5.0	9.3	18.4
Random Followup		2.9	7.7	15.5	2.9	7.7	15.5
Genetic Algo		16.8	43.1	80.7	2.9	33.3	81.8
SELC (No Forbiddance)		30.3	62.2	94.5	5.9	50.6	93.8
SELC (Unweighted Mutation)	1	17.6	43.1	84.5	2.9	31.6	82.2
	2	16.7	42.9	84.3	3.3	32.4	82.0
	3	18.5	44.5	83.5	4.7	33.6	83.4
	4	21.2	44.1	83.9	3.4	33.4	81.9
	5	16.6	47.5	83.9	3.8	34.0	84.5
SELC (Weighted Mutation)	1	28.4	66.2	94.4	6.6	45.9	93.5
	2	26.0	66.2	92.8	7.5	50.5	91.8
	3	31.1	63.5	92.2	7.2	49.6	93.7
	4	29.4	63.8	90.1	7.6	46.8	91.2
	5	31.9	65.3	86.1	7.1	46.9	91.3

Examples 1 and 3 are from standard test functions in the global optimization literature. By closely examining those functions, one may have some idea about the location of the

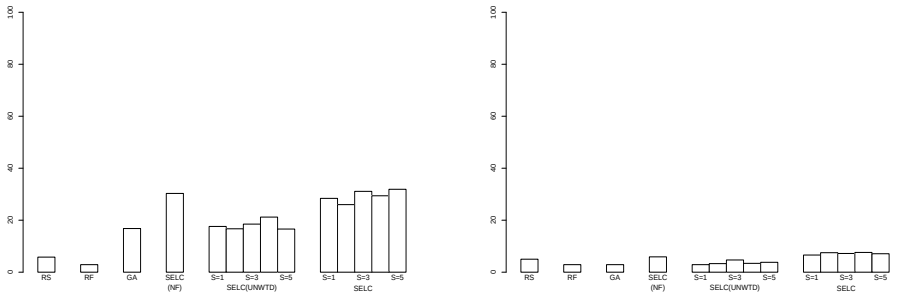
global maximum and may be able to save some computations. However, in many real life examples, (e.g., computer experiments) the analytic form of the function is either unknown or very complicated. In these situations, the function can be thought of as a "black box" and SELC method should perform well.

4.6 Application

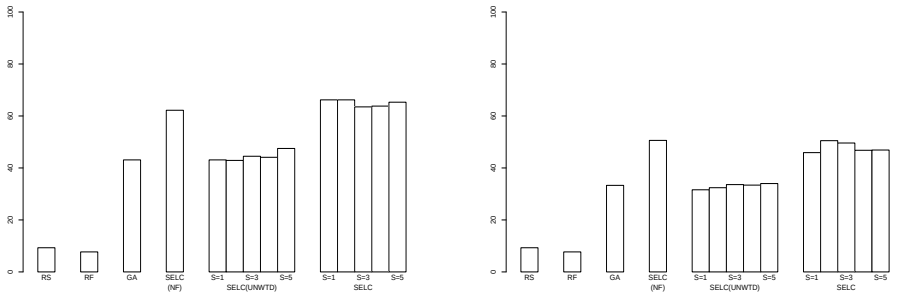
The SELC method was applied to a combinatorial chemistry problem where a combination of reagents was desired to maximize target efficacy. For this example, target efficacy is measured by a compound's percent inhibition of activity for a specific biological screen. For this screen, a percent inhibition value of 50 or greater is an indicator of a promising compound. And, percent inhibition values of 95 or greater have a high probability of exhibiting activity in confirmation screening.

Consider a core molecule onto which reagents can be added to three locations, denoted by A, B, and C. In this example, the desired compound space included two reagents at position A, 10 reagents at position B, and 14 reagents at position C. In total, the compound space contained 280 ($= 2 \times 10 \times 14$) possible chemical entities. The reagents in this application can be though of as different levels of the factors (i.e. positions) and are denoted by integers, 1, 2, etc.. In this example, 208 of the 280 chemical entities were actually created without the assistance of the SELC algorithm. To show the algorithm's benefits to the combinatorial chemistry group, we applied the SELC to this problem under the hypothetical constraint that resources were limited to creating only 25 compounds.

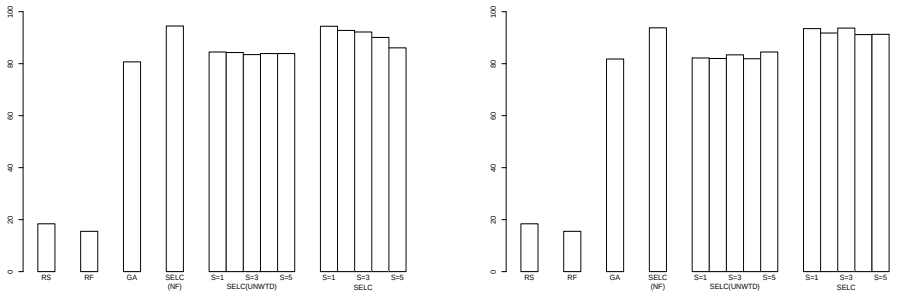
Based on prior scientific knowledge, some combinations of reagents for this experiment were known to yield unfavorable percent inhibition values. These combinations of reagents were used to focus the initial starting design, and were placed into the forbidden array prior to the experiment. Tables 12 and 13 present the relative frequency of occurrence of the individual levels of factors B and C, respectively in the forbidden array. Because we were limited to creating 25 total compounds, we chose a $2 \times 2 \times 3$ orthogonal array to initialize the experiment. Using Tables 6 and 7, in conjunction with scientific guidance, the initial orthogonal array included levels 8 and 9 of Factor B, and levels 3, 4, and 8 of Factor C.



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 300



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 500



Starting Design : $OA(121, 11^4)$ Starting Design : $OA(242, 11^4)$
total run size = 1000

Figure 3: Example 3 : % of success in identifying global maximum for different methods

Table 12: Factor B

Level	1	2	3	4	5	6	7	8	9	10
Relative Freq. (in %)	3	3	26	4	29	5	10	1	5	14

Table 13: Factor C

Level	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Relative Freq. (in %)	8	7	7	4	5	4	4	3	8	5	16	11	8	8

The results from the initial orthogonal array are presented in Table 14 (upper half).

Upon completing the initial orthogonal array, subsequent design points needed to be chosen. Because not all levels of factors B and C were explored in the initial experiment, the SELC algorithm was slightly modified to enable it to explore other parts of the chemistry space. Specifically, if a factor was found to be significantly associated with an improvement in response (at 5% level), the levels of that factor received probabilities proportional to their individual association with the response. However, if a factor was not significantly associated with improvement in response, then all levels of the factor received equal probability for the weighted mutation. In this application, factor B was significantly associated with the response after the 13th compound was created. The probabilities of mutation to the levels of factor B are as follows:

$$\begin{aligned} p_8 &= \frac{24+34+63+2+5+49+83+56+14+83}{1016} \times 0.75 + \frac{1}{10} \times 0.25, \\ p_9 &= \frac{0+12+21+9+0+5}{1016} \times 0.75 + \frac{1}{10} \times 0.25, \\ p_j &= \frac{1}{10} \times 0.25, \qquad \text{for all } j \neq 8, 9. \end{aligned}$$

The denominator, 1016, is the sum of positive responses and the weights of 0.75 and 0.25 are arbitrary. The 10 levels of B account for the 1/10 in the above expression. As it was desired to maximize the target efficacy, only positive values of the response were considered in calculating p_j's. The results from the subsequent runs of the experiment can be found in Table 14 (bottom half). A fully sequential SELC method has been employed here.

All compounds run in the experiment were analyzed in a follow-up experiment where

Table 14: Combinatorial Chemistry Example

#	A	B	C	y	
1	1	8	8	24	
2	1	9	8	-23	
3	2	8	8	34	
4	2	9	8	12	
5	1	8	3	63	*
6	1	9	3	21	
7	2	8	3	2	
8	2	9	3	9	
9	1	8	4	5	
10	1	9	4	-16	
11	2	8	4	49	*
12	2	9	4	5	
13	2	8	10	83	*
14	2	3	4	65	*
15	2	1	4	107	*
16	2	2	10	49	
17	2	8	2	56	*
18	1	6	10	19	
19	2	2	4	60	*
20	2	10	10	39	
21	1	8	10	14	
22	2	6	8	90	*
23	2	6	10	64	*
24	2	1	1	-3	
25	2	2	5	63	*

their IC_{50} values [1] were determined. Compounds that were judged to be acceptable by the chemists are indicated with an asterisk in Table 14. Clearly, the SELC method succeeded in identifying a rich set of promising compounds.

4.7 Summary and Conclusions

The problem of searching for an optimal design setting in a relatively large space is not easy. The SELC method does this job efficiently. Relaxing the condition of orthogonality, GA is flexible enough to explore more design points, which enhances the chance of finding the best

[1] IC_{50} assays (assays to determine the concentration of a drug-like compound resulting in a 50% reduction in activity of a disease target) are a commonly used method for assessing drug efficacy in pharmaceutical screening regimens. Typically, these assays are performed via serial dilutions of Dimethyl Sulfoxide (DMSO) compound libraries to achieve dilutions of 2×10^7 in 100% DMSO. Subsequent to the DMSO serial dilution, assays are performed with each dilution to ascertain the IC_{50} of the compound of interest.

setting in relatively fewer runs, particularly in the presence of interaction effects. Because forbidden array can make use of prior knowledge to rule out unfavorable settings, the SELC is particularly well-suited for scientific problems in which such knowledge is available.

The by-product of SELC algorithm, discussed in the Appendix is also of interest. If there are many factors, the experimenter can get an insight by employing the Bayesian approach. The posterior probabilities identify the important factors and interactions clearly. This approach will result in a more comprehensive search of the model space. A system can have a large number of factors, of which only a handful are important. A major use of experimental design is screening, in which experimenters seek to identify significant effects (both main effects and potentially interactions) from a large set of candidate effects. The Bayesian variable selection helps in identifying the important factors and understanding the impact of a large number of factors in relatively fewer runs.

The novel idea of forbidden array and weighted mutation enables SELC to find the optimal solution more efficiently than GA. The improvement on performance, however, depends on the nature of the response surface. If the response surface is very smooth, any reasonable search algorithm should work satisfactorily. For an extremely complicated surface, almost complete enumeration might be needed irrespective of the efficiency of the search methods. For response surfaces whose ruggedness lies in between the two, SELC is expected to perform well.

4.8 References

Bates, R. A., Buke, R. J., Riccomagno, E., and Wynn, H. P. (1996), "Experimental Design and Observation for Large Systems," *J. R. Statist. Soc. B*, 58, 77-94.

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors," *Canadian Journal of Statistics*, 24, 17–36.

Chipman, H., Hamada, M., and Wu, C. F. J. (1997), "A Bayesian Variable Selection Approach for Analyzing Designed Experiments with Complex Aliasing," *Technometrics*, 39, 372–381.

Dixon, L. C. W., and Szego, G. P. (1978), "The Global Optimization Problem: An Introduction," In *Towards Global Optimization 2* (Edited by L. C. W. Dixon and G. P. Szego), 1-15. North Holland, Amsterdam.

Gasteiger, J. and Engel, T. (eds.) (2003), *Chemoinformatics : a Textbook*, Wiley-VCH, Weinheim.

George, E. I., and McCulloch, R. E. (1993), "Variable Selection Via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881–889.

Hamada, M., Martz, H.D., Reese, C.S. and Wilson, A.G. (2001), "Finding Near Optimal Bayesian Designs Via Genetic Algorithms," *The American Statistician* 55, 175-181.

Hedayat, A. S., Sloane, N. J. A., and Stufken, J. (1999), *Orthogonal Arrays : Theory and Applications*, Springer-Verlag, New York, Inc.

Heredia-Langner, A., Carlyle, W.M., Montgomery, D.C., Borror, C.M. and Runger, G.C. (2003), "Genetic Algorithms for the Construction of D-optimal Designs," *Journal of Quality Technology* 35, 28-46.

Heredia-Langner, A., Montgomery, D. C., Carlyle, W. M., Borror, C. M. (2004), "Model-Robust Optimal Designs: A Genetic Algorithm Approach," *Journal of Quality Technology*, 36, **3**, 263–279.

Holland, J. M. (1975), *Adaptation in Natural and Artificial Systems*, The University of Michigan Press, Ann Arbor.

Holland, J. M. (1992), *Adaptation in Natural and Artificial Systems*, The MIT Press, Cambridge, MA.

Leach, A. R., and Gillet, V. J. (2003), *An Introduction to Chemoinformatics*, Kluwer Academic Publishers, London.

Levy, A. V., and Montalvo, A. (1985), "The Tunnelling Algorithm for the Global Minimization of Functions," *SIAM Journal of Scientific and Statistical Computing*, 6, 15–29.

Reeves, C. L., and Wright, C. C. (1999), "Genetic Algorithms and the Design of Experiments," In Davis, L. D.; DeJong, K.; Vose, M. D. and Whitley, L. D. (1999) (Ed.), Springer-Verlag, New York, Inc. pp 207-226.

Rouhi, A. M. (2003), "Custom Synthesis for Drug Discovery," *Chemical & Engineering News*, 81 , **7**, 75–78.

Santner, T. J., Williams B. J., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, Springer-Verlag, New York, Inc.

Wu, C. F. J., Mao, S. S. and Ma, F. S. (1990), "SEL : A Search Method Based on Orthogonal Arrays," In S. Ghosh (1990), (Ed.), *Statistical Design and Analysis of Industrial Experiments*, Marcel Dekker, Inc., New York, 279 - 310.

Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, John Wiley & Sons.

APPENDIX A

DESIGN EFFICIENCY UNDER MODEL UNCERTAINTY FOR NONREGULAR FRACTIONS OF GENERAL FACTORIALS

A.1 Proof of Lemma 1

For $1 \leq j \leq m$, let 1_j be the $s_j \times 1$ vector with all elements unity, I_j be the identity matrix of order s_j, and $P_j = [\, p_j(0),\ \ldots,\ p_j(s_j - 1)\,]$ be a matrix of order $(s_j - 1) \times s_j$, such that the $s_j \times s_j$ matrix

$$Q \;=\; \begin{pmatrix} 1_{s_j}^T \\ P_j \end{pmatrix} \tag{A.1}$$

satisfies

$$Q_j^T Q_j = s_j I_{s_j} \tag{A.2}$$

The general mean will contribute a column $1_{(N)}$ to the design matrix arising from A. The main effect of F_j will cotribute an $N \times (s_j - 1)$ submatrix

$$Z_j \;=\; \begin{pmatrix} p_j(a_{1j})^T \\ \vdots \\ p_j(a_{Nj})^T \end{pmatrix} \tag{A.3}$$

to the design matrix arising from A.

Furthermore, if the 2fi $F_j F_k (1 \leq j < k \leq m)$ is included in the model then it would contribute an $N \times \{(s_j - 1)(s_k - 1)\}$ submatrix

$$Z_{jk} \;=\; \begin{pmatrix} p_j(a_{1j})^T \otimes p_k(a_{1k})^T \\ \vdots \\ p_j(a_{Nj})^T \otimes p_k(a_{Nk})^T \end{pmatrix} \tag{A.4}$$

to the design matrix arising from A. Then following Mukerjee (1999) (see also Xu and Wu (2001); Cheng and Ye (2004)),

$$X(h) = [1_{(N)}, Z_1, \ldots, Z_m, \ldots, Z_{jk}, \ldots], \tag{A.5}$$

where $1_{(N)}$ is the $N \times 1$ vector with all elements unity, and any Z_{jk} is included in $X(h)$ if and only if the 2fi F_jF_k belongs to h. In (A.5), $1_{(N)}$ corresponds to the general mean, any Z_j corresponds to the main effect of F_j and any Z_{jk} corresponds to the 2fi F_jF_k. With all factors at two levels, one can take $P_j = [-1 \;\; 1]$ in view of (A.2), and then (A.5) agrees with CDT. For general factorials, the specific choice of the matrices P_j, subject to (A.2), does not affect our results.

We propose to find an expression for

$$\sum_{h \in H(w)} tr\left[\{X(h)^T X(h)\}^2\right]$$

By (A.5), $X(h)X(h)^T = 1_N 1_N^T + \sum_{j=1}^m Z_j Z_j^T + \sum_h^* Z_{jk} Z_{jk}^T$ where $\sum_h^*$ denotes sum over ordered pairs over jk such that $F_jF_k \in h$. Hence

$$\begin{aligned}
\{X(h)^T X(h)\}^2 &= \left[X(h)X(h)^T\right]\left[X(h)X(h)^T\right] \\
&= \left[1_N 1_N^T + \sum_{j=1}^m Z_j Z_j^T + \sum_h^* Z_{jk} Z_{jk}^T\right]\left[1_N 1_N^T + \sum_{j=1}^m Z_j Z_j^T + \sum_h^* Z_{jk} Z_{jk}^T\right] \\
&= N 1_N 1_N^T + \sum_{u=1}^N 1_N 1_N^T Z_u Z_u^T + \sum_h^* 1_N 1_N^T Z_{uv} Z_{uv}^T \\
&\quad + \sum_{j=1}^m Z_j Z_j^T 1_N 1_N^T + \sum_{j=1}^m \sum_{u=1}^m Z_j Z_j^T Z_u Z_u^T + \sum_{j=1}^m \sum_h^* Z_j Z_j^T Z_{uv} Z_{uv}^T \\
&\quad + \sum_h^* Z_{jk} Z_{jk}^T 1_N 1_N^T + \sum_h^* \sum_{u=1}^m Z_{jk} Z_{jk}^T Z_u Z_u^T + \sum_h^* \sum_h^* Z_{jk} Z_{jk}^T Z_{uv} Z_{uv}^T.
\end{aligned} \tag{A.6}$$

Since A is an orthogonal array of strength two, from (A.1), (A.2),(A.3) and (A.4), the following are obvious :

$$1_{(N)}^T Z_u = 0 \; (1 \le u \le m), \tag{A.7}$$

$$1_{(N)}^T Z_{uv} = 0 \; (1 \le u < v \le m), \tag{A.8}$$

$$Z_j^T Z_j = \frac{N}{s_j}\sum_{\alpha=0}^{s_j-1} p_j(\alpha)p_j(\alpha)^T = NI_{s_j-1}(1 \le j \le m), \tag{A.9}$$

$$Z_j^T Z_u = \frac{N}{s_j s_u}\sum_{\alpha=0}^{s_j-1}\sum \beta = 0^{s_u-1} p_j(\alpha)p_u(\beta)^T = 0\ (1 \le j \ne u \le m). \tag{A.10}$$

Hence (A.6) yields

$$\begin{aligned}\left[X(h)X(h)^T\right]\left[X(h)X(h)^T\right] &= NJ_{NN} + \sum_{j=1}^{m} Z_j Z_j^T Z_j Z_j^T + \sum_{j=1}^{m}\sum_{h}^{*} Z_j Z_j^T Z_{uv} Z_{uv}^T \\ &+ \sum_{h}^{*}\sum_{u=1}^{m} Z_{jk} Z_{jk}^T Z_u Z_u^T + \sum_{h}^{*}\sum_{h}^{*} Z_{jk} Z_{jk}^T Z_{uv} Z_{uv}^T. \end{aligned}\tag{A.11}$$

Also note that

$$tr\left[\{X(h)^T X(h)\}^2\right] = tr\left[X(h)X(h)^T X(h)X(h)^T\right], \tag{A.12}$$

and that

$$\begin{aligned}\sum_{h\in H(w)} NJ_{NN} &= \begin{pmatrix} W \\ w \end{pmatrix} NJ_{NN}, \\ tr\left[\sum_{h\in H(w)} NJ_{NN}\right] &= \begin{pmatrix} W \\ w \end{pmatrix} N^2, \end{aligned}\tag{A.13}$$

$$\begin{aligned} tr\left[Z_j Z_j^T Z_j Z_j^T\right] &= tr\left[(Z_j^T Z_j)^2\right] = tr\left[N^2 I_{s_j-1}\right] = N^2(s_j-1), \\ tr\left[\sum_{h\in H(w)}\sum_{j=1}^{m} Z_j Z_j^T Z_j Z_j^T\right] &= \sum_{h\in H(w)}\sum_{j=1}^{m} tr\left[(Z_j^T Z_j)^2\right] \\ &= N^2 \begin{pmatrix} W \\ w \end{pmatrix} \sum_{j=1}^{m}(s_j-1). \end{aligned}\tag{A.14}$$

Also,

$$\begin{aligned} tr\left[\sum_{j=1}^{m}\sum_{h}^{*} Z_j Z_j^T Z_{uv} Z_{uv}^T\right] &= tr\left[\sum_{j=1}^{m}\sum_{h}^{*} Z_{uv} Z_{uv}^T Z_j Z_j^T\right] \\ &= tr\left[\sum_{u=1}^{m}\sum_{h}^{*} Z_{jk} Z_{jk}^T Z_u Z_u^T\right]. \end{aligned}\tag{A.15}$$

By $(A.11) - (A.15)$,

$$\begin{aligned}\sum_{h\in H(w)} tr\left[\{X(h)^T X(h)\}^2\right] &= c + 2\sum_{h\in H(w)}\sum_{j=1}^{m}\sum_{h}^{*} tr\left[Z_j Z_j^T Z_{uv} Z_{uv}^T\right] \\ &+ \sum_{h\in H(w)}\sum_{h}^{*}\sum_{h}^{*} tr\left[Z_{jk} Z_{jk}^T Z_{uv} Z_{uv}^T\right], \end{aligned}\tag{A.16}$$

where $c = N^2 \binom{W}{w} \left\{1 + \sum_{j=1}^{m}(s_j - 1)\right\}$ is same for all designs.

Now

$$\begin{aligned}
&\sum_{h\in H(w)} \overset{*}{\sum_{h}} \overset{*}{\sum_{h}} tr\left[Z_{jk}Z_{jk}^T Z_{uv}Z_{uv}^T\right] \\
&= \overset{*}{\sum_{h}} \overset{*}{\sum_{h}} \sum_{D:D\ni F_jF_k,F_uF_v} tr\left[Z_{jk}Z_{jk}^T Z_{uv}Z_{uv}^T\right] \qquad \left[\overset{*}{\sum_{h}} \text{ is sum over } \textit{all} \text{ the } W \text{ 2fi's}\right] \\
&= \overset{*}{\sum_{h}} \sum_{D:D\ni F_jF_k} tr\left[Z_{jk}Z_{jk}^T Z_{jk}Z_{jk}^T\right] + \overset{*}{\sum} \overset{*}{\sum_{jk\neq uv}} \sum_{D:D\ni F_jF_k,F_uF_v} tr\left[Z_{jk}Z_{jk}^T Z_{uv}Z_{uv}^T\right] \\
&= \binom{W-1}{w-1} \overset{*}{\sum} tr\left[Z_{jk}Z_{jk}^T Z_{jk}Z_{jk}^T\right] + \binom{W-2}{w-2} \overset{*}{\sum} \overset{*}{\sum_{jk\neq uv}} tr\left[Z_{jk}Z_{jk}^T Z_{uv}Z_{uv}^T\right] \qquad \text{(A.17)}
\end{aligned}$$

Since A is an orthogonal array of strenght two, from (A.1), (A.2) and (A.4), we get

$$\begin{aligned}
Z_{jk}^T Z_{jk} &= \frac{N}{s_j s_k} \sum_{\alpha=0}^{s_j-1} \sum_{\beta=0}^{s_k-1} \{p_j(\alpha)\otimes p_k(\beta)\}\{p_j(\alpha)\otimes p_k(\beta)\}^T \\
&= \frac{N}{s_j s_k} \left\{\sum_{\alpha=0}^{s_j-1} p_j(\alpha)p_j(\alpha)^T\right\} \otimes \left\{\sum_{\beta=0}^{s_k-1} p_k(\beta)p_k(\beta)^T\right\} \\
&= \frac{N}{s_j s_k} \left\{s_j I_{s_j-1}\right\} \otimes \{s_k I_{s_k-1}\} \\
&= N I_{(s_j-1)(s_k-1)}. \qquad \text{(A.18)}
\end{aligned}$$

Hence

$$tr\left[Z_{jk}Z_{jk}^T Z_{jk}Z_{jk}^T\right] = tr\left[N^2 I_{(s_j-1)(s_k-1)}\right] = N^2(s_j-1)(s_k-1).$$

Therefore

$$\begin{aligned}
&\binom{W-1}{w-1} \qquad \overset{*}{\sum} tr\left[Z_{jk}Z_{jk}^T Z_{jk}Z_{jk}^T\right] \\
&\qquad = \binom{W-1}{w-1} N^2 \overset{*}{\sum}(s_j-1)(s_k-1) = c_1, (\text{say}), \qquad \text{(A.19)}
\end{aligned}$$

where c_1 does not depend on the design.

Consider now the second term in (A.17). If $jk \neq uv$ then the set $\{j,k,u,v\}$ contains either three or four distinct elements. Let $\Delta(3)$ be the set of all triplets (a,b,c) such that

$1 \le a < b < c \le m$ and $\Delta(4)$ be the set of all (a,b,c,d) such that $1 \le a < b < c < d \le m$. Then

$$\begin{aligned}
& \sum^{*}\sum^{*}_{jk \neq uv} tr\left[Z_{jk}Z_{jk}^T Z_{uv}Z_{uv}^T\right] \\
= & \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\beta\gamma}Z_{\beta\gamma}^T + Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\alpha\gamma}Z_{\alpha\gamma}^T + Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\beta\gamma}Z_{\beta\gamma}^T \\
& \qquad + Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\alpha\beta}Z_{\alpha\beta}^T + Z_{\beta\gamma}Z_{\beta\gamma}^T Z_{\alpha\beta}Z_{\alpha\beta}^T + Z_{\beta\gamma}Z_{\beta\gamma}^T Z_{\alpha\gamma}Z_{\alpha\gamma}^T\Big] \\
& + \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\gamma\delta}Z_{\gamma\delta}^T + Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\beta\delta}Z_{\beta\delta}^T + Z_{\alpha\delta}Z_{\alpha\delta}^T Z_{\beta\gamma}Z_{\beta\gamma}^T \\
& \qquad + Z_{\beta\gamma}Z_{\beta\gamma}^T Z_{\alpha\delta}Z_{\alpha\delta}^T + Z_{\beta\delta}Z_{\beta\delta}^T Z_{\alpha\gamma}Z_{\alpha\gamma}^T + Z_{\gamma\delta}Z_{\gamma\delta}^T Z_{\alpha\beta}Z_{\alpha\beta}^T\Big] \\
= & \; 2\sum_{(\alpha,\beta,\gamma)\in\Delta(3)} tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\beta\gamma}Z_{\beta\gamma}^T + Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\alpha\gamma}Z_{\alpha\gamma}^T + Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\beta\gamma}Z_{\beta\gamma}^T\Big] \\
& + 2\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\gamma\delta}Z_{\gamma\delta}^T + Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\beta\delta}Z_{\beta\delta}^T + Z_{\alpha\delta}Z_{\alpha\delta}^T Z_{\beta\gamma}Z_{\beta\gamma}^T\Big]
\end{aligned} \tag{A.20}$$

Now for $(\alpha,\beta,\gamma) \in \Delta(3)$, by (A.4),

$$\begin{aligned}
Z_{\alpha\beta}^T Z_{\beta\gamma} &= \sum_{u=1}^{N}\Big[p_\alpha(a_{u\alpha}) \otimes p_\beta(a_{u\beta})\Big]\Big[p_\beta(a_{u\beta}) \otimes p_\gamma(a_{u\gamma})\Big]^T \\
&= \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)}\Big[p_\alpha(i) \otimes p_\beta(j)\Big]\Big[p_\beta(j) \otimes p_\gamma(k)\Big]^T,
\end{aligned}$$

where $n_{ijk}^{(\alpha,\beta,\gamma)}$ is the number of times the ordered triplet (i,j,k) occurs as a row in the $N \times 3$ subarray of A given by the αth, βth and γth columns. Therefore,

$$\begin{aligned}
& tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\beta\gamma}Z_{\beta\gamma}^T\Big] \\
= & \; tr\Big[(Z_{\alpha\beta}^T Z_{\beta\gamma})(Z_{\beta\gamma}^T Z_{\alpha\beta})\Big] \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} tr\Big([p_\alpha(i) \otimes p_\beta(j)][p_\beta(j) \otimes p_\gamma(k)]^T \\
& \qquad [p_\beta(u) \otimes p_\gamma(v)][p_\alpha(r) \otimes p_\beta(u)]^T\Big) \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)}(s_\beta\delta_{ju}-1)(s_\gamma\delta_{kv}-1)
\end{aligned}$$

$$
\begin{aligned}
& tr\Big([p_\alpha(i)\otimes p_\beta(j)][p_\gamma(r)\otimes p_\beta(u)]^T\Big) \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} (s_\beta\delta_{ju}-1)(s_\gamma\delta_{kv}-1) \\
& tr\Big([p_\gamma(r)\otimes p_\beta(u)]^T[p_\alpha(i)\otimes p_\beta(j)]\Big) \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} (s_\beta\delta_{ju}-1)^2(s_\gamma\delta_{kv}-1)(s_\alpha\delta_{ir}-1) \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} (s_\beta^2\delta_{ju}-2s_\beta\delta_{ju}+1) \\
& (s_\alpha s_\gamma\delta_{ir}\delta_{kv}-s_\alpha\delta_{ir}-s_\gamma\delta_{kv}+1) \\
= & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} (s_\alpha s_\beta^2 s_\gamma\delta_{ir}\delta_{ju}\delta_{kv}-2s_\alpha s_\beta s_\gamma\delta_{ir}\delta_{ju}\delta_{kv} \\
& +s_\alpha s_\gamma\delta_{ir}\delta_{kv}-s_\alpha s_\beta^2\delta_{ir}\delta_{ju}+2s_\alpha s_\beta\delta_{ir}\delta_{ju}-s_\alpha\delta_{ir}-s_\beta^2 s_\gamma\delta_{ju}\delta_{kv}+2s_\beta s_\gamma\delta_{ju}\delta_{kv} \\
& -s_\gamma\delta_{kv}+s_\beta^2\delta_{ju}-2s_\beta\delta_{ju}+1) \\
= & s_\alpha s_\beta^2 s_\gamma \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 - 2s_\alpha s_\beta s_\gamma \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 \\
& +s_\alpha s_\gamma\frac{N^2}{s_\alpha s_\gamma}+2s_\alpha s_\beta\frac{N^2}{s_\alpha s_\beta}-s_\alpha\frac{N^2}{s_\alpha}-s_\alpha s_\beta^2\frac{N^2}{s_\alpha s_\beta^2} \\
& -s_\beta^2 s_\gamma\frac{N^2}{s_\beta^2 s_\gamma}+2s_\beta s_\gamma\frac{N^2}{s_\beta s_\gamma}-s_\gamma\frac{N^2}{s_\gamma}+s_\beta^2\frac{N^2}{s_\beta^2}+2s_\beta\frac{N^2}{s_\beta}+N^2 \\
= & (s_\beta-2)\phi(\alpha\beta\gamma)+N^2-N^2s_\beta+2N^2-N^2-N^2s_\beta+2N^2-N^2+N^2s_\beta \\
& -2N^2+N^2 \\
= & (s_\beta-2)\phi(\alpha\beta\gamma)+2N^2-N^2s_\beta \\
= & (s_\beta-2)\Big\{\phi(\alpha\beta\gamma)-N^2\Big\},
\end{aligned}
$$

where

$$
\phi(\alpha\beta\gamma) = s_\alpha s_\beta s_\gamma \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2.
$$

Similarly

$$
\begin{aligned}
tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\alpha\gamma}Z_{\alpha\gamma}^T\Big] &= (s_\alpha-2)\Big\{\phi(\alpha\beta\gamma)-N^2\Big\}, \\
tr\Big[Z_{\alpha\gamma}Z_{\alpha\gamma}^T Z_{\beta\gamma}Z_{\beta\gamma}^T\Big] &= (s_\gamma-2)\Big\{\phi(\alpha\beta\gamma)-N^2\Big\}. \qquad \text{(A.21)}
\end{aligned}
$$

Hence, the first term of (A.20) equals

$$2 \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big(s_\alpha + s_\beta + s_\gamma - 6\Big)\Big\{\phi(\alpha\beta\gamma) - N^2\Big\}.$$

Consider now the second term in (A.20). Let $n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}$ be defined in the same way as $n_{ijk}^{(\alpha,\beta,\gamma)}$. For $(\alpha,\beta,\gamma,\delta) \in \Delta(4)$,

$$\begin{aligned} Z_{\alpha\beta}^T Z_{\gamma\delta} &= \sum_{u=1}^{N} \Big[p_\alpha(a_{u\alpha}) \otimes p_\beta(a_{u\beta})\Big]\Big[p_\gamma(a_{u\gamma}) \otimes p_\delta(a_{u\delta})\Big]^T \\ &= \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1} n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\Big[p_\alpha(i) \otimes p_\beta(j)\Big]\Big[p_\gamma(k)^T \otimes p_\delta(a_l)^T\Big]. \end{aligned}$$

This gives

$$\begin{aligned} & tr\Big[Z_{\alpha\beta}Z_{\alpha\beta}^T Z_{\gamma\delta}Z_{\gamma\delta}^T\Big] \\ = \; & tr\Big[(Z_{\alpha\beta}^T Z_{\gamma\delta})(Z_{\alpha\beta}^T Z_{\gamma\delta})^T\Big] \\ = \; & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1}\sum_{y=0}^{s_\delta-1} n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} n_{ruvy}^{(\alpha,\beta,\gamma,\delta)} \\ & \qquad\qquad (s_\alpha\delta_{ir}-1)(s_\beta\delta_{ju}-1)(s_\gamma\delta_{kv}-1)(s_\delta\delta_{ly}-1) \\ = \; & s_\alpha s_\beta s_\gamma s_\delta \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\Big\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\Big\}^2 \\ & \quad -\Big[s_\beta s_\gamma s_\delta \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\sum_{r=0}^{s_\alpha-1} n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} n_{rjkl}^{(\alpha,\beta,\gamma,\delta)} \\ & \qquad\qquad + \text{three other similar terms}\Big] \\ & \quad +\Big[s_\alpha s_\beta \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\sum_{v=0}^{s_\gamma-1}\sum_{y=0}^{s_\delta-1} n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} n_{ijvy}^{(\alpha,\beta,\gamma,\delta)} \\ & \qquad\qquad + \text{five other similar terms}\Big] \\ & \quad -\Big[s_\alpha \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1}\sum_{y=0}^{s_\delta-1} n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} n_{iuvy}^{(\alpha,\beta,\gamma,\delta)} \\ & \qquad\qquad + \text{three other similar terms}\Big] + N^2 \\ = \; & \phi(\alpha\beta\gamma\delta) - \Big[s_\beta s_\gamma s_\delta \sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\{n_{jkl}^{(\beta,\gamma,\delta)}\}^2 + \text{ three other similar terms}\Big] \\ & \quad +\Big[s_\alpha s_\beta \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\Big(\frac{N}{s_\alpha s_\beta}\Big)\Big(\frac{N}{s_\alpha s_\beta}\Big) + \text{ five other similar terms}\Big] \end{aligned}$$

$$
\begin{aligned}
& -\Big[s_\alpha \sum_{i=0}^{s_\alpha - 1} + \left(\frac{N}{s_\alpha}\right)\left(\frac{N}{s_\alpha}\right) + \text{ three other similar terms } \Big] + N^2 \\
= \;& \phi(\alpha\beta\gamma\delta) - \Big[\phi(\beta\gamma\delta) + \phi(\alpha\beta\gamma) + \phi(\alpha\gamma\delta) + \phi(\alpha\beta\delta)\Big] + 6N^2 - 4N^2 + N^2 \\
= \;& \phi(\alpha\beta\gamma\delta) - \Big[\phi(\beta\gamma\delta) + \phi(\alpha\beta\gamma) + \phi(\alpha\gamma\delta) + \phi(\alpha\beta\delta)\Big] + 3N^2
\end{aligned}
$$

Hence the second term in (A.20) equals

$$
\begin{aligned}
& 6 \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \Big[\phi(\alpha\beta\gamma\delta) - \phi(\beta\gamma\delta) - \phi(\alpha\beta\gamma) - \phi(\alpha\gamma\delta) - \phi(\alpha\beta\delta) + 3N^2\Big] \\
= \;& 6 \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \phi(\alpha\beta\gamma\delta) + 18N^2 \binom{m}{4} - 6(m-3) \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \phi(\alpha\beta\gamma).
\end{aligned}
$$

Summarizing the above,

$$
\begin{aligned}
& \sum^{*}\sum^{*}_{jk\neq uv} tr\Big[Z_{jk} Z_{jk}^T Z_{uv} Z_{uv}^T\Big] \\
= \;& 2 \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} (s_\alpha + s_\beta + s_\gamma - 6)\phi(\alpha\beta\gamma) - 6(m-3) \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \phi(\alpha\beta\gamma) \\
& +6 \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \phi(\alpha\beta\gamma\delta) + c^*, \qquad \text{(A.22)}
\end{aligned}
$$

where c^* is a constant that does not depend on the origin.

We next consider the second term in the RHS of (A.16).

$$
\begin{aligned}
& \sum_{h\in H(w)} \sum_{j=1}^{m} \sum_{h}^{*} tr\left[Z_j Z_j^T Z_{uv} Z_{uv}^T\right] \\
& = \sum^{*} \sum_{j=1}^{m} \sum_{D: D\ni F_u F_v} tr\left[Z_j Z_j^T Z_{uv} Z_{uv}^T\right] \\
& = \binom{W-1}{w-1} \sum^{*} \sum_{j=1}^{m} tr\left[Z_j Z_j^T Z_{uv} Z_{uv}^T\right] \\
& = \binom{W-1}{w-1} \Bigg[\sum \sum_{1\le\alpha<\beta\le m} \Big\{ tr\left(Z_\alpha Z_\alpha^T Z_{\alpha\beta} Z_{\alpha\beta}^T\right) + tr\left(Z_\beta Z_\beta^T Z_{\alpha\beta} Z_{\alpha\beta}^T\right) \Big\} \\
& + \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big\{ tr\left(Z_\alpha Z_\alpha^T Z_{\beta\gamma} Z_{\beta\gamma}^T\right) + tr\left(Z_\beta Z_\beta^T Z_{\alpha\gamma} Z_{\alpha\gamma}^T\right) \\
& \qquad + tr\left(Z_\gamma Z_\gamma^T Z_{\alpha\beta} Z_{\alpha\beta}^T\right) + \Big\} \Bigg]. \qquad \text{(A.23)}
\end{aligned}
$$

Now for $1 \leq \alpha < \beta \leq m$, by (A.3) and (A.4),

$$Z_\alpha^T Z_{\alpha\beta} = \sum_{u=1}^{N} p_\alpha(a_{u\alpha})\Big\{p_\alpha(a_{u\alpha})^T \otimes p_\beta(a_{u\beta})^T\Big\} = 0.$$

Similarly, $Z_\beta^T Z_{\alpha\beta} = 0$. Also, for $(\alpha, \beta, \gamma) \in \Delta(3)$, we have

$$\begin{aligned} Z_\alpha^T Z_{\beta\gamma} &= \sum_{u=1}^{N} p_\alpha(a_{u\alpha})\Big\{p_\beta(a_{u\beta})^T \otimes p_\gamma(a_{u\gamma})^T\Big\} \\ &= \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} p_\alpha(i)\Big\{p_\beta(j) \otimes p_\gamma(k)\Big\}^T. \end{aligned}$$

This implies

$$\begin{aligned} & tr\Big[Z_\alpha Z_\alpha^T Z_{\beta\gamma} Z_{\beta\gamma}^T\Big] \\ = & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} tr\Big[p_\alpha(i)\Big\{p_\beta(j) \otimes p_\gamma(k)\Big\}^T \\ & \qquad\qquad \Big\{p_\beta(u) \otimes p_\gamma(v)\Big\} p_\alpha(r)^T\Big] \\ = & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} (s_\alpha\delta_{ir} - 1)(s_\beta\delta_{ju} - 1)(s_\gamma\delta_{kv} - 1) \\ = & \sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{r=0}^{s_\alpha-1}\sum_{u=0}^{s_\beta-1}\sum_{v=0}^{s_\gamma-1} n_{ijk}^{(\alpha,\beta,\gamma)} n_{ruv}^{(\alpha,\beta,\gamma)} \Big[s_\alpha s_\beta s_\gamma \delta_{ir}\delta_{ju}\delta_{kv} \\ & \qquad -\Big\{s_\alpha s_\beta \delta_{ir}\delta_{ju} + \text{ two similar terms}\Big\} - \Big\{s_\alpha\delta_{ir} + \text{ two similar terms}\Big\} - 1\Big] \\ = & \phi(\alpha\beta\gamma) + \text{terms do not depend on the design.} \end{aligned}$$

Similar conditions apply to $tr\Big[Z_\beta Z_\beta^T Z_{\alpha\gamma} Z_{\alpha\gamma}^T\Big]$ etc. Hence the RHS of (A.23) equals

$$3\binom{W-1}{w-1} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \phi(\alpha\beta\gamma) + \text{ terms that do not depend on the design .} \qquad \text{(A.24)}$$

By (A.16), (A.17), (A.22), (A.23) and (A.24)

$$\sum_{h\in H(w)} tr\Big\{\Big(X(h)^T X(h)\Big)^2\Big\} = C_0^* + 6\binom{W-1}{w-1} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \phi(\alpha\beta\gamma)$$

$$
\begin{aligned}
&+\binom{W-2}{w-2}\Big[2\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha,\beta,\gamma)(s_\alpha+s_\beta+s_\gamma-6)\\
&\qquad -6(m-3)\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha\beta\gamma)+6\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\phi(\alpha,\beta,\gamma,\delta)\Big]\\
=\;& C_0^*+\binom{W-1}{w-1}\Big[6\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha\beta\gamma)+\\
&\qquad \frac{w-1}{W-1}\Big\{2\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha,\beta,\gamma)(s_\alpha+s_\beta+s_\gamma-6)\\
&\qquad -6(m-3)\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha\beta\gamma)+6\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\phi(\alpha\beta\gamma\delta)\Big\}\Big]\\
=\;& C_0^*+\binom{W-1}{w-1}\Big[\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\xi(\alpha,\beta,\gamma)\phi(\alpha\beta\gamma)+\frac{6(w-1)}{W-1}\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\phi(\alpha\beta\gamma\delta)\Big],
\end{aligned}
$$

where C_0^* does not depend on the design and

$$
\begin{aligned}
\xi(\alpha,\beta,\gamma) &= 6+\frac{w-1}{W-1}\Big\{2(s_\alpha+s_\beta+s_\gamma-6)-6(m-3)\Big\}\\
&= 6+\frac{2(w-1)}{W-1}\Big\{s_\alpha+s_\beta+s_\gamma-3m+3)\Big\}
\end{aligned}
$$

Hence the objective function can be written as

$$
E_w^* = \sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\xi(\alpha,\beta,\gamma)\phi(\alpha\beta\gamma)+\frac{6(w-1)}{W-1}\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\phi(\alpha\beta\gamma\delta), \qquad \text{(A.25)}
$$

where

$$
\xi(\alpha,\beta,\gamma) = 6+\frac{2(w-1)}{W-1}\Big\{s_\alpha+s_\beta+s_\gamma-3m+3)\Big\}. \qquad \text{(A.26)}
$$

Special Case : $s_1 = s_2 = \ldots = s_m = s$

$$
\xi(\alpha,\beta,\gamma) = 6-\frac{6(w-1)}{W-1}(m-s-1) = \xi_0 \text{ (say)}
$$

Also,

$$
\phi(\alpha\beta\gamma) = s^3\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2
$$

$$\phi(\alpha\beta\gamma\delta) \;=\; s^4 \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\left\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\right\}^2$$

Hence, then

$$\begin{aligned} E_w^* &= \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \xi_0 s^3 \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\left\{n_{ijk}^{(\alpha,\beta,\gamma)}\right\}^2 + \\ &\quad \frac{6(w-1)}{W-1}\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} s^4 \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\left\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\right\}^2 \\ &= \left[6 - \frac{6(w-1)}{W-1}(m-s-1)\right]s^3 T_3 + \frac{6(w-1)}{W-1}s^4 T_4 \end{aligned} \tag{A.27}$$

where

$$\begin{aligned} T_3 &= \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\left\{n_{ijk}^{(\alpha,\beta,\gamma)}\right\}^2 \\ T_4 &= \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\left\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\right\}^2 \end{aligned} \tag{A.28}$$

In particular, consider $\mathbf{s_1} = \mathbf{s_2} = \ldots = \mathbf{s_m} = \mathbf{s} = \mathbf{2}$.

Then

$$\begin{aligned} E_w^* &= \left[6 - \frac{6(w-1)}{W-1}(m-s-1)\right]8T_3 + \frac{6(w-1)}{W-1}16T_4 \\ &= 48\left[1 - \frac{(w-1)}{W-1}(m-s-1)\right]T_3 + \frac{96(w-1)}{W-1}T_4, \end{aligned} \tag{A.29}$$

where T_3 and T_4 are given by (A.28).

Do (A.28) and (A.29) agree with Cheng, Deng and Tang (2002)?

In their notation, $B_s(d) = N^{-2}\sum_{S:|S|=s}\left[\sum_{i=1}^N \prod_{j\in S} x_{ij}(d)\right]^2$ where $x_{ij}(d)$ is the (i,j)th entry of $X(d)$. Since we have an OA of strength two, in our case we have $\sum_{i=1}^N \prod_{j\in S} x_{ij}(d) = 0$ whenever $|S| = 2$. Hence $B_2(d) = 0$ in our situation.

We now note that

$$n_{001}^{(\alpha,\beta,\gamma)} \;=\; (n_{001}^{(\alpha,\beta,\gamma)} + n_{000}^{(\alpha,\beta,\gamma)}) - n_{000}^{(\alpha,\beta,\gamma)} = n_{00}^{(\alpha,\beta)} - n_{000}^{(\alpha,\beta,\gamma)} = \frac{N}{4} - n_{000}^{(\alpha,\beta,\gamma)},$$

$$
\begin{aligned}
n_{110}^{(\alpha,\beta,\gamma)} &= (n_{110}^{(\alpha,\beta,\gamma)} + n_{010}^{(\alpha,\beta,\gamma)}) - n_{010}^{(\alpha,\beta,\gamma)} = n_{10}^{(\alpha,\beta)} - n_{010}^{(\alpha,\beta,\gamma)} = \frac{N}{4} - \left(\frac{N}{4} - n_{000}^{(\alpha,\beta,\gamma)}\right) \\
&\qquad\qquad = n_{000}^{(\alpha,\beta,\gamma)}, \\
n_{111}^{(\alpha,\beta,\gamma)} &= (n_{111}^{(\alpha,\beta,\gamma)} + n_{110}^{(\alpha,\beta,\gamma)}) - n_{110}^{(\alpha,\beta,\gamma)} = n_{11}^{(\alpha,\beta)} - n_{110}^{(\alpha,\beta,\gamma)} = \frac{N}{4} - n_{000}^{(\alpha,\beta,\gamma)}.
\end{aligned}
$$

Next,

$$
\begin{aligned}
B_3(d) &= N^{-2} \sum_{S:|S|=3} \Big[\sum_{i=1}^{N} \prod_{j\in S} x_{ij}(d)\Big]^2 \\
&= N^{-2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big[\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma}\Big]^2 \\
&= N^{-2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big[n_{000}^{(\alpha,\beta,\gamma)} - n_{001}^{(\alpha,\beta,\gamma)} - n_{010}^{(\alpha,\beta,\gamma)} - n_{100}^{(\alpha,\beta,\gamma)} + n_{110}^{(\alpha,\beta,\gamma)} \\
&\qquad\qquad + n_{101}^{(\alpha,\beta,\gamma)} + n_{011}^{(\alpha,\beta,\gamma)} - n_{111}^{(\alpha,\beta,\gamma)}\Big]^2 \\
&= N^{-2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big[n_{000}^{(\alpha,\beta,\gamma)} - 3\left(\frac{N}{4} - n_{000}^{(\alpha,\beta,\gamma)}\right) + 3n_{000}^{(\alpha,\beta,\gamma)} - \left(\frac{N}{4} - n_{000}^{(\alpha,\beta,\gamma)}\right)\Big]^2 \\
&= N^{-2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big[8n_{000}^{(\alpha,\beta,\gamma)} - N\Big]^2 \\
&= 64N^{-2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big[n_{000}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big]^2.
\end{aligned}
$$

Again,

$$
\begin{aligned}
&\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \Big\{n_{ijk}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big\}^2 \\
= \;& \Big(n_{000}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 + \Big\{\Big(n_{001}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 + \text{two similar terms}\Big\} \\
&\quad + \Big\{\Big(n_{110}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 + \text{two similar terms}\Big\} + \Big(n_{111}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 \\
= \;& \Big(n_{000}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 + 3\Big(\frac{N}{8} - n_{000}^{(\alpha,\beta,\gamma)}\Big)^2 + 3\Big(n_{000}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2 + \Big(\frac{N}{8} - n_{000}^{(\alpha,\beta,\gamma)}\Big)^2 \\
= \;& 8\Big(n_{000}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big)^2.
\end{aligned}
$$

Therefore

$$
B_3(d) = \frac{8}{N^2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \Big\{n_{ijk}^{(\alpha,\beta,\gamma)} - \frac{N}{8}\Big\}^2
$$

$$
\begin{aligned}
&= \frac{8}{N^2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \left\{ n_{ijk}^{(\alpha,\beta,\gamma)2} - \frac{N}{8} n_{ijk}^{(\alpha,\beta,\gamma)} + \frac{N^2}{64} \right\} \\
&= \frac{8}{N^2} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \left\{ \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} n_{ijk}^{(\alpha,\beta,\gamma)2} - \frac{N^2}{8} \right\} \\
&= \frac{8}{N^2} T_3 - \binom{m}{3}. \qquad \text{(A.30)}
\end{aligned}
$$

Also

$$
\begin{aligned}
B_4(d) &= N^{-2} \sum_{S:|S|=4} \Big[\sum_{i=1}^{N} \prod_{j\in S} x_{ij}(d) \Big]^2 \\
&= N^{-2} \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \Big[\sum_{i=1}^{N} x_{i\alpha} x_{i\beta} x_{i\gamma} x_{i\delta} \Big]^2 .
\end{aligned}
$$

For any $(\alpha,\beta,\gamma,\delta) \in \Delta(4)$, $\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma}x_{i\delta} = \sum_i \sum_j \sum_k \sum_l n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} (-1)^{i+j+k+l} = n^T\xi$ where n is a 16×1 vector with elements $n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}$ arranged lexicographically, and $\xi^T = (1-1)\otimes(1-1)\otimes(1-1)\otimes(1-1)$. Hence $(\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma}x_{i\delta})^2 = (n^T\xi)^2 = n^T\xi\xi^T n$. But

$$
(1-1)^T \otimes (1-1) = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = 2I - J
$$

gives

$$
\begin{aligned}
\xi\xi^T &= (2I-J)\otimes(2I-J)\otimes(2I-J)\otimes(2I-J) \\
&= \Big[16 I\otimes I\otimes I\otimes I - 8\Big\{ I\otimes I\otimes I\otimes J + \text{ three similar terms } \Big\} \\
&\quad +4\Big\{ I\otimes I\otimes J\otimes J + \text{ five similar terms } \Big\} - 2\Big\{ I\otimes J\otimes J\otimes J \\
&\quad + \text{three similar terms } \Big\} + J\otimes J\otimes J\otimes J \Big],
\end{aligned}
$$

and recall that the elements of $(I\otimes I\otimes I\otimes 1^T)n$ are $n_{ijk}^{(\alpha,\beta,\gamma)}$. Then

$$
\begin{aligned}
n^T(I\otimes I\otimes I\otimes I)n &= \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1} \Big\{ n_{ijkl}^{(\alpha,\beta,\gamma,\delta)} \Big\}^2, \\
n^T(I\otimes I\otimes I\otimes J)n &= n^T(I\otimes I\otimes I\otimes 1)(I\otimes I\otimes I\otimes 1^T)n \\
&= \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1} \Big\{ n_{ijk}^{(\alpha,\beta,\gamma)} \Big\}^2,
\end{aligned}
$$

$$
\begin{aligned}
n^T(I\otimes I\otimes J\otimes J)n &= n^T(I\otimes I\otimes 1\otimes 1)(I\otimes I\otimes 1^T\otimes 1^T)n \\
&= \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\Big\{n_{ij}^{(\alpha,\beta)}\Big\}^2 \\
&= \sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\Big(\frac{N}{4}\Big)^2 \\
&= \frac{N^2}{4}, \\
n^T(I\otimes J\otimes J\otimes J)n &= \sum_{i=0}^{s-1}\Big\{n_{i}^{(\alpha)}\Big\}^2 \\
&= \sum_{i=0}^{s-1}\Big(\frac{N}{2}\Big)^2 \\
&= \frac{N^2}{2}, \\
n^T(J\otimes J\otimes J\otimes J)n &= N^2.
\end{aligned}
$$

Thus

$$
\begin{aligned}
\Big(\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma}x_{i\delta}\Big)^2 &= 16\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\Big\}^2 \\
&\quad -8\Big[\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2+\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ijl}^{(\alpha,\beta,\delta)}\Big\}^2 \\
&\quad +\sum_{i=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ikl}^{(\alpha,\gamma,\delta)}\Big\}^2+\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{jkl}^{(\beta,\gamma,\delta)}\Big\}^2\Big]+4.6.\frac{N^2}{4} \\
&\quad -2.4.\frac{N^2}{2}+N^2 \\
&= 16\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\Big\}^2-8\Big[\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 \\
&\quad +\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ijl}^{(\alpha,\beta,\delta)}\Big\}^2+\sum_{i=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ikl}^{(\alpha,\gamma,\delta)}\Big\}^2 \\
&\quad +\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{jkl}^{(\beta,\gamma,\delta)}\Big\}^2\Big]+3N^2.
\end{aligned}
$$

Hence

$$
\begin{aligned}
B_4(d) &= N^{-2}\Big[16T_4-8\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\Big(\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2+\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ijl}^{(\alpha,\beta,\delta)}\Big\}^2 \\
&\quad +\sum_{i=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{ikl}^{(\alpha,\gamma,\delta)}\Big\}^2+\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\sum_{l=0}^{s-1}\Big\{n_{jkl}^{(\beta,\gamma,\delta)}\Big\}^2\Big)+3N^2\binom{m}{4}\Big]
\end{aligned}
$$

$$= \quad N^{-2}\Big[16T_4 - 8(m-3)T_3\Big] + 3\binom{m}{4}. \tag{A.31}$$

Alternative derivation of (A.30) using method for obtaining (A.31)

$\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma} = n^T\xi$ where n is a 8×1 vector with elements $n_{ijk}^{(\alpha,\beta,\gamma)}$ arranged lexicographically, and $\xi^T = (1-1)\otimes(1-1)\otimes(1-1)$.

$$\begin{aligned}
\Big[\sum_{i=1}^{N} x_{i\alpha}x_{i\beta}x_{i\gamma}\Big]^2 &= (n^T\xi)^2 \\
&= n^T\xi\xi^T n \\
&= n^T\Big[(2I-J)\otimes(2I-J)\otimes(2I-J)\Big]n \\
&= n^T\Big[8I\otimes I\otimes I - \Big\{4I\otimes I\otimes J + \ldots\Big\} + \Big\{2I\otimes J\otimes J \\
&\qquad + \ldots\Big\} - J\otimes J\otimes J\Big]n \\
&= 8\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 - 4.3.4.\frac{N^2}{16} + 2.3.2.\frac{N^2}{4} - N^2 \\
&= 8\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 - N^2,
\end{aligned}$$

which gives

$$\begin{aligned}
B_3(d) &= N^{-2}\sum_{S:|S|=3}\Big[\sum_{i=1}^{N}\prod_{j\in S}x_{ij}(d)\Big]^2 \\
&= N^{-2}\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\Big[\sum_{i=1}^{N}x_{i\alpha}x_{i\beta}x_{i\gamma}\Big]^2 \\
&= N^{-2}\Big[8\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\sum_{i=0}^{s-1}\sum_{j=0}^{s-1}\sum_{k=0}^{s-1}n_{ijk}^{(\alpha,\beta,\gamma)2} - N^2\binom{m}{3}\Big] \\
&= \frac{8}{N^2}T_3 - \binom{m}{3},
\end{aligned}$$

which agrees with (A.30).

Following Proposition 1 of Cheng et al. (2002), their objective function is

$$\begin{aligned} & 6\frac{w}{W}B_3(d) + 6\frac{w}{W}\frac{(w-1)}{(W-1)}B_4(d) \\ = \; & 6\frac{w}{W}\Big[\frac{8}{N^2}T_3 - \binom{m}{3}\Big] + 6\frac{w}{W}\frac{(w-1)}{(W-1)}\Big[\frac{16}{N^2}T_4 - \frac{8(m-3)}{N^2}T_3 + 3\binom{m}{4}\Big] \\ = \; & \text{Constant free from design} + \frac{w}{W}\frac{6}{N^2}\Big[8T_3 + \frac{w-1}{W-1}\Big\{16T_4 - 8(m-3)T_3\Big\}\Big]. \end{aligned}$$

Effectively, this boils down to $8T_3 + \frac{w-1}{W-1}\Big\{16T_4 - 8(m-3)T_3\Big\} = 8\Big[1 - \frac{w-1}{W-1}(m-3)\Big]T_3 + 16\frac{w-1}{W-1}T_4$, which is proportional to (A.29).

An expression for B_3 and B_4

Consider a typical member of Ω_3, say $11100\ldots0$. Then

$$\begin{aligned} V(11100\ldots0) \; = \; & \Big(I_{s_1} - \frac{1}{s_1}J_{s_1}\Big) \otimes \Big(I_{s_2} - \frac{1}{s_2}J_{s_2}\Big) \otimes \Big(I_{s_3} - \frac{1}{s_3}J_{s_3}\Big) \otimes \Big(\frac{1}{s_4}J_{s_4}\Big) \otimes \ldots \\ & \otimes \Big(\frac{1}{s_m}J_{s_m}\Big). \end{aligned}$$

Then

$$\begin{aligned} n^T V(11100\ldots0)n \; = \; & \frac{1}{v}n^T\Big[\Big(s_1I_{s_1} - J_{s_1}\Big) \otimes \Big(s_2I_{s_2} - J_{s_2}\Big) \otimes \Big(s_3I_{s_3} - J_{s_3}\Big) \otimes J_{s_4} \otimes \ldots \\ & \otimes J_{s_m}\Big]n \\ = \; & \frac{1}{v}n^T\Big[\Big\{I_{s_1} \otimes I_{s_2} \otimes I_{s_3} \otimes 1_{s_4} \otimes \ldots \otimes 1_{s_m}\Big\} \\ & \Big\{(s_1I_{s_1} - J_{s_1}) \otimes (s_2I_{s_2} - J_{s_2}) \otimes (s_3I_{s_3} - J_{s_3})\Big\} \\ & \Big\{I_{s_1} \otimes I_{s_2} \otimes I_{s_3} \otimes 1_{s_4}^T \otimes \ldots \otimes 1_{s_m}^T\Big\}\Big]n \\ = \; & \frac{1}{v}n^{(123)T}\Big[\Big(s_1I_{s_1} - J_{s_1}\Big) \otimes \Big(s_2I_{s_2} - J_{s_2}\Big) \otimes \Big(s_3I_{s_3} - J_{s_3}\Big)\Big]n^{(123)} \\ = \; & \frac{1}{v}n^{(123)T}\Big[s_1s_2s_3I_{s_1} \otimes I_{s_2} \otimes I_{s_3} - \Big\{s_1s_2I_{s_1} \otimes I_{s_2} \otimes J_{s_3} + \ldots\Big\} \\ & + \Big\{s_1I_{s_1} \otimes J_{s_2} \otimes J_{s_3} + \ldots\Big\} - J_{s_1} \otimes J_{s_2} \otimes J_{s_3}\Big]n^{(123)}. \end{aligned}$$

Note, $n^{(123)} = \Big[\Big\{I_{s_1} \otimes I_{s_2} \otimes I_{s_3} \otimes 1_{s_4}^T \otimes \ldots \otimes 1_{s_m}^T\Big\}\Big]n$ is a vector whose elements are $n_{ijk}^{(1,2,3)}$ arranged lexicographically. Now,

$$n^{(123)T}\left(I_{s_1}\otimes I_{s_2}\otimes I_{s_3}\right)n^{(123)}=\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1}\sum_{k=0}^{s_3-1}\left\{n_{ijk}^{(1,2,3)}\right\}^2.$$

Similarly,

$$\begin{aligned}n^{(123)T}\left(I_{s_1}\otimes I_{s_2}\otimes J_{s_3}\right)n^{(123)} &= n^{(123)T}\left(I_{s_1}\otimes I_{s_2}\otimes 1_{s_3}\right)\left(I_{s_1}\otimes I_{s_2}\otimes 1_{s_3}^T\right)n^{(123)}\\ &= n^{(12)T}n^{(12)}\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1}\left\{n_{ij}^{(1,2)}\right\}^2,\end{aligned}$$

which does not depend on the design as $n_{ij}^{(1,2)}=\frac{N}{s_1s_2}\forall i,j$. Similar considerations apply to other terms. Hence

$$n^TV(1110\ldots0)n = \frac{1}{v}\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1}\sum_{k=0}^{s_3-1}\left\{n_{ijk}^{(1,2,3)}\right\}^2 s_1s_2s_3+\text{constant}.$$

Hence

$$\begin{aligned}B_3 &= \sum_{x\in\Omega_3}n^TV(x)n\\ &= \frac{1}{v}\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\left\{n_{ijk}^{(\alpha,\beta,\gamma)}\right\}^2 s_\alpha s_\beta s_\gamma\\ &= \frac{1}{v}\sum_\alpha\sum_\beta\sum_\gamma\phi(\alpha\beta\gamma).\end{aligned}\tag{A.32}$$

Similarly,

$$\begin{aligned}&n^TV(111100\ldots0)n\\ =\;&\frac{1}{v}n^T\Big[\left(s_1I_{s_1}-J_{s_1}\right)\otimes\left(s_2I_{s_2}-J_{s_2}\right)\otimes\left(s_3I_{s_3}-J_{s_3}\right)\otimes\left(s_4I_{s_4}-J_{s_4}\right)\otimes J_{s_5}\otimes\\ &\qquad\qquad\ldots\otimes J_{s_m}\Big]n\\ =\;&\frac{1}{v}n^{(1234)T}\Big[\left(s_1I_{s_1}-J_{s_1}\right)\otimes\left(s_2I_{s_2}-J_{s_2}\right)\otimes\left(s_3I_{s_3}-J_{s_3}\right)\otimes\left(s_4I_{s_4}-J_{s_4}\right)\Big]n^{(1234)}\\ =\;&\frac{1}{v}n^{(123)T}\Big[s_1s_2s_3s_4I_{s_1}\otimes I_{s_2}\otimes I_{s_3}\otimes I_{s_4}-\left\{s_1s_2s_3I_{s_1}\otimes I_{s_2}\otimes I_{s_3}\otimes J_{s_4}+\ldots\right\}\\ &\quad+\left\{s_1s_2I_{s_1}\otimes I_{s_2}\otimes J_{s_3}\otimes J_{s_4}+\ldots\right\}-\left\{s_1I_{s_1}\otimes J_{s_2}\otimes J_{s_3}\otimes J_{s_4}+\ldots\right\}\\ &\quad-J_{s_1}\otimes J_{s_2}\otimes J_{s_3}\otimes J_{s_4}\Big]n^{(1234)}\\ =\;&\frac{1}{v}\Big[s_1s_2s_3s_4\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1}\sum_{k=0}^{s_3-1}\sum_{l=0}^{s_4-1}\left\{n_{ijkl}^{(1,2,3,4)}\right\}^2-\Big(s_1s_2s_3\sum_{i=0}^{s_1-1}\sum_{j=0}^{s_2-1}\sum_{k=0}^{s_3-1}\left\{n_{ijk}^{(1,2,3)}\right\}^2\\ &\qquad\qquad+\ldots\Big)\Big]+\text{constant}\\ =\;&\frac{1}{v}\Big[\phi(1,2,3,4)-\left\{\phi(1,2,3)+\phi(1,2,4)+\phi(1,3,4)+\phi(2,3,4)\right\}\Big]+\text{constant}.\end{aligned}$$

Hence

$$\begin{aligned} B_4 &= \sum_{x\in\Omega_4} n^T V(x) n \\ &= \frac{1}{v}\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\Big[\phi(1,2,3,4) - \Big\{\phi(1,2,3)+\phi(1,2,4)+\phi(1,3,4)+\phi(2,3,4)\Big\}\Big] \\ &\qquad\qquad +\text{constant} \\ &= \frac{1}{v}\Big[\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\phi(\alpha,\beta,\gamma,\delta) - (m-3)\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\phi(\alpha\beta\gamma)\Big] + \text{constant}. \end{aligned}$$

E_w^* for symmetric factorials

In particular, if $s_1 = s_2 = \ldots = s_m = s$, then

$$B_3 = \frac{s^3}{v}\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2 = \frac{s^3}{v}T_3,$$

and

$$\begin{aligned} B_4 &= \frac{s^3}{v}\Big[S^4\sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)}\sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\sum_{l=0}^{s_\delta-1}\Big\{n_{ijkl}^{(\alpha,\beta,\gamma,\delta)}\Big\}^2 - \\ &\qquad (m-3)s^3\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}\sum_{i=0}^{s_\alpha-1}\sum_{j=0}^{s_\beta-1}\sum_{k=0}^{s_\gamma-1}\Big\{n_{ijk}^{(\alpha,\beta,\gamma)}\Big\}^2\Big] \\ &= \frac{1}{v}\Big[s^4T_4 - (m-3)s^3T_3\Big]. \end{aligned}$$

Hence, solving we get,

$$\begin{aligned} T_3 &= \frac{v}{s^3}B_3, \\ T_4 &= \frac{v}{s^4}\Big[B_4 + (m-3)B_3\Big]. \end{aligned}$$

Hence for symmetrical factorials, the objective function in (A.27) becomes

$$\begin{aligned} E_w^* &= \Big[6 - \frac{6(w-1)}{W-1}(m-s-1)\Big]s^3\frac{v}{s^3}B_3(d) + \frac{6(w-1)}{W-1}s^4\frac{v}{s^4}\Big[B_4(d) + (m-3)B_3(d)\Big] \\ &= 6v\Big[\Big\{1 + \frac{w-1}{W-1}(s-2)\Big\}B_3(d) + \frac{w-1}{W-1}B_4(d)\Big]. \end{aligned} \tag{A.33}$$

In particular, if $s = 2$, then the above is proportional to $B_3(d) + \frac{w-1}{W-1}B_4(d)$, which agrees with Cheng et al. (2002).

A.2 Proof of Theorem 1

For any $(\alpha,\beta,\gamma) \in \Delta(3)$ define $B(\alpha\beta\gamma) = n^T V(x) n$ where $x = x_1 \dots x_m$ has 1 in αth, βth and γth positions and 0 elsewhere. Also for any $(\alpha,\beta,\gamma,\delta) \in \Delta(4)$ define $B(\alpha\beta\gamma\delta) = n^T V(x) n$ where $x = x_1 \dots x_m$ has 1 in αth, βth, γth and δth positions and 0 elsewhere. $B(\alpha\beta\gamma)$ measures the departure of the subarray of A given by the αth, βth and γth columns from being an OA of strength three. Similarly $B(\alpha\beta\gamma\delta)$ can be interpreted. We have already seen that

$$\begin{aligned} B(\alpha\beta\gamma) &= \frac{1}{v}\phi(\alpha,\beta,\gamma), \\ B(\alpha\beta\gamma\delta) &= \frac{1}{v}\Big[\phi(\alpha\beta\gamma\delta) - \Big\{\phi(\alpha\beta\gamma) + \phi(\alpha\beta\delta)\phi(\alpha\gamma\delta)\phi(\beta\gamma\delta)\Big\}\Big]. \end{aligned}$$

Hence,

$$\begin{aligned} \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \phi(\alpha\beta\gamma\delta) &= \sum_{(\alpha,\beta,\gamma,\delta)\in\Delta(4)} \Big[vB(\alpha\beta\gamma\delta) + \phi(\alpha\beta\gamma) + \phi(\alpha\beta\delta)\phi(\alpha\gamma\delta)\phi(\beta\gamma\delta)\Big] \\ &= vB_4(d) + (m-3) \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \phi(\alpha,\beta,\gamma) \\ &= vB_4(d) + (m-3)vB_3(d). \end{aligned} \tag{A.34}$$

Therefore, by (A.25) and (A.26)

$$\begin{aligned} \Delta(w) &= \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} v\xi(\alpha,\beta,\gamma)B(\alpha\beta\gamma) + \frac{6(w-1)}{W-1}\Big[vB_4(d) + (m-3)vB_3(d)\Big] \\ &= v\Big[\sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big\{6 + \frac{2(w-1)}{W-1}(s_\alpha + s_\beta + s_\gamma - 3m + 3)\Big\}B(\alpha\beta\gamma) \\ &\qquad + \frac{6(w-1)}{W-1}\Big\{vB_4(d) + (m-3)vB_3(d)\Big\}\Big] \\ &= v\Big[\sum_{(\alpha,\beta,\gamma)\in\Delta(3)} \Big\{6 + \frac{2(w-1)}{W-1}(3-3m)\Big\}B(\alpha\beta\gamma) \\ &\qquad + \frac{2(w-1)}{W-1} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} (s_\alpha + s_\beta + s_\gamma)B(\alpha\beta\gamma) \\ &\qquad + \frac{6(w-1)}{W-1}\Big\{B_4(d) + (m-3)B_3(d)\Big\}\Big] \\ &= v\Big[\Big\{6 + \frac{2(w-1)}{W-1}(3-3m)\Big\}B_3(d) + \frac{2(w-1)}{W-1} \sum_{(\alpha,\beta,\gamma)\in\Delta(3)} (s_\alpha + s_\beta + s_\gamma)B(\alpha\beta\gamma) \\ &\qquad + \frac{6(w-1)}{W-1}\Big\{B_4(d) + (m-3)B_3(d)\Big\}\Big] \end{aligned}$$

$$= \quad 6v\Big[\Big\{1-\frac{2(w-1)}{W-1}\Big\}B_3(d)+\frac{(w-1)}{W-1}B_4(d) \\ +\frac{(w-1)}{3(W-1)}\sum_{(\alpha,\beta,\gamma)\in\Delta(3)}(s_\alpha+s_\beta+s_\gamma)B(\alpha\beta\gamma)\Big]. \tag{A.35}$$

If $s_1 = s_2 = \ldots = s_m = s$, then (A.35) becomes

$$6v\Big[\Big\{1-\frac{2(w-1)}{W-1}\Big\}B_3(d)+\frac{(w-1)}{W-1}B_4(d)+\frac{(w-1)s}{(W-1)}B_3\Big] \\ = 6v\Big[\Big\{1+\frac{(w-1)}{W-1}(s-2)\Big\}B_3(d)+\frac{(w-1)}{W-1}B_4(d)\Big],$$

which agrees with (A.33).

APPENDIX B

BAYESIAN FACTOR SCREENING AND RESPONSE SURFACE DESIGNS

B.1 Derivation of Hellinger Distance

We now sketch the derivation of the Hellinger distance between predictive densities. We refer to the model and priors defined in Section 2 to save space.

Let Y be the $n \times 1$ vector of independent observations from the linear model in (20). The model matrix, X_i is an $n \times r_i$ matrix with the first column corresponding to the intercept and the remaining $r_i - 1$ columns corresponding to the factorial effects in model M_i. The prior specification for the coefficients vector for model M_i is $\pi(\beta_i|\sigma^2) \sim MVN(0, \sigma^2\Gamma_i)$, where Γ is defined in (32). Therefore, the predictive distribution of Y is normal with mean 0 and variance $\sigma^2\Sigma_i$, where Σ_i is defined in (34).

Following the outline in Meyer, Steinberg and Box (1996), we proceed conditionally on σ^2 and integrate out σ^2 in the last step. Let f_i and f_j be the predictive densities of models M_i and M_j respectively. The Hellinger distance between the predictive densities is

$$H(f_i, f_j) = 2 - 2\int (f_i\ f_j)^{1/2} dY.$$

We now need to integrate $(f_i\ f_j)^{1/2}$ over the data to compute the Hellinger distance.

$$\begin{aligned}
\int (f_i\ f_j)^{1/2} dY &= \int \frac{\exp\{-\frac{1}{2}(Y'\frac{\Sigma_i^{-1}}{2\sigma^2}Y + Y'\frac{\Sigma_j^{-1}}{2\sigma^2}Y)\}}{(2\ \pi)^{n/2}|\sigma^2\Sigma_i|^{1/4}|\sigma^2\Sigma_j|^{1/4}} dY \\
&= \int \frac{\exp\{-\frac{1}{2}Y'(\frac{(\Sigma_i^{-1}}{2\sigma^2} + \frac{\Sigma_j^{-1}}{2\sigma^2})Y\}}{(2\ \pi)^{n/2}|\sigma^2\Sigma_i|^{1/4}|\sigma^2\Sigma_j|^{1/4}} dY \\
&= \frac{\left|\left(\frac{\Sigma_i^{-1}}{2\sigma^2} + \frac{\Sigma_j^{-1}}{2\sigma^2}\right)^{-1}\right|^{1/2}}{|\sigma^2\Sigma_i|^{1/4}|\sigma^2\Sigma_j|^{1/4}} \int \frac{\exp\{-\frac{1}{2}Y'(\frac{(\Sigma_i^{-1}}{2\sigma^2} + \frac{\Sigma_j^{-1}}{2\sigma^2})Y\}}{(2\ \pi)^{n/2}\left|\left(\frac{\Sigma_i^{-1}}{2\sigma^2} + \frac{\Sigma_j^{-1}}{2\sigma^2}\right)^{-1}\right|^{1/2}} dY
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\left|\left(\frac{\Sigma_i^{-1}}{2\sigma^2}+\frac{\Sigma_j^{-1}}{2\sigma^2}\right)^{-1}\right|^{1/2}}{|\sigma^2\Sigma_i|^{1/4}|\sigma^2\Sigma_j|^{1/4}} \\
&= \frac{1}{|\sigma^2\Sigma_i|^{1/4}|\sigma^2\Sigma_j|^{1/4}\left|\frac{\Sigma_i^{-1}}{2\sigma^2}+\frac{\Sigma_j^{-1}}{2\sigma^2}\right|^{1/2}} \\
&= \frac{1}{|\sigma^2\Sigma_i|^{1/4}\left|\frac{1}{2}\left(\frac{\Sigma_i^{-1}}{\sigma^2}+\frac{\Sigma_j^{-1}}{\sigma^2}\right)\right|^{1/2}|\sigma^2\Sigma_j|^{1/4}} \\
&= \frac{1}{\left|\frac{1}{2}\left(\Sigma_i^{-1/2}\Sigma_j^{1/2}+\Sigma_i^{1/2}\Sigma_j^{-1/2}\right)\right|^{1/2}}
\end{aligned}
$$

Lastly, we need to integrate over σ^2. Notice that σ^2 cancels out from the derivation of the Hellinger distance. Therefore, selecting any proper prior distribution (say an inverted gamma distribution) will simply result in the Hellinger distance in (31). If an improper prior is selected, then the Hellinger distance may not be bounded above by 2. Substituting the expression in the last step for $\int(f_i\ f_j)^{1/2}dY$ in the Hellinger distance, we get the expression in (33).

B.2 Expected Number of Active Effects

Here, the expected number of active effects in a model is derived for a general form of prior (24). There are q factors and $\binom{q}{2}$ two-way interactions being considered in the model. Let p be the probability that a specific main effect is active. The probability $p_{AB,i}$ that a specific interaction (say AB) is active, given i main effect parents are active is

$$
p_{AB,i} = \begin{cases} c_1 p & \text{if} \quad i = 0 \\ c_2 p & \text{if} \quad i = 1 \\ c_3 p & \text{if} \quad i = 2. \end{cases} \tag{B.1}
$$

The probability $p_{A^2,A}$ that a specific quadratic effect (say A^2) is active, given the main effect A is active is

$$
p_{A^2,A} = \begin{cases} d_1 p & \text{if} \quad \delta_A = 0 \\ d_2 p & \text{if} \quad \delta_A = 1. \end{cases} \tag{B.2}
$$

Conditional on f active main effects, the expected number of active effects (main effects plus two-way interactions) is

$$f + \binom{f}{2} c_3 p + f(q-f)c_2 p + \binom{q-f}{2} c_1 p + f d_2 p + (q-f) d_1 p \tag{B.3}$$

This is because $\binom{f}{2}$ interactions will have two active parents, $f(q-f)$ interactions will have one active parent, and $\binom{q-f}{2}$ interactions will have no active parents. Also, f quadratic effects correspond to the f main effects and there are $(q-f)$ quadratic effects which does not have the corresponding main effect in the model. Expanding (B.3) yields an expected total number of terms (including main effects) as

$$\begin{aligned} \mathrm{E}(\#\text{effects} \quad &| \quad f \text{ active main effects}) \\ &= c_1 pq(q-1)/2 + qpd_1 \\ &+ f\left[1 + \frac{p}{2}(c_1 - c_3) + pq(c_2 - c_1) + (d_2 - d_1)p\right] \\ &+ f^2 \frac{p}{2}\left[c_1 - 2c_2 + c_3\right]. \end{aligned} \tag{B.4}$$

Since f is Binomial with q trials and probability of success p, we have $\mathrm{E}(f) = pq$ and $\mathrm{E}(f^2) = pq(1 - p + pq)$. Taking the expectation of (B.4) with respect to f yields

$$\begin{aligned} \mathrm{E}(\#\text{effects}) &= c_1 pq(q-1)/2 + pqd_1 \\ &+ pq\left[1 + \frac{p}{2}(c_1 - c_3) + pq(c_2 - c_1) + (d_2 - d_1)p\right] \\ &\quad + pq(1 - p + pq)\frac{p}{2}\left[c_1 - 2c_2 + c_3\right]. \end{aligned}$$

Further simplification yields

$$\begin{aligned} \mathrm{E}(\#\text{effects}) &= pq + p\binom{q}{2}\left\{c_1 + 2p(c_2 - c_1) + p^2(c_1 - 2c_2 + c_3)\right\} \\ &+ pq\left[(d_2 - d_1)p + d_1)\right]. \end{aligned} \tag{B.5}$$

For specified values of q, c_1, c_2, c_3 and an expected number of effects, this cubic in p can easily be solved for p. Note that the expected number of main effects is pq, the first term of (B.5).

B.3 HD Optimal Design

Table 15: 18-run HD optimal design

1	2	0	0	2	2	2
0	0	0	0	0	0	0
0	0	2	0	2	2	0
0	1	0	2	0	2	0
1	0	2	0	0	0	2
0	2	1	2	0	2	2
0	2	2	0	2	1	2
2	0	2	2	2	2	2
0	2	2	2	2	0	0
2	0	0	0	2	1	0
2	2	2	0	0	2	0
2	1	1	1	1	2	1
2	0	0	2	0	2	2
2	2	0	0	0	0	2
0	0	0	2	1	0	2
2	2	2	2	0	0	1
2	2	0	2	2	0	0
1	0	0	1	2	0	1

APPENDIX C

ALIASING RELATIONS OF MIXED FACTORIALS IN THE FORM OF PRODUCT ARRAYS

Definition 1 Following Bose (1947), we give the definition for treatment contrasts belonging to factorial effects for the general case of an $s_1 \times \ldots \times s_n$ factorials. A treatment contrast

$$\sum_{j_1=0}^{s_1-1} \cdots \sum_{j_n=0}^{s_n-1} \ell(j_1 \ldots j_n)\tau(j_1 \ldots j_n)$$

belongs to the *factorial effect* $F_{i_1} \ldots F_{i_g}$ $(1 \leq i_1 < \ldots < i_g \leq n; 1 \leq g \leq n)$ if

(i) $\ell(j_1 \ldots j_n)$ depends only on $j_{i_1}, \ldots, j_{i_g}$, and

(ii) writing $\ell(j_1 \ldots j_n) = \overline{\ell}(j_{i_1} \ldots j_{i_g})$ in view of (i) above, the sum of $\overline{\ell}(j_{i_1} \ldots j_{i_g})$ separately over each of the arguments $j_{i_1}, \ldots, j_{i_g}$, is zero.

The following two lemmas are immediate.

Lemma A.1 *Let* $(a, b) = (a_1, \ldots, a_{n_1}, b_1, \ldots, b_{n_2})'$ *be any fixed nonnull vector where* $a_i \in GF(s_1)$ *and* $b_j \in GF(s_2)$. *Then each of the sets*

$$V_{i,j}(a, b) = \{(x, y) = (x_1, \ldots, x_{n_1}, y_1, \ldots, y_{n_2},)' : a'x = \alpha_i, b'y = \beta_j\} \tag{C.1}$$

$0 \leq i \leq s_1 - 1$, $0 \leq j \leq s_2 - 1$ *has cardinality* $s_1^{n_1-1} s_2^{n_2-1}$.

Lemma A.2 *If* $(a^{(1)}, b^{(1)})$ *and* $(a^{(2)}, b^{(2)})$ *are distinct pencils, then for every* $(i, j), (i', j')$, $(0 \leq i, i' \leq s_1 - 1, 0 \leq j, j' \leq s_2 - 1)$, *the set* $V_{i,j}(a^{(1)}, b^{(1)}) \cap V_{i',j'}(a^{(2)}, b^{(2)})$ *has cardinality* $s_1^{n_1-2} s_2^{n_2-2}$.

C.1 Proof of Result 1.

(a)

Consider distinct pencils, (a, b) and (a^*, b^*). Let

$$L = \sum_{i=0}^{s_1-1} \sum_{j=0}^{s_2-1} l(i,j)\{ \sum_{(x,y)\in V_{i,j}(a,b)} \tau(x,y)\}, \tag{C.2}$$

and

$$L^* = \sum_{k=0}^{s_1-1} \sum_{l=0}^{s_2-1} l^*(k,l)\{ \sum_{(x,y)\in V_{k,l}(a^*,b^*)} \tau(x,y)\}, \tag{C.3}$$

be the treatment contrasts belonging to (a, b) and (a^*, b^*), respectively. Here

$$\sum_{i=0}^{s_1-1} l(i,j) = \sum_{j=0}^{s_2-1} l(i,j) = 0, \tag{C.4}$$

and

$$\sum_{k=0}^{s_1-1} l^*(k,l) = \sum_{l=0}^{s_2-1} l^*(k,l) = 0. \tag{C.5}$$

Consider the scalar product of the coefficient vectors in (C.2) and (C.3). Observe that, for any (i, j) and (k, l), if $(x, y) \in V_{i,j}(a, b) \cap V_{i,j}(a^*, b^*)$, then the contribution of $\tau(x, y)$ to this scalar product equals $l(i, j)l^*(k, l)$. Hence the scalar product equals

$$\sum\sum l(i,j)l^*(k,l)\#\{V_{i,j}(a,b) \cap V_{k,l}(a^*,b^*)\}$$

which is zero by (C.4), (C.5) and Lemma A.2.

(b)

Without loss of generality, let $i_1 = 1, \ldots, i_g = g$ and $j_1 = 1, \ldots, j_h = h$. Then $a_1, \ldots, a_g$ are nonzero while $a_{g+1} = \ldots = a_{n_1} = 0$ and $b_1, \ldots, b_h$ are nonzero while $b_{h+1} = \ldots = b_{n_2} = 0$, so that

$$V_{i,j}(a,b) = \{(x,y) : \sum_{k=1}^{g} a_k x_k = \alpha_i, \sum_{l=1}^{h} b_l y_l = \beta_j\}$$

$0 \le i \le s_1 - 1$ and $0 \le j \le s_2 - 1$. Recalling the definition of a treatment contrast L in equations (1) and (2), it is easy to see that for and (x, y), the coefficient of $\tau(x, y)$ in L depends

on (x, y) only through $x_1, \ldots, x_g$ and $y_1, \ldots, y_h$. In fact, writing $\bar{l}(x_1, \ldots, x_g, y_1, \ldots, y_h)$ for the coefficient of $\tau(x, y)$ in L, one gets

$$\begin{aligned} &\bar{l}(x_1, \ldots, x_g, y_1, \ldots, y_h) = l(i, j) \\ &\text{if } \sum_{k=1}^{g} a_k x_k = \alpha_i, \sum_{l=1}^{h} b_l y_l = \beta_j, \end{aligned} \tag{C.6}$$

$0 \leq i \leq s_1 - 1$ and $0 \leq j \leq s_2 - 1$. Now, as $a_1 \neq 0$, the quantity $\sum_{k=1}^{g} a_k x_k$ equals each of $\alpha_0, \alpha_1, \ldots, \alpha_{s_1-1}$ once as x_1 assumes all possible values over $GF(s_1)$, each exactly once, for any fixed $x_2, \ldots, x_g, y_1, \ldots, y_h$. Hence by (C.6)

$$\sum_{x_1 \in GF(s_1)} \bar{l}(x_1, \ldots, x_g, y_1, \ldots, y_h) = 0,$$

for any fixed $x_2, \ldots, x_g, y_1, \ldots, y_h$. Similar arguments for other x_k and y_l's complete the proof.

C.2 Proof of Lemma 1.

$$L(B) = \sum_{i=0}^{s_1-1} \sum_{j=0}^{s_2-1} l(i, j) \{ \sum_{(x,y) \in V_{i,j}((a,b),B)} \tau(x, y) \},$$

and

$$L^*(B) = \sum_{i=0}^{s_1-1} \sum_{j=0}^{s_2-1} l^*(i, j) \{ \sum_{(x,y) \in V_{i,j}((a^*,b^*),B)} \tau(x, y) \}.$$

Therefore, it is enough to show that

$$V_{i,j}((a, b), B) = V_{i,j}((a^*, b^*), B) \forall i, j.$$

Since (a, b) and (a^*, b^*) are are aliases of each other, we have $a - a^* = B_1'\lambda$ and $b - b^* = B_2'\xi$ for suitable $\lambda \in GF(s_1)$ and $\xi \in GF(s_2)$. Now,

$$\begin{aligned} & V_{i,j}((a, b), B) \\ = \; & \{(x, y) : a'x = \alpha_i, b'y = \beta_j, B_1 x = 0, B_2 y = 0\} \\ = \; & \{(x, y) : (a^* + B_1'\lambda)'x = \alpha_i, (b^* + B_2'\xi)'y = \beta_j, \\ & \qquad B_1 x = 0, B_2 y = 0\} \end{aligned}$$

$$
\begin{aligned}
&= \{(x,y) : a^{*\prime}x + \lambda' B_1 x = \alpha_i, b^{*\prime}y + \xi' B_2 y = \beta_j, \\
&\qquad\qquad B_1 x = 0, B_2 y = 0\} \\
&= \{(x,y) : a^{*\prime}x = \alpha_i, b^{*\prime}y = \beta_j, B_1 x = 0, B_2 y = 0\} \\
&= V_{i,j}((a^*, b^*), B)
\end{aligned}
$$

which completes the proof of the lemma.

Table 16: Design Matrix, Paint Experiment

Run	A	B	C	D	E	F
1	1	1	0	1	1	2
2	0	1	1	1	1	2
3	1	0	1	1	1	2
4	0	0	0	1	1	2
5	1	1	0	2	1	0
6	0	1	1	2	1	0
7	1	0	1	2	1	0
8	0	0	0	2	1	0
9	1	1	0	0	1	1
12	0	1	1	0	1	1
11	1	0	1	0	1	1
12	0	0	0	0	1	1
13	1	1	0	1	2	0
14	0	1	1	1	2	0
15	1	0	1	1	2	0
16	0	0	0	1	2	0
17	1	1	0	2	2	1
18	0	1	1	2	2	1
19	1	0	1	2	2	1
20	0	0	0	2	2	1
21	1	1	0	0	2	2
22	0	1	1	0	2	2
23	1	0	1	0	2	2
24	0	0	0	0	2	2
25	1	1	0	1	0	1
26	0	1	1	1	0	1
27	1	0	1	1	0	1
28	0	0	0	1	0	1
29	1	1	0	2	0	2
30	0	1	1	2	0	2
31	1	0	1	2	0	2
32	0	0	0	2	0	2
33	1	1	0	0	0	0
34	0	1	1	0	0	0
35	1	0	1	0	0	0
36	0	0	0	0	0	0

APPENDIX D

SEQUENTIAL ELIMINATION OF LEVEL COMBINATIONS BY MEANS OF MODIFIED GENETIC ALGORITHMS

D.1 Identification of Significant factors : A Bayesian Approach

The model selection problem amounts to identifying a subset of predictors as active. In this setting there are typically many parameters to estimate. Here we review a stochastic variable selection method, based on Gibbs sampler. Starting with the given design and the corresponding responses, the linear regression with normal errors is

$$y = X\beta + \sigma\epsilon, \qquad \epsilon \sim N(0,1) \tag{D.1}$$

where β contains linear and quadratic main effects and linear-by-linear interaction effects. Importance of effects is captured via an unobserved vector δ of zeros and ones where $\delta_i = I\{\beta_i \neq 0\}$. A normal mixture prior is used for the coefficients β :

$$f(\beta_i|\delta_i) = \begin{cases} N(0,\tau_i^2), & \text{if } \delta_i = 0, \\ N(0,(c_i\tau_i)^2), & \text{if } \delta_i = 1. \end{cases} \tag{D.2}$$

When $\delta_i = 0$, β_i has a large mass around zero and thereby, is not likely to have a large effect. On the other hand, when $\delta_i = 1$, a large value of c_i ensures that the variable is likely to have a large influence.

The Bayesian method finds posterior probabilities of β's. Details are given by Chipman et al. (1997). The hierarchical priors on the linear, quadratic and linear-by-linear interaction effects of the factors reflect the common beliefs like *effect sparsity, effect hierarchy* and *effect inheritance* (Wu and Hamada, 2000). Priors used in the current analysis are given next:

$$P(\delta_A = 1) = p, \tag{D.3}$$

$$P(\delta_{A^2} = 1|\delta_A) = \begin{cases} 0.1p & \text{if} \quad \delta_A = 0, \\ p & \text{if} \quad \delta_A = 1, \end{cases} \tag{D.4}$$

$$P(\delta_{AB} = 1|\delta_A, \delta_B) = \begin{cases} 0.1p & \text{if} \quad \delta_A + \delta_B = 0, \\ 0.5p & \text{if} \quad \delta_A + \delta_B = 1, \\ p & \text{if} \quad \delta_A + \delta_B = 2. \end{cases} \tag{D.5}$$

In the current analysis, $p = 0.25$ is chosen. Following George and McCulloch (1993), $\tau_j = \frac{\Delta y}{3\Delta X_j}$ is taken where Δy represents a "small" change in y, and ΔX_j represents a large change in X_j. In our example, $\Delta X_j = \max(X_j) - \min(X_j)$ and $\Delta y = \sqrt{\text{Var}(y)}/5$ is used. The posterior probabilities of β's are computed using Gibbs sampler.

VITA

Abhyuday Mandal was born in India. He got a Bachelor's and a Master's degree in Statistics from Indian Statistical Institute, Calcutta, India and another Master's degree in Statistics from University of Michigan, USA. He received the Ph.D. degree in 2005 from Georgia Institute of Technology. In August 2005, he will join University of Georgia as an Assistant Professor at the Department of Statistics.