

# SLICED DESIGNS FOR MULTI-PLATFORM ONLINE EXPERIMENTS

by

Soheil Sadeghi

A dissertation submitted in partial fulfillment of  
the requirements for the degree of

Doctor of Philosophy

(Statistics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 04/19/2017

The dissertation is approved by the following members of the Final Oral Committee:

Zhiguang Qian, Professor, Statistics

Neeraj Arora, Professor, Marketing

Kam-Wah Tsui, Professor, Statistics

Anru Zhang, Assistant Professor, Statistics

Paul Hoban, Assistant Professor, Marketing

© Copyright by Soheil Sadeghi 2017  
All Rights Reserved

*To my mom*

## ACKNOWLEDGMENTS

---

*Thanks to Peter Qian and Neeraj Arora for being exceptional advisors and mentors.*

*Thanks to UW–Madison Statistics department for providing an enjoyable experience to further my research.*

*I thank my college advisor, Hashem Mahlooji, for being a wonderful support of all my professional endeavors.*

*Finally, I appreciate my mom, Shahnaz Salimian, for supporting my natural curiosity and for encouraging me to follow my passion.*

— SOHEIL SADEGHI 2017

## CONTENTS

---

Contents iii

List of Tables iv

List of Figures v

Abstract vi

1 Introduction 1

2 Factorial Designs at Two Levels 8

2.1 *Model-Dependent vs. Model-Free Criteria* 12

2.2 *Properties and Principles of Factorial Designs* 17

2.3 *Construction of Fractional Factorial Designs* 18

3 Formulation of Designs for Multi-Platform Experiments 24

3.1 *Illustration* 29

3.2 *Sliced Effect Hierarchy and Existing Design Methods* 31

4 Sliced Factorial Designs for Two-Platform Experiments 33

5 Sliced Factorial Designs for Four-Platform Experiments 40

6 Two Algorithms to Construct Sliced Factorial Designs 49

7 Empirical Application 58

8 Conclusion 71

References 74

## LIST OF TABLES

---

2.1	Model Matrix and Data for $2^3$ Design . . . . .	11
2.2	Constructing the $2^{4-1}$ Fractional Factorial Design . . . . .	19
2.3	Confounded Effects for the $2^{4-1}$ Design <b>I = 1234</b> . . . . .	20
3.1	Effect Hierarchy Principle for Each Sub Design of the Experiment in Example 1 . . . . .	30
3.2	Sliced Effect Hierarchy Principle for the Complete Design of the Experiment in Example 1 . . . . .	30
5.1	Rule for Replacing any Three Columns of the Form (a, b, ab) by the 4-level Column S . . . . .	41
5.2	Sliced Minimum Aberration Designs with 16 Runs, $S = (s_1, s_2, s_1s_2)$ . .	48
5.3	Sliced Minimum Aberration Designs with 32 Runs, $S = (s_1, s_2, s_1s_2)$ . .	48
5.4	Sliced Minimum Aberration Designs with 64 Runs, $S = (s_1, s_2, s_1s_2)$ . .	48
6.1	Rank of $2^{9-3}$ Orthogonal Sliced Factorial Designs for Example 3 with Three Generators $g_1$ : <b>13458</b> , $g_2$ : <b>1247</b> , and $g_3$ : <b>1236</b> . . . . .	57
7.1	Six Binary Design Factors for Update Email Optimization . . . . .	67
7.2	Versions One, Two, Three, and Four - Update Spring 2015 . . . . .	68
7.3	Versions Five, Six, Seven, and Eight - Update Spring 2015 . . . . .	69
7.4	Descriptive Statistics for Empirical Application . . . . .	70
7.5	Confounded Effects . . . . .	70
7.6	Results for $P_1$ . . . . .	70
7.7	Results for $P_2$ . . . . .	70
7.8	Slice Factor Behavior . . . . .	70

**LIST OF FIGURES**

---

1.1	Liberty Mutual Insurance Homepage. . . . .	3
3.1	Design Set d of the Experiment in Definition 3.1. . . . .	27
6.1	Flow Chart of Algorithm 2. . . . .	55
7.1	Next Issue of Update After the Study- Fall 2015. . . . .	66

## ABSTRACT

---

Multivariate testing is a popular method to improve websites, mobile apps, and email campaigns. A unique aspect of testing in the online space is that it needs to be conducted across multiple platforms such as a desktop and a smartphone. The existing experimental design literature does not offer precise guidance for such a multi-platform context. In this paper, we introduce a multi-platform design framework that allows us to measure the effect of the design factors for each platform and the interaction effect of the design factors with platforms. Substantively, the resulting designs are of great importance for testing digital campaigns across platforms. We illustrate this in an empirical email application to maximize engagement for a digital magazine. We introduce a novel “sliced effect hierarchy principle” and develop design criteria to generate factorial designs for multi-platform experiments. To help construct such designs, we prove two theorems that connect the proposed designs to the well-known minimum aberration designs. We find that experimental versions made for one platform should be similar to other platforms. From the standpoint of real world application, such homogeneous sub-designs are cheaper to implement. To assist practitioners, we provide two algorithms to construct the designs that we propose. We also tabulate sliced factorial designs with 16, 32, and 64 runs for four-platform experiments.

# 1 INTRODUCTION

---

A highly desirable characteristic of marketing activities online is that they lend themselves to rapid and extensive testing. Online retailers routinely improve the layout of their website to maximize profitability. Websites intended for the purpose of educating visitors evaluate user engagement on an ongoing basis and attempt to improve metrics that include page views and time spent per page. The simplest form of online experimentation is A/B testing. It has become quite popular because tools such as Google Analytics make it easy to implement. A more informative form of online testing is multivariate testing because it allows one to assess the individual effect of multiple factors at the same time. The primary focus of this paper is on constructing new designs for multivariate testing for online experiments.

Online testing is a popular and rapidly growing method in the digital world for applications that include website and email optimization. A report by Forrester evaluates a variety of suppliers that include Adobe, Maxymiser and Opimizely which are well known in the online testing area (Stanhope, 2013). For website optimization, consider, for example, the homepage of Liberty Mutual Insurance shown in Figure 1.1. A typical A/B testing may begin by creating two versions of the website: version A and version B. To implement the test, the website traffic is divided into two disjoint sets. Engagement metrics such as the number of pages navigated determine the winner of the test.

A more informative form of testing involves multiple attributes of the website. In a very simple example, consider the following four website factors, each at two levels (present versus absent): the login tab on top right, the 'Resources tab' below it, 'find a local agent' tab in the middle, and the 'Auto Savings and Discounts' visual at

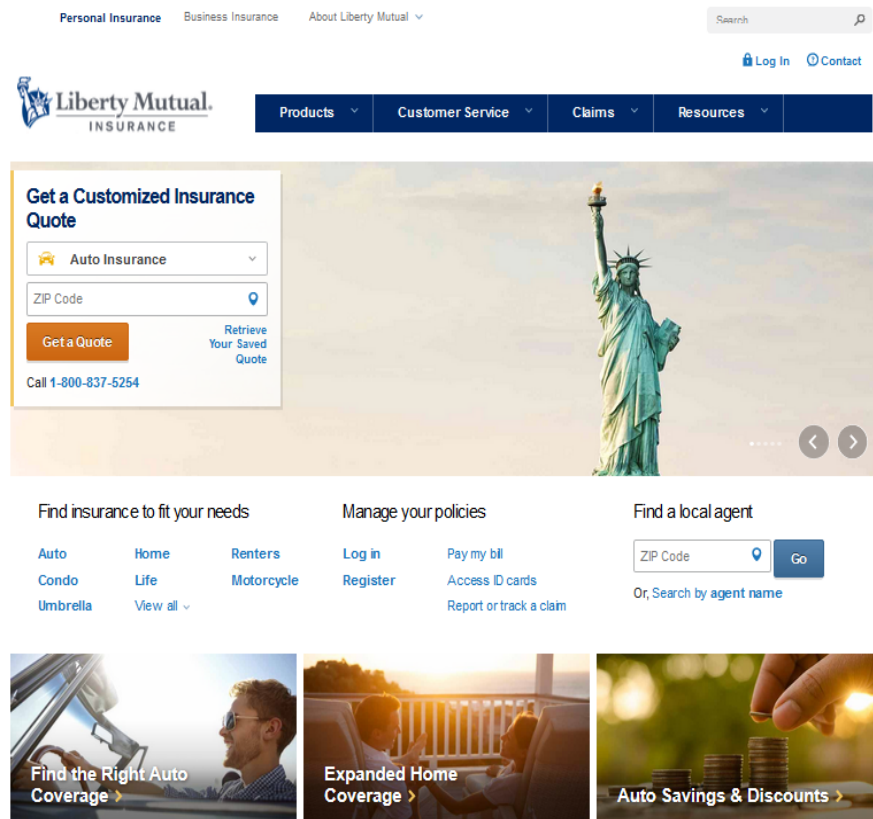


Figure 1.1: Liberty Mutual Insurance Homepage.

bottom. To find the best combination of these four attributes, one can create sixteen versions of the website for a full factorial design. For an engagement metric such as page views, this multivariate test can help uncover the incremental contribution of each factor and as a result help improve the website layout. As the number of attributes increase, a fraction of all combinations could be used in order to perform the multivariate test (Wu and Hamada, 2011). This is often necessary to ensure feasibility and to address cost constraints.

Multivariate experiments in the online space present a new design challenge:

experiment needs to be conducted across multiple platforms that include laptops, tablets, smartphones, and smartwatches. Such experiments are important because a different set of attribute combinations may be optimal for each platform. For example, presence of multiple images may be best for a laptop and in contrast a list of links may be more effective for a smartphone. In this paper, we introduce a new design, called the *sliced factorial design*, for such multi-platform, multivariate online experiments. We develop optimality criteria to construct such designs. This complete design we propose can be partitioned into sub-designs or slices such that each slice is used for one platform.

We have chosen to use the term “slice” because this is how it has been used in the design literature historically. For example, a sliced Latin hypercube design in Qian (2012) can be divided into slices of smaller Latin hypercube designs. Such a design is attractive for building emulators for computer models and performing numerical integration. Similarly, sliced designs constructed by Xu et al. (2011) and Qian (2012) are intended for running simulation experiments and building Gaussian process prediction models. In contrast with extant work, the sliced factorial design proposed by us is intended for performing multi-platform experiments in the digital space.

To establish terminology and to build upon existing literature on the topic, we begin the paper with a detailed review of resolution and aberration based optimality criteria for factorial designs. Both resolution and aberration criteria are built upon three fundamental and widely used principles: effect hierarchy, effect sparsity, and effect heredity (Wu and Hamada, 2011). In our design framework we ensure that each slice or sub-design follows these fundamental principles. In order to account

for the unique multi-platform context, we generalize the effect hierarchy principle to the *sliced effect hierarchy* principle. Our main idea is that each sub design allows accurate assessment of factors that are significant for the corresponding platform and the complete design allows us to identify the interaction between the design factors and the platforms. We impose a reasonable ordinal structure on the effects the experiment is trying to uncover in the cross-platform context. Based on the sliced effect hierarchy principle, we develop extended design criteria called *sliced resolution* and *sliced aberration*. These criteria are then used to generate the sliced factorial designs to perform a multi-platform, multivariate test.

Minimum aberration designs are popular among practitioners and tables to construct them are readily available in software and textbooks (Wu and Hamada, 2011). In an effort to build upon the rich results of the aberration literature, we prove two theorems that connect sliced factorial designs to minimum aberration designs. Acting as the bridge between what we already know, these theorems help construct sliced factorial designs that we propose. Using the extended design criteria and the theorems, we develop two algorithms to construct sliced factorial designs. The first algorithm is for the case where all the combinations of design factors are feasible for all the platforms. This algorithm works for both symmetric and asymmetric designs. The second algorithm is usable in situations where some combinations of design factors may be infeasible for some platforms. This incorporates a typical design constraint wherein not all factors combinations may be feasible for all platforms. An example of such a design constraint on a smartphone would be the inability to include two image based factors, a brand logo and the picture of a child, on one

screen. The same combination may work quite well on a desktop.

Based on our new design criteria, we find that it is desirable to have the sliced factorial design be divided into homogeneous slices: experimental versions made for a slice should be as similar as possible to the ones for other slices. In terms of resources, the multivariate design we propose requires a smaller number of versions compared to other designs. We illustrate that although slices are used to uncover factor effects within each platform, the sliced factorial design can be used to compare the results across different platforms.

Finally, we illustrate our novel design framework in the context of an empirical email optimization application intended to maximize engagement for a digital magazine. In this application we selected six binary attributes and measured design effects across two platforms. The dependent variable of interest was page views and involved over 25,000 users. Because of extensive programming related constraints, the maximum number of versions was restricted to eight. Algorithm 1 was used to generate a sliced factorial design for this study. Google Analytics was used to record the number of page views for each version. Lenth's test, which is well suited for such unreplicated data, revealed interesting insights about how different factors were effective for different platforms. Based on the results for this multi-platform experiment, the expected gain in page views for the two platforms was 16% and 7%.

The remainder of the paper is organized as follows. In Section 2 we lay a foundation for our work by reviewing key design concepts upon which we build our framework, and highlight how our research differs from the literature. In

Section 3 we formulate the multi-platform experiments and explain the sliced hierarchy principle. The following three sections introduce two design criteria that are useful to construct the new designs we propose and present two algorithms to construct them. In Section 7, we present an empirical application as an illustration for our design framework. Finally, we offer our conclusions in the last section of the paper.

## 2 FACTORIAL DESIGNS AT TWO LEVELS

---

Before presenting our design framework in the next section, we lay the foundation by beginning with a description of what factorial designs are, the fundamental principles that guide factorial designs and formal criteria to construct them. Factorial designs are often used in studies where the interest is to model the effects of more than one factor simultaneously. These designs are well studied in statistics. Wu and Hamada (2011), Montgomery (2008), and Box et al. (1978) are excellent references on this subject. Ledolter and Swersey (2007) is also a great reference that aims to provide a review of experimental designs with applications in marketing. To serve as a basis for further development, in this section we provide a brief review of factorial designs and the relevant concepts. For ease of presentation, we begin with the case that includes  $k$  two-level factors, denoted by  $1, \dots, k$ . For each factor, its linear effect on a response variable, denoted by  $y$ , is investigated. A *full factorial design* requires  $2^k$  runs for the  $k$  factors. Denote  $\mathbf{y}$  as the vector of responses of length  $2^k$ , and  $\mathbf{X}$  as the  $2^k \times 2^k$  model matrix for which the  $2^k$  columns consist of the column of ones  $\mathbf{x}_0$ ,  $k$  columns of  $\pm 1$  values  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$  representing the design vectors,  $\binom{k}{2}$  columns of two-way interactions  $\mathbf{x}_{12}, \mathbf{x}_{13}, \dots$  (each a product of two design columns),  $\binom{k}{3}$  columns of three-way interactions  $\mathbf{x}_{123}, \mathbf{x}_{124}, \dots$  (each a product of three design columns), ..., all the way to the column of  $k$ -way interaction  $\mathbf{x}_{12\dots k}$  (the product of all  $k$  design columns).

The regression model can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}, \quad (2.1)$$

which does not include the error term as this is a fully saturated model, with the

same number of coefficients as number of observations. The regression coefficient  $\beta$  consists of elements  $\beta_0$  (the intercept),  $\beta_1, \dots, \beta_k$  (main effects),  $\beta_{12}, \beta_{13}, \dots$  (two-way interaction effects), ...,  $\beta_{12\dots k}$  (the  $k$ -way interaction effect). Obviously, not all effects need to be included in the model. For instance, a model of main effects only includes  $k$  main effects and a noise component that reflects all the interaction effects. The least square estimate of the regression coefficient  $\beta$  is given by

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \quad (2.2)$$

where  $\mathbf{X}'$  is the transpose of the matrix  $\mathbf{X}$ , and  $(\mathbf{X}'\mathbf{X})^{-1}$  is the inverse of the matrix  $\mathbf{X}'\mathbf{X}$ . For a  $2^k$  full factorial design, all diagonal elements of  $\mathbf{X}'\mathbf{X}$  are equal to  $2^k$ , and off-diagonal elements are zero. Therefore, the estimate of an element of  $\beta$ , e.g.  $\hat{\beta}_{ij\dots l}$ , is given by

$$\hat{\beta}_{ij\dots l} = \frac{1}{2^k} \sum_{t=1}^{2^k} c_t y_t = \frac{1}{2} \left( \frac{1}{2^{k-1}} \sum_{t=1}^{2^k} c_t y_t \right), \quad (2.3)$$

where the weights  $c_t$ 's are the  $\pm 1$  elements of the corresponding vector column  $\mathbf{x}_{ij\dots l}$ . For a  $2^k$  full factorial design, half of the elements of each column  $\mathbf{x}_{ij\dots l}$  (excluding the column of ones  $\mathbf{x}_0$ ) are equal to  $+1$ , and the rest half of elements are equal to  $-1$ . The model matrix (excluding the column of ones) of a full  $2^3$  factorial design, provided in the first seven columns of Table 2.1, is an example. Hence, the estimate

$\hat{\beta}_{ij\dots l}$  can be written as

$$\hat{\beta}_{ij\dots l} = \frac{1}{2} \left( \frac{1}{2^{k-1}} \sum_{t=1}^{2^k} c_t y_t \right) = \frac{1}{2} \left( \bar{y}(\mathbf{x}_{ij\dots l+}) - \bar{y}(\mathbf{x}_{ij\dots l-}) \right), \quad (2.4)$$

where  $\bar{y}(\mathbf{x}_{ij\dots l+})$  is the average of the  $y_t$  values at (+) level of column  $\mathbf{x}_{ij\dots l}$ , and  $\bar{y}(\mathbf{x}_{ij\dots l-})$  is the average of the  $y_t$  values at (-) level of column  $\mathbf{x}_{ij\dots l}$ . In the factorial design literature, usually  $\bar{y}(\mathbf{x}_{ij\dots l+}) - \bar{y}(\mathbf{x}_{ij\dots l-})$  is reported as the estimate, denoted by **ij ... k** effect, and not the regression coefficient estimate  $\hat{\beta}_{ij\dots l}$ , i.e.,

$$\mathbf{ij \dots k \text{ effect}} = \bar{y}(\mathbf{x}_{ij\dots l+}) - \bar{y}(\mathbf{x}_{ij\dots l-}). \quad (2.5)$$

For example, in Table 2.1, main effect **1** is estimated by  $\frac{1}{4}(y_5 + y_6 + y_7 + y_8) - \frac{1}{4}(y_1 + y_2 + y_3 + y_4)$ , and interaction effect **23** is estimated by  $\frac{1}{4}(y_1 + y_4 + y_5 + y_8) - \frac{1}{4}(y_2 + y_3 + y_6 + y_7)$ .

Table 2.1: Model Matrix and Data for  $2^3$  Design

<b>1</b>	<b>2</b>	<b>3</b>	<b>12</b>	<b>13</b>	<b>23</b>	<b>123</b>	Data
$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_{12}$	$\mathbf{x}_{13}$	$\mathbf{x}_{23}$	$\mathbf{x}_{123}$	$\mathbf{y}$
-	-	-	+	+	+	-	$y_1$
-	-	+	+	-	-	+	$y_2$
-	+	-	-	+	-	+	$y_3$
-	+	+	-	-	+	-	$y_4$
+	-	-	-	-	+	+	$y_5$
+	-	+	-	+	-	-	$y_6$
+	+	-	+	-	-	-	$y_7$
+	+	+	+	+	+	+	$y_8$

Full factorial designs are rarely used in practice for large  $k$ . Instead, a fraction

of a full factorial design, called a *fractional factorial design* denoted by  $2^{k-p}$ , is often used. In general,  $2^{k-p}$  denotes a  $(\frac{1}{2})^p$  fraction of a  $2^k$  factorial design. The optimal fraction can be selected according to optimality criteria that we explain next. A design can be replicated more than once and a design with a single replicate is called an unreplicated design.

## 2.1 Model-Dependent vs. Model-Free Criteria

In constructing any experimental designs, it is important to search for the optimal design. In the design literature, there are two major types of criteria that are used for this purpose: model-dependent criteria, e.g. D-optimality, and model-free criteria, e.g. resolution and aberration. To use a model-dependent criterion, only one model is considered. Designs for a single model are usually based on alphabetic criteria such as D-optimality, A-optimality, and E-optimality. A summary of these criteria can be found in Atkinson et al. (2007). Extensions for the case in which the model is uncertain exist in the literature although most of these works focus on small number of alternative models and do not consider many models that can be indeed true. To mention some, Läuter (1974) suggested a weighted average criterion over several models, Zhou et al. (2003) provides a Bayesian interpretation of this for A-optimality, and Heredia-Langner et al. (2004) constructed the optimal design based on this weighted average criterion.

In the Marketing literature, there is a large subfield of experimental designs for discrete choice models which has been significantly grown over the past 20 years

(Louviere and Woodworth, 1983; Huber and Zwerina, 1996; Arora and Huber, 2001; Sándor and Wedel, 2001, 2002, 2005; Kessels et al., 2006; Toubia and Hauser, 2007; Yu et al., 2009; Kessels et al., 2009; Liu and Arora, 2011). The primary focus of this literature is on a particular model, mostly the multinomial logit. For this model, the choice design construction process includes searching for the best design in a high dimensional space using model-dependent criteria and gets conflated with the difficulty that the information matrix is a function of the model parameters. For example, Toubia and Hauser (2007) proposed M-efficiency as a generalization of alphabetic optimality for the case where the focus is on a function of parameters rather than the direct estimates of the parameters in the assumed model. It is usually used when some managerial decisions are of greater interest for managers than others.

One drawback of standard model-dependent criteria is that, in practice, the model is almost never known in advance, and these criteria do not consider the confounding of pre-specified parameters with the potentially significant ones that are missing in the model. To motivate, consider a researcher who is deciding about the design of an  $n$  run experiment with  $k$  design factors. The researcher considers the main effects in the model, but she is worried if the potential significant two-way interactions bias the results. Let  $\mathbf{X}_1$  denote the  $n \times k$  model matrix for main effects and  $\mathbf{X}_2$  denote the  $n \times \binom{k}{2}$  model matrix for the two-way interactions. Suppose the true model is

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}, \quad (2.6)$$

where  $\boldsymbol{\epsilon}$  represents the vector of residuals with mean zero and identity covariance

matrix  $\mathbf{I}$ . Considering the model with main effects only for estimation, i.e.,

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\epsilon}^*, \quad (2.7)$$

the researcher can select the design according to a model-dependent criterion. A-optimality requires minimization of trace  $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$ ; For D-optimality, the optimization is defined over the determinant of  $\mathbf{X}_1'\mathbf{X}_1$ ; the largest eigenvalue of  $(\mathbf{X}_1'\mathbf{X}_1)^{-1}$  is minimized for E-optimality; and M-efficiency considers an alphabetic optimality criterion over  $\mathbf{M}(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{M}'$  which is the covariance matrix of a linear function of main effects, i.e.  $\mathbf{M}\boldsymbol{\beta}_1$ .

The researcher's choice of main effects model in Equation 2.7 provides a least square estimate for  $\boldsymbol{\beta}_1$ ,  $\hat{\boldsymbol{\beta}}_1$ , with the expected value:

$$E(\hat{\boldsymbol{\beta}}_1) = \boldsymbol{\beta}_1 + \mathbf{A}\boldsymbol{\beta}_2, \quad (2.8)$$

where  $\mathbf{A} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$  is known as the alias matrix in the design literature. Therefore, the estimate of main effects will be biased if any two-way interaction terms are significant although it is optimal in terms of variance. The relative tradeoff between bias and variance in design construction has been actively studied in the design literature. Box and Draper, 1959, 1963; Draper and Guttman, 1992; Bursztyn and Steinberg, 2006; Jones and Nachtsheim, 2011 are some examples.

Following Montepiedra and Fedorov (1997), a suitable criterion of goodness of

the estimate  $\hat{\beta}_1$  is provided by the mean squared error matrix

$$\text{MSE}(\hat{\beta}_1|\beta_2) = E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)] = (\mathbf{X}_1' \mathbf{X}_1)^{-1} + \mathbf{A}' \beta_2 \beta_2' \mathbf{A}, \quad (2.9)$$

in which the first term is the variance matrix and the second term is squared bias matrix. Montepiedra and Fedorov (1997) considers minimization of one of the two terms subject to a constraint on the other term for a single known value of  $\beta_2$ . However,  $\beta_2$  is almost never known in advance. Draper and Guttman (1992) suggested a novel idea to turn the bias term into variance by putting a prior distribution on  $\beta_2$ . Draper and Guttman (1992) assumes that  $\beta_2$  has a normal distribution with mean zero and covariance matrix  $\sigma_{\beta_2}^2 \mathbf{I}$ , and shows that the mean squared error matrix only includes the variance term which equals to

$$\text{Var}(\hat{\beta}_1|\sigma_{\beta_2}^2) = (\mathbf{X}_1' \mathbf{X}_1)^{-1} + \sigma_{\beta_2}^2 \mathbf{A}' \mathbf{A}. \quad (2.10)$$

Draper and Guttman (1992) studied the properties of optimal designs as a function of  $\sigma_{\beta_2}^2$ . Using a similar prior assumption on  $\beta_2$ , Bursztyn and Steinberg (2006) suggested the expected value of the squared norm of bias vector as a measure of the potential bias:

$$E\|E(\hat{\beta}_1|\beta_2) - \beta_1\|^2 = E(\beta_2' \mathbf{A}' \mathbf{A} \beta_2) = E[\text{Trace}(\mathbf{A}' \mathbf{A} \beta_2 \beta_2')] = \sigma_{\beta_2}^2 \text{Trace}(\mathbf{A}' \mathbf{A}). \quad (2.11)$$

Therefore, Draper and Guttman (1992) focuses on minimization of  $\text{Trace}(\mathbf{A}' \mathbf{A})$  for the purpose of minimizing the potential bias. In a recent work, Jones and

Nachtsheim (2011) seeks to drive the potential bias toward zero while keeping high D-efficiency. For standard regular factorial designs, Jones and Nachtsheim (2011) shows that this approach leads to an ordering of orthogonal designs which is the same as the way resolution and aberration criteria do. This guides us to introducing model-free criteria that are largely neglected, over the past 20 years, in the marketing literature, especially design literature on ratings based conjoint analysis (Green and Rao, 1971; Green and Srinivasan, 1990, 1978; Elrod et al., 1992). When ratings based conjoint was first used in marketing (Green and Rao, 1971), the default design criterion was orthogonality. In Green (1974), the author discusses a variety of designs relevant to ratings based conjoint: orthogonal, Latin square and blocked design. To the best of our knowledge this paper is the last work in marketing literature on model-free criteria, and orthogonality still continues to be the default design criterion (see for example Orthoplan for conjoint analysis in SPSS). Since the inception of the orthogonality criterion, the statistical design literature for linear models has progressed significantly, with the most widely used design optimality criterion today being minimum aberration (Mukerjee and Wu, 2007, p. 3). In this dissertation, we introduce the rich design literature on model-free criteria to the marketing literature, and build upon it to develop the proposed design framework for our multi-platform context.

In the case of model-free design criteria, resolution and aberration are usually used for regular factorial designs and generalized aberration is used to construct irregular fractions. These criteria focus on estimation of lower-order effects and assume that the true model will be one of the many models containing some or all

of factorial effects. Fries and Hunter (1980) introduced minimum aberration as a natural generalization of maximum resolution, and since then many works have been devoted on finding minimum aberration designs. Fang and Mukerjee, 2000; Butler, 2003; Jacroux, 2004; Cheng and Tang, 2005 are some of recent works on this path. In terms of model robustness, Cheng et al. (1999) studied the performance of minimum aberration designs and showed that this criterion indirectly takes efficiency into account and provides a good surrogate for model robustness in terms of estimation capacity. Before we review the construction of factorial designs using minimum aberration and maximum resolution, we first define some properties and principles of factorial designs that guide their constructions.

## 2.2 Properties and Principles of Factorial Designs

Two key properties of factorial designs are *balance* and *orthogonality*. A design is balanced if each factor level appears in the same number of experimental runs. Two factors are called orthogonal if all their different combinations appear in the same number of experimental runs. A design for which all pair of its factors are orthogonal is called an orthogonal design. As one would expect, a full factorial design is balanced and orthogonal.

After defining the properties, an important question is to understand the relative importance and relationship between the effects. Below we list three fundamental principles that serve as the foundation for factorial designs (Wu and Hamada, 2011, p. 172).

1. **Effect Hierarchy:** This principle indicates that lower-order effects are more likely to be important than higher-order effects. For example, main effects are more likely to be important than interaction effects. Further, effects of the same order are equally likely to be important.
2. **Effect Sparsity:** It indicates that number of relatively important effects in a factorial design is small.
3. **Effect Heredity:** In order for an interaction to be significant, at least one of its parent main effects should be significant.

These fundamental principles guide the construction and analysis of factorial designs. In most situations, the effects of lower order interactions are believed to be more important than the higher order effects. For this reason, a fractional factorial design can be generated by confounding the effects of higher order interactions with the lower order ones. Lower order effects therefore can be estimated by assuming that higher order interactions are negligible. These principles are also invoked in ratings based conjoint applications in marketing. We now review construction methods for fractional factorial designs.

## **2.3 Construction of Fractional Factorial Designs**

*Fractional Factorial Designs Definition:* consider a one-half fraction of the  $2^4$  factorial design. To construct this fractional design, we can first write down a  $2^3$  full factorial design using three factors 1, 2, and 3. The – and + elements associated with the

**123** interaction column then can be used to identify the  $-$  and  $+$  versions of main effect **4** (Table 2.2). The result is a particular half fractional of the  $2^4$  full factorial design. In general,  $2^{k-p}$  denotes a  $(\frac{1}{2})^p$  fraction of a  $2^k$  factorial design. Therefore, the design in Table 2.2 is a  $2^{4-1}$  fractional factorial.

Table 2.2: Constructing the  $2^{4-1}$  Fractional Factorial Design

<b>1</b> $x_1$	<b>2</b> $x_2$	<b>3</b> $x_3$	<b>4 = 123</b> $x_4 = x_{123}$	Data <b>y</b>
$-$	$-$	$-$	$-$	$y_1$
$-$	$-$	$+$	$+$	$y_2$
$-$	$+$	$-$	$+$	$y_3$
$-$	$+$	$+$	$-$	$y_4$
$+$	$-$	$-$	$+$	$y_5$
$+$	$-$	$+$	$-$	$y_6$
$+$	$+$	$-$	$-$	$y_7$
$+$	$+$	$+$	$+$	$y_8$

Next we define several terms and criteria that are necessary to understand the construction process of fractional factorial designs.

*Defining relation and Resolution:* Using the design in Table 2.2 with eight observations, eight estimates can be calculated. Each estimate is actually the estimate of the sum of two effects that are confounded. By construction, **4** is confounded with **123**. Multiplying the elements of a column by the same column will result in a column of plus signs which corresponds to the identity **I**. Therefore, multiplying both sides of **4 = 123** by **4** will result in **I = 1234** which is called the *defining relation* of the design. The interaction **1234** is called the *generator* of the design. Table 2.3 shows all confounded pairs for the  $2^{4-1}$  design **I = 1234** where the *word 1234* has length

4. We will frequently use the fundamental design terms word, length, defining relation, and generator in the remainder of the paper.

Table 2.3: Confounded Effects for the  $2^{4-1}$  Design **I = 1234**

<b>I = 1234</b>	<b>4 = 123</b>
<b>1 = 234</b>	<b>12 = 34</b>
<b>2 = 134</b>	<b>13 = 24</b>
<b>3 = 124</b>	<b>14 = 23</b>

For the construction of the  $2^{4-1}$  fractional factorial design, we have chosen the interaction effect **123** to be confounded with the effect **4**. Any one of the interactions **12**, **13**, **23**, and **123** could be confounded with **4**. We chose **123** based on the *maximum resolution* criterion that we define next.

To construct a  $2^{k-p}$  fractional factorial design, first let **1**, ..., **k - p** denote the  $k - p$  independent columns that generate the  $2^{k-p}$  factorial design. The remaining  $p$  columns, **k - p + 1**, ..., **k** can be generated as interactions of the first  $k - p$  columns. Choice of these  $p$  columns determines the generators and the defining relation of the design. The defining relation of the design consists of the identity element **I** plus the group formed by the  $p$  generators ( $2^p - 1$  words in the group). For example, the defining relation of the  $2^{4-1}$  fractional factorial design in Table 2.2 includes the identity element **I** plus one word **1234**. For a  $2^{k-p}$  design, let  $A_i$  be the number of words of length  $i$  in its defining relation. The *wordlength pattern* of the design is

$$W = (3^{A_3}, \dots, k^{A_k}). \quad (2.12)$$

The resolution, suggested by Box and Hunter (1961), of a  $2^{k-p}$  design is defined

to be the smallest  $r$  such that  $A_r \geq 1$  which is the length of the shortest word in the defining relation. In general, a design of resolution  $R$  is one in which no  $p$  factor effect is confounded with any other effect containing less than  $R - p$  factors.

The maximum resolution design is the  $2^{k-p}$  design with the highest resolution. It is justified by the effect hierarchy principle that one is interested in choosing a design which confounds higher-order effects compared to a design which confounds lower-order effects. As a lower-resolution design has words with shorter length, which implies the confounding of lower order effects, it is preferable to choose the maximum resolution design. Note that the definition of  $W$  starts with  $A_3$  as a design with nonzero values for  $A_1$  and  $A_2$  is undesirable as main effects cannot be confounded with each other.

In our earlier example, the  $2^{4-1}$  fractional design with the defining relation  $\mathbf{I} = \mathbf{1234}$  has resolution IV. The  $2^{4-1}$  fractional design with the defining relation  $\mathbf{I} = \mathbf{124}$  has resolution III. Therefore, the design with the defining relation  $\mathbf{I} = \mathbf{1234}$  is preferred as it is the maximum resolution design.

*Minimum Aberration Designs:* For a  $2^{4-1}$  design, choice of the best design can be made based on resolution alone. However, resolution is not always enough to select the best design. Consider two  $2^{7-2}$  designs  $d_1 : \mathbf{I} = \mathbf{4567} = \mathbf{12346} = \mathbf{12357}$  and  $d_2 : \mathbf{I} = \mathbf{1236} = \mathbf{1457} = \mathbf{234567}$ . The word  $\mathbf{12357}$  is simply obtained by multiplying the two generators  $\mathbf{4567}$  and  $\mathbf{12346}$  of  $d_1$ . The defining relation of  $d_2$  is obtained by a similar mechanism. The wordlength pattern  $W(d_1) = (4^1, 5^2, 6^0, 7^0)$  is different from  $W(d_2) = (4^2, 5^0, 6^1, 7^0)$  although they both have resolution IV. Since  $d_1$  has one word of length 4, it has three confounded pairs of two-factor interactions ( $\mathbf{45} = \mathbf{67}$ ,

46 = 57, 47 = 56). In contrast,  $d_2$  has six confounded pairs of two-factor interactions as it has two words of length 4 (12 = 36, 13 = 26, 16 = 23, 14 = 57, 15 = 47, 17 = 45). Thus,  $d_1$  is the design which minimizes the number of minimum-length words in the defining relation. It is called a *minimum aberration design*.

The resolution criterion defined earlier considers the lengths of the shortest words in the defining relation. Judging based on resolution alone, the two designs  $d_1$  and  $d_2$  in the example above are equivalent, because they both have the maximum resolution  $R_{\max}$ . The minimum aberration criterion, suggested by Fries and Hunter (1980), searches for a design with the minimum number of words of length  $R_{\max}$ . Therefore, the concept of aberration is a natural extension of the concept of resolution.

More formally, suppose two  $2^{k-p}$  designs  $d_g$  and  $d_h$  are to be compared. Let  $r$  be the smallest integer such that  $A_r(d_g) \neq A_r(d_h)$ . Design  $d_g$  is said to have less aberration if  $A_r(d_g) < A_r(d_h)$ . If there is no design with less aberration than  $d_g$ , then  $d_g$  is called the minimum aberration design (Wu and Hamada, 2011). For a given pair of  $k$  and  $p$ , a minimum aberration design always exists. The minimum aberration criterion can be used to rank any two designs. Like the maximum resolution criterion, it can be justified by the effect hierarchy principle.

To construct minimum aberration designs, the most intuitive approach is to write down all possible sets of  $p$  generators, the resulting wordlength patterns, and choose the set of generators that yields the minimum aberration design. However, this approach is not practical for large  $k$  and  $p$ . More sophisticated algorithms are needed for constructing the minimum aberration designs for such large prob-

lems. Wu and Hamada (2011) tabulates some minimum aberration designs for practical use.

The concepts laid out above will be useful for us to introduce a new design for performing multi-platform, multivariate experiments. In particular, we will build upon two concepts defined above: maximum resolution and minimum aberration. The main point as it relates to the former is that the maximum resolution criterion maximizes the length of the shortest word. That is, it is desirable that we confound main effects with higher order effects than with lower order effects. Along the same lines, the minimum aberration criterion minimizes the number of shortest words in a defining relation of a factorial design. That is, it is desirable that the number of such confounds are as few as possible.

### 3 FORMULATION OF DESIGNS FOR MULTI-PLATFORM EXPERIMENTS

---

As stated earlier, a highly desirable characteristic of marketing activities online is that they are testable. Online retailers routinely test and improve the layout of their website to maximize profitability. Websites intended for the purpose of educating visitors evaluate user engagement on an ongoing basis and attempt to improve metrics that include page views and time spent per page. Multivariate testing is a popular method in the digital world for applications that include website and email optimization. However, multivariate experiments in the online space present a new design challenge: experiments need to be conducted across multiple platforms that include desktops, laptops, tablets, and smartphones. Such multivariate experiments are important because a different set of attribute combinations may be optimal for different platforms.

Next, we propose a rigorous statistical design framework to address the multi-platform problem. As far as we are aware, this is the first systematic statistical work in this direction. Our basic idea is that the statistical design in the multi-platform context should permit an assessment of factor effects for each platform and the interaction effect between a factor and the platforms. We expect to incorporate the following order of effects in the multi-platform experiment we design. To begin, the response variable of interest (e.g. page views) is expected to vary by platform. Accurate assessment of difference in page views by platform is an important first step before we assess which factors are significant and for which platform. It is necessary to control for between-platform difference in page views to be able to accurately assess factor effects. This is easily seen in a context where certain factors are only significant for a smaller platform that gets lower web traffic. In such a

context, failure to accurately assess between-platform differences in page views could very easily mask such factor effects for the smaller platform. Recent industry reports (e.g. Adobe Digital Index) further emphasize this point by documenting the substantial gap in page views by platform type. One such report (Chaffey, 2016) notes that a majority of site visits across a wide variety of industries are from desktops thus making it quite clear that the between-platform difference is expected to be large and significant. Our empirical application reported later further illustrates this point.

After controlling for the between-platform difference effect, the effects associated with the design factors are the central focus of the multi-platform experiment. As only some of the design factors may be significant, accurate assessment of the magnitude of the significant design factors is important. Finally, the effects of the design factors likely vary by platform. The design needs to accurately assess such platform-by-design factor interactions. Among all the effects that an experiment can uncover, our focus is on the following effects and in this order: the between-platform difference effect, the design factor effects and the interaction between the two.

To facilitate development of the ideas we propose, we formally define a multi-platform experiment as follows.

**Definition 3.1 (Multi-platform Experiment).** *Consider a multi-platform experiment for studying  $k$  two-level design factors, denoted by  $1, \dots, k$ , on  $s$  platforms  $P_1, \dots, P_s$ . The complete design set  $d$  of the experiment consists of  $s$  sub designs,  $d_1, \dots, d_s$ , with  $d_j$  associated with  $P_j$ . To quantify the difference among the platforms, let  $S$  denote a categorical*

factor, called the slice factor, with  $s$  levels. The  $j$ th level of  $S$  is associated with  $P_j$ .

Figure 3.1 visually displays the design set  $d$  of the experiment in Definition 3.1. To construct suitable designs for the experiment in Definition 3.1, we introduce two guiding properties:

*Property 1.* For  $j = 1, \dots, s$ , the sub design  $d_j$  should achieve desirable estimation capacity for the design factors on platform  $P_j$ .

*Property 2.* Combined together, the complete design  $d$  should achieve desirable estimation capacity for the slice factor  $S$  and the two-way interactions between  $S$  and the design factors.

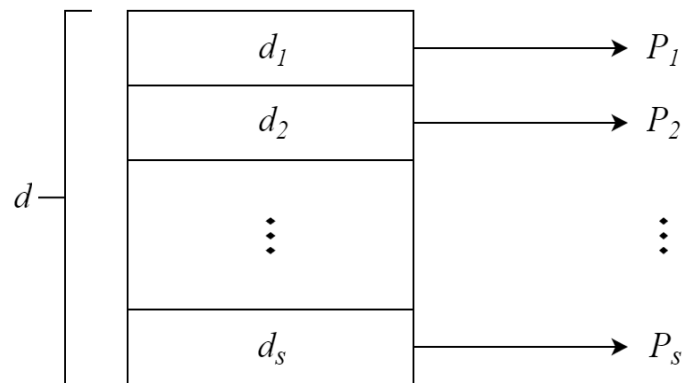


Figure 3.1: Design Set  $d$  of the Experiment in Definition 3.1.

In a standard factorial design, the effect hierarchy principle implies that a lower-order effect is more likely to be important than a higher-order effect, and that all effects of the same order are equally likely to be important. Therefore, to construct the sub design  $d_j$ , it is desirable to focus on main effects of the design factors. This

follows Property 1. Further, the focus of the complete design  $d$  should be on the main effect of the slice factor  $S$  and two-way interactions between  $S$  and the design factors. This follows Property 2.

The set of effects in a multi-platform experiment are expected to follow an order that we formalize next. We formally state the *sliced effect hierarchy principle* for  $d$  to incorporate the slice factor  $S$  in the design construction process. To do so, we divide the effects of  $d$  into two disjoint sets.

For the design  $d$  in Definition 3.1, let  $E_I$  be the set of all effects with words that exclude the slice factor  $S$  and  $E_S$  be the set of all effects with words that include the slice factor  $S$ . Using this notation, we define the sliced effect hierarchy principle as follows:

**Sliced Effect Hierarchy Principle:**

- (i) For  $E_I$  or  $E_S$ , the lower-order effects are more likely to be important than the higher-order effects.
- (ii) For  $E_I$  or  $E_S$ , effects of the same order are equally likely to be important.
- (iii) Any effect in the set  $E_S$  is likely to be more important than an effect in  $E_I$  that is of the same order.
- (iv) Any effect in the set  $E_S$  is likely to be less important than an effect in  $E_I$  that is of a lower order.

With regards to the difference between the slice factor and design factors in a multi-platform experiment, the slice factor is distinct from the design factors in two

major perspectives. First, a multi-platform experiment aims to detect what level of the design factors should be chosen to improve the layout of the design for each platform, and is not trying to select between platforms. Second, according to the sliced effect hierarchy principle for a multi-platform experiment, the importance of the effects related to the slice factor is higher than the importance of same-order effects of the design factors. A design set of a multi-platform experiment should be able to distinguish between the slice factor effects and the effects of design factors. Therefore, treating the slice factor as another design factor not only contaminates the experiment by selecting between platforms instead of improving for all of them, but also hinders the distinction of the slice factor effects and the effects of the design factors.

### 3.1 Illustration

Next we use a simple example to illustrate the sliced effect hierarchy principle and how it differs from the effect hierarchy principle defined in Section 2.

*Example 1.* For the experiment in Definition 3.1, let  $k = 3$  and  $s = 2$  for two platforms  $P_1$  and  $P_2$ . The slice factor  $S$  with two levels,  $-$  and  $+$ , represents  $P_1$  and  $P_2$ , respectively. Sub designs  $d_1$  and  $d_2$  are factorial designs with three factors such that each includes seven factorial effects that are ranked in Table 3.1 following the effect hierarchy principle.

The complete design  $d$  is a factorial design with three design factors and the slice factor  $S$ , and includes fifteen factorial effects (see Table 3.2). The two sets  $E_1$

Table 3.1: Effect Hierarchy Principle for Each Sub Design of the Experiment in Example 1

Rank	Effects
(i)	<b>1, 2, 3</b>
(ii)	<b>12, 13, 23</b>
(iii)	<b>123</b>

Table 3.2: Sliced Effect Hierarchy Principle for the Complete Design of the Experiment in Example 1

Rank	$E_S$	$E_I$
(i)	<b>S</b>	
(ii)		<b>1, 2, 3</b>
(iii)	<b>1S, 2S, 3S</b>	
(iv)		<b>12, 13, 23</b>
(v)	<b>12S, 13S, 23S</b>	
(vi)		<b>123</b>
(vii)	<b>123S</b>	

and  $E_S$  are  $\{1, 2, 3, 12, 13, 23, 123\}$  and  $\{S, 1S, 2S, 3S, 12S, 13S, 23S, 123S\}$ , respectively. Following the sliced effect hierarchy principle defined above, Table 3.2 ranks all fifteen factorial effects.

Example 1 illustrates that the sliced effect hierarchy principle assumes a certain order of effects for the design. As we will show next, this principle guides the design that is best suited for a multi-platform experiment.

## 3.2 Sliced Effect Hierarchy and Existing Design

### Methods

The sliced effect hierarchy is a new principle and existing methods cannot be used to construct designs that conform with it. We illustrate this point by discussing how some design methods that appear to be reasonable choices fail to work in the multi-platform contexts. First, consider a random splitting approach that would begin by generating a factorial design for  $d$  and then randomly split it into sub designs for  $d_j$ 's. Although simple to implement, such an approach cannot guarantee that each  $d_j$  follows Property 1. Second, another method could be to independently generate all sub designs  $d_j$ 's and then put them together to form  $d$ . Unfortunately, such an approach cannot guarantee that  $d$  has a desirable structure according to Property 2. Third, blocking is a standard method to form blocks of homogeneous units in a factorial design. While this method works well for agriculture and engineering applications where treatment-blocking interaction is negligible (Wu and Hamada, 2011), it is ill-suited for our new design problem. If one uses the slice factor  $S$  as a block factor to construct a blocked factorial design  $d$  with blocks  $d_1, \dots, d_s$ , then  $S$  would be confounded with the higher-order interaction effects of the design factors. This assumes that the slice factor  $S$  has a negligible interaction with the design factors. This assumption contradicts Property 2 and is counter to the primary goal of how design factor effects may interact with platforms.

In view of the drawbacks of the aforementioned methods, next we explain how to construct new designs for the experiment in Definition 3.1. Our basic idea is that

each sub design  $d_j$  should follow the effect hierarchy principle and the complete design  $d$  should follow the sliced effect hierarchy principle defined above. We construct the designs by extending resolution and aberration based criteria and call them *sliced factorial designs*.

## 4 SLICED FACTORIAL DESIGNS FOR TWO-PLATFORM EXPERIMENTS

---

In this section, we propose a method for constructing the sliced factorial designs for the experiment in Definition 3.1 with two platforms. We first extend the maximum resolution and minimum aberration criteria to accommodate the slice factor. Assume that the complete design  $d$  is a  $2^{k+1-p}$  fractional factorial design that consists of two sub designs,  $d_1$  and  $d_2$ , each of which is a  $2^{k-p}$  fractional factorial design. The maximum resolution and minimum aberration criteria, as defined in Section 2 for standard fractional factorial designs, can be directly used to judge the statistical quality of each sub design. The resolution-based criterion maximizes the length of the shortest word and the aberration-based criterion minimizes the number of shortest words in a defining relation of a factorial design. For the complete design  $d$ , we need to modify these criteria to accommodate the sliced effect hierarchy principle. To accomplish this, we first provide some useful notation and definitions.

**Definition 4.1 (Sliced Wordlength Pattern).** *For a  $2^{k+1-p}$  complete design  $d$  in Definition 3.1 with  $s = 2$ ,*

1. *The sliced defining relation is obtained by multiplying the defining relation of  $d$  by the slice factor  $S$ .*
2. *A sliced word is any word in the sliced defining relation except for the slice factor  $S$ .*
3. *The sliced wordlength pattern is the vector*

$$SW = (3^{B_3}, \dots, (k+1)^{B_{k+1}}) \quad (4.1)$$

where  $B_j$  is the number of sliced words of length  $j$  and  $k$  is the number of design factors.

We now revisit Example 1 to illustrate Definition 4.1.

*Example 1 (Continued).* Building upon this example that we set up earlier, assume that the complete design is a  $2^{4-1}$  fractional factorial design that consists of two sub designs, each of which is a  $2^{3-1}$  fractional factorial design. The design  $d^{(1)}: \mathbf{I} = \mathbf{123S}$  is a minimum aberration design with four factors that can be used as the complete design of this multi-platform experiment. As a result of  $d^{(1)}$ , the sub designs on  $P_1$  and  $P_2$  are  $d_1^{(1)}: \mathbf{I} = -\mathbf{123}$  and  $d_2^{(1)}: \mathbf{I} = \mathbf{123}$ , respectively. Both  $d_1^{(1)}$  and  $d_2^{(1)}$  are minimum aberration designs with three factors. For the complete design  $d^{(1)}$ , the sliced defining relation is  $\mathbf{S} = \mathbf{123}$  which is obtained by multiplying  $\mathbf{I} = \mathbf{123S}$  by the slice factor  $S$ . The design  $d^{(1)}$  has one sliced word  $\mathbf{123}$ , which is of length three. The sliced wordlength pattern is then  $SW(d^{(1)}) = (3^1)$ .

Following the sliced effect hierarchy principle, we now extend resolution and aberration to accommodate Property 1 and Property 2. We use the sliced wordlength pattern to define *sliced resolution* as follows:

**Sliced Resolution:** The sliced resolution of a  $2^{k+1-p}$  complete design  $d$  in Definition 3.1 with  $s = 2$  is defined to be the smallest  $sr$  such that  $B_{s,r} \geq 1$ , i.e., the length of the shortest sliced word in the sliced defining relation.

Following the sliced effect hierarchy principle, one is interested in maximizing the sliced resolution. Maximizing the sliced resolution ensures that the slice factor

$S$  and its interaction with the design factors can be best estimated. Further, for situations where the complete design  $d$  cannot be judged by its sliced resolution alone, we extend the minimum aberration criterion to a new concept called the *sliced minimum aberration*:

**Sliced Minimum Aberration Designs:** Suppose that, for the experiment in Definition 3.1 with  $s = 2$ , two  $2^{k+1-p}$  complete designs  $d^{(1)}$  and  $d^{(2)}$  are to be compared. Let  $sr$  be the smallest integer such that  $B_{sr}(d^{(1)}) \neq B_{sr}(d^{(2)})$ . Design  $d^{(1)}$  is said to have less sliced aberration if  $B_{sr}(d^{(1)}) < B_{sr}(d^{(2)})$ . If there is no design with less sliced aberration than  $d^{(1)}$ , then  $d^{(1)}$  is called a sliced minimum aberration design.

For a given pair of  $k$  and  $p$ , a sliced minimum aberration design always exists. Built upon the sliced effect hierarchy principle, the sliced minimum aberration criterion allows any two complete designs to be ranked.

Having defined a suitable design criterion for a multi-platform context, we are now ready to construct sliced minimum aberration designs. Lemma 1 below establishes a connection between the sliced minimum aberration criterion and the minimum aberration criterion. Our intention here is to guide the construction of the sliced minimum aberration designs using readily available minimum aberration designs of fewer number of factors.

**Lemma 1.** *A  $2^{k+1-p}$  sliced minimum aberration design corresponds to a sliced defining relation in which all sliced words contain  $S$ .*

**Proof.** A sliced defining relation with all sliced words containing  $S$  means the slice factor  $S$  is not involved in any word of the defining relation for the sliced minimum

aberration design. Therefore, no generator uses the slice factor  $S$  to generate the sliced minimum aberration design. For fixed  $k$  and  $p$ , it is sufficient to show that any sliced defining relation with at least a sliced word excluding  $S$  is inferior to a sliced defining relation that includes  $S$  in all its sliced words. Any sliced factorial design with a sliced word excluding  $S$  has  $S$  involved in the defining relation of the sliced factorial design, and hence  $S$  is used in the generators to generate the sliced factorial design. We prove for the case where one generator uses  $S$ . The proof can be easily generalized to the case where more than one generator uses  $S$ . Consider a  $2^{k+1-p}$  sliced factorial design  $d$  that has  $p - 1$  generators not involving  $S$  and one generator  $g$  involving  $S$ . It is sufficient to show that a design with all generators excluding  $S$  is better according to the sliced aberration criterion. Form a new design  $d_{new}$  by removing  $S$  from  $g$ . Call the new generator  $g_{new}$ . As  $S$  only appears in  $g$ , the product of  $g_{new}$  with other generators will result in a word excluding  $S$ . Comparing the sliced defining relation of  $d$  with  $d_{new}$ 's shows that the length of all the sliced words formed by  $g_{new}$  has increased by one, and the lengths of all other sliced words remain the same. Therefore,  $d_{new}$  is better according to the sliced aberration criterion.

□

We now revisit Example 1 to illustrate Lemma 1.

*Example 1 (Continued).* Let us revisit the design  $d^{(1)}$  that we discussed earlier. We note that  $d^{(1)}$  with the sliced defining relation  $\mathbf{S} = \mathbf{123}$  is not a sliced minimum aberration design because the sliced word  $\mathbf{123}$  does not contain  $S$ . The only generator of  $d^{(1)}$ , as defined in Section 2, is  $\mathbf{123S}$  which uses  $S$ . Following the proof of Lemma 1,

removing  $S$  from this generator results in a sliced defining relation of  $\mathbf{S} = \mathbf{123S}$  that is superior to the sliced defining relation of  $d^{(1)}$ .

To find a sliced minimum aberration design, it is sufficient to search among possible complete designs for which the sliced defining relations contain  $S$  in all their sliced words. As all sliced words contain  $S$ , removing  $S$  from the sliced defining relation of such designs result in a defining relation with all words excluding  $S$ . Therefore, minimizing the number of shortest sliced words in the sliced defining relation of a  $2^{k+1-p}$  complete design  $d$  is equivalent to minimizing the number of shortest words in the defining relation of a  $2^{k-p}$  fractional design consisting of design factors only. This leads to Theorem 1.

**Theorem 1.** *The sliced defining relation of a  $2^{k+1-p}$  sliced minimum aberration design can be generated by multiplying the slice factor  $S$  to the defining relation of a  $2^{k-p}$  minimum aberration design consisting of the design factors only.*

An extremely useful outcome of Theorem 1 is that it provides an easy way to construct sliced minimum aberration designs by using existing minimum aberration designs of fewer number of factors. To construct  $2^{k+1-p}$  sliced minimum aberration designs by using this theorem, one can begin by generating a  $2^{k-p}$  minimum aberration design consisting of the design factors for one platform and then repeat the same design for the other platform. To illustrate this point, we now use Theorem 1 to construct a sliced minimum aberration design for Example 1.

*Example 1 (Continued).* The design  $\mathbf{I} = \mathbf{123}$  is a  $2^{3-1}$  minimum aberration design consisting of design factors that can be used for both platforms. Multiplying  $\mathbf{I} = \mathbf{123}$

by  $S$  provides the sliced defining relation  $\mathbf{S} = \mathbf{123S}$  of the sliced minimum aberration design  $d^{(2)}$ . As  $d^{(2)}$  has the defining relation  $\mathbf{I} = \mathbf{123}$ , it is not a  $2^{4-1}$  minimum aberration design consisting of all four factors. Design  $d^{(2)}$  has one sliced word  $\mathbf{123S}$  of length four. The sliced wordlength pattern is then  $SW(d^{(2)}) = (4^1)$ . To illustrate that  $d^{(2)}$  is a better design than  $d^{(1)}$ , we compare their confounded effects involving the slice factor  $S$ . These confounded effects for  $d^{(1)}$  are  $\mathbf{S} = \mathbf{123}$ ,  $\mathbf{1S} = \mathbf{23}$ ,  $\mathbf{2S} = \mathbf{13}$ , and  $\mathbf{3S} = \mathbf{12}$ . In contrast, the confounded effects of  $d^{(2)}$  are  $\mathbf{S} = \mathbf{123S}$ ,  $\mathbf{1S} = \mathbf{23S}$ ,  $\mathbf{2S} = \mathbf{13S}$ , and  $\mathbf{3S} = \mathbf{12S}$ . This provides more estimation capacity for the slice factor  $S$  because, according to sliced effect hierarchy principle,  $\mathbf{123S}$  is less likely to be important than  $\mathbf{123}$ . Along the same lines,  $d^{(2)}$  confounds  $\mathbf{1S}$  with  $\mathbf{23S}$  and this provides more estimation capacity than confounding  $\mathbf{1S}$  with  $\mathbf{23}$ . Using the same argument, it is more desirable to confound  $\mathbf{2S}$  with  $\mathbf{13S}$  and  $\mathbf{3S}$  with  $\mathbf{12S}$ .

## 5 SLICED FACTORIAL DESIGNS FOR FOUR-PLATFORM EXPERIMENTS

---

First, we briefly describe the construction of a full factorial design set  $d$  for the experiment in Definition 3.1 with  $s = 4$  using the method of replacement. Consider a saturated  $2^{N-1}$  design with  $N = 2^l$  runs where  $l = k + 2$ . We can represent the  $N - 1$  columns of this saturated design by  $l$  independent columns denoted by  $1, \dots, l$  and their interactions of order 2 to  $l$ , that is  $12, 13, \dots, 12 \dots l$  (Wu and Zhang, 1993). Any three columns of the form  $(a, b, ab)$ , where  $ab$  is the interaction column between columns  $a$  and  $b$ , can be replaced by the 4-level column representing the slice factor  $S$  without affecting orthogonality (Addelman, 1963). This replacement can be done according to the rule shown in Table 5.1.

Table 5.1: Rule for Replacing any Three Columns of the Form  $(a, b, ab)$  by the 4-level Column  $S$

a	b	ab		4-level column S
0	0	0		0
0	1	1	→	1
1	0	1		2
1	1	0		3

Next, we describe the construction of a design set  $d$  with  $2^{k+2-p}$  runs for the experiment in Definition 3.1 with  $s = 4$ . Consider a full factorial design with  $2^{k+2}$  runs, whose 4-level column is represented by  $S = (s_1, s_2, s_3)$ , with  $s_3 = s_1 s_2$ , and 2-level columns are represented by  $1, \dots, k$ . For a two-platform experiment, in Section 4, we defined the sliced wordlength pattern by extending the wordlength pattern definition to the aliasing relation of the slice factor  $S$ . For a four-platform experiment, however, the same definition does not work because there are three aliasing relations for the slice factor  $S$ : aliasing relations of  $s_1$ ,  $s_2$ , and  $s_3$ . The

aliasing relation of  $s_j$  is obtained by multiplying the defining relation of  $d$  by  $s_j$ . Therefore, a word  $\omega$  in the defining relation of  $d$  appears in three aliasing relations for the slice factor  $S$  as  $s_1\omega$ ,  $s_2\omega$ , and  $s_3\omega$ . We extend the definition of the sliced wordlength pattern to be defined over the minimum length of  $s_1\omega$ ,  $s_2\omega$ , and  $s_3\omega$ . This extension follows the main idea of the sliced minimum aberration criterion, defined in Section 4, which is minimizing the number of shortest length of a sliced wordlength pattern. Defining the sliced wordlength pattern over the minimum length of  $s_1\omega$ ,  $s_2\omega$ , and  $s_3\omega$  is to make sure that the sliced minimum aberration is taking care of the worst case (Minimax principle).

We use Wu and Zhang (1993)'s definition of wordlength pattern for designs with two-level and four-level factors to define the sliced wordlength pattern. For the design set  $d$  with  $2^{k+2-p}$  runs, there are two types of words in its defining relation. The first involves only the design factors  $1, \dots, k$ , which is called type 0, and the second involves one of the  $s_j$ 's and some of the design factors  $1, \dots, k$ , which is called type 1. Because of the relation  $s_1s_2s_3 = I$ , any two  $s_j$ 's that appear in a word can be replaced by the third  $s_j$ . Therefore, these two types considers all possible combinations. Following Wu and Zhang (1993), the vector

$$W(d) = ([A_{i0}(d), A_{i1}(d)])_{i \geq 3}, \quad (5.1)$$

is the wordlength pattern of  $d$  in which  $A_{i0}(d)$  and  $A_{i1}(d)$  respectively are the number of type 0 and type 1 words of length  $i$  in the defining relation of  $d$ . The term  $[A_{20}(d), A_{21}(d)]$  is not considered in Equation 5.1 because any design  $d$  with positive  $[A_{20}(d), A_{21}(d)]$  is not useful as two of its main effects are confounded. We

use the concepts laid out in Wu and Zhang (1993) to formally define the sliced wordlength pattern of a design set  $d$  for a four-platform experiment as follows:

**Definition 5.1 (Sliced Wordlength Pattern).** For a design set  $d$  with the wordlength pattern  $W(d) = ([A_{i0}(d), A_{i1}(d)])_{i \geq 3}$  for the experiment in Definition 3.1 with  $s = 4$ , the sliced wordlength pattern is the vector  $SW(d) = ([B_{i0}(d), B_{i1}(d)])_{i \geq 2}$ , where

$$- B_{i0}(d) = A_{(i+1)1}(d) \text{ where } i \geq 2$$

$$- B_{i1}(d) = A_{(i-1)0}(d) \text{ where } i \geq 4$$

$$- B_{i1}(d) = 0 \text{ where } i = 2, 3,$$

i.e.  $SW = ([A_{31}, 0]_2, [A_{41}, 0]_3, [A_{51}, A_{30}]_4, [A_{61}, A_{40}]_5, \dots)$ .

A type 0 word in the defining relation of  $d$  appears as a type 1 word in the aliasing relations of the sliced factor  $S$ . Hence, it is counted as a type 1 word in the sliced wordlength pattern resulting in  $B_{i1}(d) = A_{(i-1)0}(d)$ . On the other hand, a type 1 word in the defining relation of  $d$  appears as a type 1 word in the aliasing relations of two  $s_j$ 's and as a type 0 word in the aliasing relation of the third  $s_j$ . It is counted as a type 0 word in the sliced wordlength pattern resulting in  $B_{i0}(d) = A_{(i+1)1}(d)$  because the sliced wordlength pattern is defined over the minimum length of a word in the three aliasing relations.

The sliced resolution of  $d$  is defined to be the smallest  $i$  for which at least one of  $B_{i0}(d)$  and  $B_{i1}(d)$  is positive. Additional discrimination among designs with the same sliced resolution is made by sliced minimum aberration that we define next. To define the sliced minimum aberration, the two types of words of the design set

$d$  are not treated the same. According to the sliced effect hierarchy, a type 1 word in the aliasing relations of the slice factor  $S$  is more serious because it involves one  $s_j$ . This is consistent with Wu and Zhang (1993) justification that considers a type 0 word in the defining relation of  $d$  to be more important than a type 1 because a type 0 word in the defining relation appears as a type 1 word in the aliasing relations of the slice factor  $S$ . Therefore, it is usually more important to require a smaller  $B_{i1}(d)$  than a smaller  $B_{i0}(d)$  for the same  $i$ . We thus define sliced minimum aberration designs for a four-platform experiment as follows:

**Sliced Minimum Aberration Designs:** Suppose that, for the experiment in Definition 3.1 with  $s = 4$ , two design sets  $d^{(1)}$  and  $d^{(2)}$  with  $2^{k+2-p}$  runs are to be compared. Let  $r$  be the smallest integer such that  $[B_{r0}(d^{(1)}), B_{r1}(d^{(1)})] \neq [B_{r0}(d^{(2)}), B_{r1}(d^{(2)})]$ . If  $B_{r1}(d^{(1)}) < B_{r1}(d^{(2)})$ , or  $B_{r1}(d^{(1)}) = B_{r1}(d^{(2)})$  but  $B_{r0}(d^{(1)}) < B_{r0}(d^{(2)})$ , then  $d^{(1)}$  is said to have less sliced aberration than  $d^{(2)}$ . If there is no design with less sliced aberration than  $d^{(1)}$ , then  $d^{(1)}$  is called a sliced minimum aberration design.

To illustrate the above two definitions, we consider the following example:

*Example 2.* For the experiment in Definition 3.1, let  $k = 5$  and  $s = 4$  for four platforms  $P_1, P_2, P_3$  and  $P_4$ . Also, let  $s_1, s_2, 1, 2$ , and  $3$  be five independent columns of the 32-run  $2^{31}$  design. Consider the following two designs:

$$d^{(1)} : S, 1, 2, 3, 12, 13$$

$$d^{(2)} : S, 1, 2, 3, 123s_1, 23s_2,$$

where  $S = (s_1, s_2, s_1s_2)$ , and the last two columns represent the last two design factors. For example, in  $d^{(1)}$ , we have  $4 = 12$  and  $5 = 13$  although they are  $4 = 123s_1$  and  $5 = 23s_2$  in  $d^{(2)}$ . Therefore, the defining relations of  $d^{(1)}$  and  $d^{(2)}$  are as follows:

$$d^{(1)} : \mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{2345}$$

$$d^{(2)} : \mathbf{I} = \mathbf{1234s_1} = \mathbf{235s_2} = \mathbf{145s_3}.$$

The defining relation of  $d^{(1)}$  has three words of type 0: two words of length three and one word of length four. Therefore, the wordlength pattern of  $d^{(1)}$  is  $W(d^{(1)}) = ([2, 0]_3, [1, 0]_4)$ . Multiplying the defining relation of  $d^{(1)}$  by  $s_j$ 's provides three aliasing relations for the slice factor  $S$  as follows:

$$s_1 = \mathbf{124s_1} = \mathbf{135s_1} = \mathbf{2345s_1}$$

$$s_2 = \mathbf{124s_2} = \mathbf{135s_2} = \mathbf{2345s_2}$$

$$s_3 = \mathbf{124s_3} = \mathbf{135s_3} = \mathbf{2345s_3}.$$

Each type 0 word in the defining relation of  $d^{(1)}$  appears with the same length and as type 1 word in all three aliasing relations of  $s_j$ 's. Therefore, the sliced wordlength pattern of  $d^{(1)}$  is  $SW(d^{(1)}) = ([0, 0]_2, [0, 0]_3, [0, 2]_4, [0, 1]_5)$ . On the other hand, the defining relation of  $d^{(2)}$  has three words of type 1: two words of length four and one word of length five. Therefore, the wordlength pattern of  $d^{(2)}$  is  $W(d^{(2)}) = ([0, 0]_3, [0, 2]_4, [0, 1]_5)$ . Multiplying the defining relation of  $d^{(2)}$  by  $s_j$ 's

provides three aliasing relations for the slice factor  $S$  as follows:

$$s_1 = \underline{1234} = 235s_3 = 145s_2$$

$$s_2 = 1234s_3 = \underline{235} = 145s_1$$

$$s_3 = 1234s_2 = 235s_1 = \underline{145}.$$

The type 1 word  $1234s_1$  in the defining relation of  $d^{(2)}$  appears as type 0 word of length four in the aliasing relation of  $s_1$  (because  $s_1s_1 = I$ ) and as type 1 word of length five in the aliasing relations of  $s_2$  and  $s_3$  (because  $s_1s_2 = s_3$  and  $s_1s_3 = s_2$ ). It is hence recorded as length four of type 0 in the sliced wordlength pattern. Similar explanation can be used for the other two words in the defining relation of  $d^{(2)}$ . Therefore, the sliced wordlength pattern of  $d^{(2)}$  is  $SW(d^{(2)}) = ([0,0]_2, [2,0]_3, [1,0]_4, [0,0]_5)$ . To compare the two design sets  $d^{(1)}$  and  $d^{(2)}$ , note that 3 is the smallest integer such that  $[0,0]_3(d^{(1)}) \neq [2,0]_3(d^{(2)})$ . The design set  $d^{(1)}$  has less sliced aberration than  $d^{(2)}$  because  $B_{31}(d^{(1)}) = B_{31}(d^{(2)}) = 0$  and  $B_{30}(d^{(1)}) = 0 < 2 = B_{30}(d^{(2)})$ . Later, we will show that  $d^{(1)}$  is a sliced minimum aberration design with 32 runs. Note that  $d^{(2)}$  is a minimum aberration design with 32 runs according to Wu and Zhang (1993) which is inferior to a sliced minimum aberration design for a four-platform experiment.

Having defined a suitable design criterion for a four-platform experiment, we are now ready to construct sliced minimum aberration designs. Theorem 2 below guides the construction of the sliced minimum aberration designs using readily available minimum aberration designs of fewer number of factors.

**Theorem 2.** *A sliced minimum aberration design corresponds to a defining relation in which all words are type 0.*

**Proof.** It is sufficient to show that any defining relation with at least a type 1 word is inferior to a defining relation in which all words are type 0. Any design with a type 1 word has at least one  $s_j$  involved in the generators of the design. We prove for the case where one generator uses one  $s_j$ . The proof can be easily generalized to the case where more than one generator uses  $s_j$ 's. Consider a design set  $d$  with  $2^{k+2-p}$  runs that has  $p-1$  generators not involving  $s_j$ 's and one generator  $g$  involving  $s_j$ . It is sufficient to show that a design with all generators excluding  $s_j$ 's is better according to the sliced minimum aberration criterion. Form a new design set  $d_{new}$  by removing  $s_j$  from  $g$ . Call the new generator  $g_{new}$ . As  $s_j$  only appears in  $g$ , the product of  $g_{new}$  with other generators will result in a type 0 word in the defining relation of  $d$ . Therefore, the length of all type 0 words formed by  $g_{new}$  in  $d_{new}$  has decreased by one compared to the length of all type 1 words formed by  $g$  in  $d$ . As a result, all these words formed by  $g_{new}$  of  $d_{new}$  appear as type 1 words in defining relations of  $s_j$ 's and are recorded with higher length in the sliced wordlength pattern compared to the ones in  $d$ . The lengths of all other words of  $d_{new}$  not formed by  $g_{new}$  remain the same as the ones in  $d$  not formed by  $g$  in their sliced wordlength patterns. Therefore,  $d_{new}$  is better according to the sliced minimum aberration criterion.

□

As a result of Theorem 2, to find a sliced minimum aberration design, it is sufficient to search among possible design sets for which all the words are type 0

in the defining relations. Therefore, minimizing the number of shortest length in the sliced wordlength pattern of  $d$  with  $2^{k+2-p}$  runs is equivalent to minimizing the number of shortest length in the wordlength pattern of a  $2^{k-p}$  fractional design consisting of design factors only. For a four-platform experiment, we thus use Theorem 2 to provide sliced minimum aberration designs with 16, 32, and 64 runs in Tables 5.2, 5.3, and 5.4 receptively.

Table 5.2: Sliced Minimum Aberration Designs with 16 Runs,  $S = (s_1, s_2, s_1s_2)$

k	Design set d	$SW(d)_{i \geq 4}$
3	S, 1, 2, 12	$([0, 1]_4)$

Table 5.3: Sliced Minimum Aberration Designs with 32 Runs,  $S = (s_1, s_2, s_1s_2)$

k	Design set d	$SW(d)_{i \geq 4}$
4	S, 1, 2, 3, 123	$([0, 0]_4, [0, 1]_5)$
5	S, 1, 2, 3, 12, 13	$([0, 2]_4, [0, 1]_5)$
6	S, 1, 2, 3, 12, 13, 23	$([0, 4]_4, [0, 3]_5)$
7	S, 1, 2, 3, 12, 13, 23, 123	$([0, 7]_4, [0, 7]_5, [0, 0]_6, [0, 0]_7, [0, 1]_8)$

Table 5.4: Sliced Minimum Aberration Designs with 64 Runs,  $S = (s_1, s_2, s_1s_2)$

k	Design set d	$SW(d)_{i \geq 4}$
5	S, 1, 2, 3, 4, 1234	$([0, 0]_4, [0, 0]_5, [0, 1]_6)$
6	S, 1, 2, 3, 4, 123, 124	$([0, 0]_4, [0, 3]_5)$
7	S, 1, 2, 3, 4, 123, 124, 134	$([0, 0]_4, [0, 7]_5)$
8	S, 1, 2, 3, 4, 123, 124, 134, 234	$([0, 0]_4, [0, 14]_5, [0, 0]_6, [0, 0]_7, \dots)$
9	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234	$([0, 4]_4, [0, 14]_5, [0, 8]_6, [0, 0]_7, \dots)$
10	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34	$([0, 8]_4, [0, 18]_5, [0, 16]_6, [0, 8]_7, \dots)$
11	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34, 24	$([0, 12]_4, [0, 26]_5, [0, 28]_6, [0, 24]_7, \dots)$
12	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34, 24, 14	$([0, 16]_4, [0, 39]_5, [0, 48]_6, [0, 48]_7, \dots)$
13	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34, 24, 14, 23	$([0, 22]_4, [0, 55]_5, [0, 72]_6, [0, 96]_7, \dots)$
14	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34, 24, 14, 23, 13	$([0, 28]_4, [0, 77]_5, [0, 112]_6, [0, 168]_7, \dots)$
15	S, 1, 2, 3, 4, 123, 124, 134, 234, 1234, 34, 24, 14, 23, 13, 12	$([0, 33]_4, [0, 105]_5, [0, 168]_6, [0, 280]_7, \dots)$

## 6 TWO ALGORITHMS TO CONSTRUCT SLICED FACTORIAL DESIGNS

---

To construct a sliced minimum aberration design for the complete design  $d$  in Definition 3.1, we use the idea in Theorems 1 and 2 to provide a simple algorithm for the case where all combinations of the design factors are feasible within all platforms. Later, we propose a second algorithm for the more general situation where some combinations of the design factors may be infeasible for some platforms.

**Algorithm 1.**

*Step 1: For platform  $P_1$  of the experiment in Definition 3.1, find a minimum aberration design of the design factors and use it for all sub designs  $d_1, d_2, \dots, d_s$ .*

*Step 2: For  $j = 1, \dots, s$ , add the  $j$ th level of the slice factor to experimental versions of  $d_j$ .*

*Step 3: Combine all sub designs in Step 2 to obtain a sliced minimum aberration design  $d$ .*

Next we use an example involving a greater number of factors to illustrate Algorithm 1.

*Example 3.* For the experiment in Definition 3.1, let  $k = 8$  and  $s = 2$  for two platform  $P_1$  and  $P_2$ . The slice factor  $S$  is defined with two levels  $+$  and  $-$ , representing platforms  $P_1$  and  $P_2$ , respectively. Consider a  $2^{9-3}$  complete design  $d$  that consists of  $d_1$  and  $d_2$ , each being a  $2^{8-3}$  factorial design. Following Step 1 of Algorithm 1, three generators  $g_1$ : **13458**,  $g_2$ : **1247**, and  $g_3$ : **1236** provide the  $2^{8-3}$  minimum aberration design of the design factors (Wu and Hamada, 2011, p. 254). Using the same design

for both platforms results in the defining relation of  $\mathbf{I} = 1236 = 1247 = 3467 = 13458 = 24568 = 23578 = 15678$  for both  $d_1$  and  $d_2$ .

The words **3467**, **24568**, **23578**, and **15678** are simply obtained by the products  $g_2 \times g_3$ ,  $g_1 \times g_3$ ,  $g_1 \times g_2$ , and  $g_1 \times g_2 \times g_3$  respectively. The resulting experimental versions for both  $d_1$  and  $d_2$  are {8, 167, 2678, 12, 36, 1378, 237, 12368, 47, 1468, 246, 12478, 34678, 134, 2348, 123467, 5, 15678, 2567, 1258, 3568, 1357, 23578, 12356, 4578, 1456, 24568, 12457, 34567, 13458, 2345, 12345678}, where 24568, for instance, represents the version that has five factors 2, 4, 5, 6, and 8 at + and the other three factors 1, 3, and 7 at -. Adding S to all experimental versions of  $d_2$  and combining it with  $d_1$  provides the  $2^{9-3}$  sliced minimum aberration design with the sliced defining relation  $\mathbf{S} = 1236\mathbf{S} = 1247\mathbf{S} = 3467\mathbf{S} = 13458\mathbf{S} = 24568\mathbf{S} = 23578\mathbf{S} = 15678\mathbf{S}$  that has the sliced wordlength pattern  $(5^3, 6^4)$ .

For the experiment in Definition 3.1, for situations where some combinations of the design factors may be infeasible for some platforms, exact sliced minimum aberration designs cannot be guaranteed. We provide a second algorithm that searches all designs that are orthogonal and balanced in main effects until it reaches the best possible sliced factorial design. This algorithm assumes that  $s = 2$ . By applying Step 1 of Algorithm 1, there are  $2^p$  fractions to be selected for sub design  $d_1$ . It is assumed that all experimental versions of at least one of the fractions are feasible within each platform. We first introduce some useful notation.

For the experiment in Definition 3.1 with  $s = 2$ , let  $M$  be the set of all  $2^{k+1-p}$  complete designs  $d$ 's. We partition the set  $M$  into three disjoint sets,  $M_1$ ,  $M_2$ , and  $M_3$ , based on the defining relations of the sub designs  $d_j$ 's:

$M_1$ : The set of  $2^{k+1-p}$  complete designs  $d$ 's for which the defining relations of both sub designs  $d_j$ 's are the same.

$M_2$ : The set of  $2^{k+1-p}$  complete designs  $d$ 's for which the defining relations of both sub designs  $d_j$ 's have the same words but different use of  $+/-$ .

$M_3$ : The set of  $2^{k+1-p}$  complete designs  $d$ 's for which the defining relations of sub designs  $d_j$ 's have different words.

A sliced minimum aberration design generated by Algorithm 1 belongs to the set  $M_1$ . It is easy to show that any design from the set  $M_3$  is neither balanced nor orthogonal in main effects. Therefore, no designs from the set  $M_3$  is considered. For the case where some combinations of design factors may be infeasible and no sliced minimum aberration design from the set  $M_1$  is possible, one should search for a design from the set  $M_2$ . Definition 6.1 provides a design from the set  $M_2$ .

**Definition 6.1.** Consider the experiment in Definition 3.1 with  $s = 2$ , where the sub design  $d_1$  is a  $2^{k-p}$  factorial design with  $p$  generators  $g_1, \dots, g_p$ . Let  $g_{i_1}, \dots, g_{i_t}, t \in \{1, \dots, p\}$ , denote a sub selection from  $g_1, \dots, g_p$ . Then, we define

1. The generator  $-g_j$  is obtained by changing the  $+/-$  sign of the generator  $g_j, j = 1, \dots, p$ .
2. Among all the words in the group formed by all  $p$  generators  $g_1, \dots, g_p$ , the set  $\langle g_{i_1}, \dots, g_{i_t} \rangle$  represents the sub selection of words formed by  $g_{i_1}, \dots, g_{i_t}$ .

3. *Slicing of the complete design  $d$  is determined by  $g_{i_1}, \dots, g_{i_t}$  if the sub design  $d_2$  is a  $2^{k-p}$  factorial design with  $p$  generators consisting of  $-g_{i_1}, \dots, -g_{i_t}$  and the rest  $p - t$  generators being the same as  $d_1$ .*

We now revisit Example 3 to illustrate Definition 6.1.

*Example 3 (Continued).* Construct  $d_1$  using three generators  $g_1$ : **13458**,  $g_2$ : **1247**, and  $g_3$ : **1236**, and  $d_2$  using three generators  $-g_1$ : **-13458**,  $g_2$ : **1247**, and  $g_3$ : **1236**. As a result,  $d_1$  and  $d_2$  have defining relations  $\mathbf{I} = \mathbf{1236} = \mathbf{1247} = \mathbf{3467} = \mathbf{13458} = \mathbf{24568} = \mathbf{23578} = \mathbf{15678}$  and  $\mathbf{I} = \mathbf{1236} = \mathbf{1247} = \mathbf{3467} = \mathbf{-13458} = \mathbf{-24568} = \mathbf{-23578} = \mathbf{-15678}$ , respectively. By combining  $d_1$  and  $d_2$ , the sliced defining relation  $\mathbf{S} = \mathbf{1236S} = \mathbf{1247S} = \mathbf{3467S} = \mathbf{13458} = \mathbf{24568} = \mathbf{23578} = \mathbf{15678}$  is provided for  $d$  whose slicing is determined by  $g_1$  and has the sliced wordlength pattern  $(5^7)$ . The set of words formed by  $g_1$  is  $\langle g_1 \rangle = \{\mathbf{13458}, \mathbf{24568}, \mathbf{23578}, \mathbf{15678}\}$ . It is important to note that these words have different use of  $+/-$  in the defining relations of  $d_1$  and  $d_2$ . The goal is to have the words in this set as long as possible because this will result in increase of the sliced aberration of  $d$ . These four words change the sliced wordlength pattern  $(5^3, 6^4)$  of the sliced minimum aberration design to  $(5^7)$  for the design whose slicing is determined by  $g_1$ . The design whose slicing is determined by  $g_1$  belongs to set  $M_2$  as the defining relations of both its sub designs have the same words but with different use of  $+/-$ . Similarly, the design whose slicing is determined by  $g_1, g_2$  is constructed by using three generators  $g_1, g_2,$  and  $g_3$  for  $d_1$  and three generators  $-g_1, -g_2,$  and  $g_3$  for  $d_2$ . The set of words formed by  $g_1, g_2$  is  $\langle g_1, g_2 \rangle = \{\mathbf{1247}, \mathbf{3467}, \mathbf{13458}, \mathbf{24568}\}$ . The design whose slicing is determined by

$g_1, g_2$  belongs to set  $M_2$ .

Algorithm 2 constructs the complete design  $d$  for the experiment in Definition 3.1 with  $s = 2$  as follows:

**Algorithm 2.**

*Step 1: Find all  $p$  generators  $g_i$ 's of a  $2^{k-p}$  minimum aberration design of  $k$  design factors.*

*Step 2: Among all  $2^p$  fractions created by  $g_i$ 's, if there is a fraction for which all experimental versions are feasible for both platforms, choose it for both sub designs,  $d_1$  and  $d_2$ , to form a sliced minimum aberration design; otherwise go to Step 3.*

*Step 3: For all sub selections of generators from  $g_1, \dots, g_p$ , create all  $2^p$  sets  $\langle g_{i_1}, \dots, g_{i_t} \rangle$ 's as defined in Definition 6.1,  $t \in \{1, \dots, p\}$ .*

*Step 4: Rank all the sets formed in Step 3 with the highest rank being the set of longest words and the lowest rank being the set of shortest words.*

*Step 5: For the set with the highest rank, generate the design whose slicing is determined by the relevant generators.*

*Step 6: If sub designs of the design generated in Step 5 have all experimental versions feasible on the platforms, assign one sub design to each platform, add the associated level of the slice factor to the experimental versions, and combine them to form  $d$ ; otherwise remove this set from the ranked list of Step 4 and continue from Step 5.*

Figure 6.1 displays the flow chart of Algorithm 2. We now revisit Example 3 to illustrate Algorithm 2.

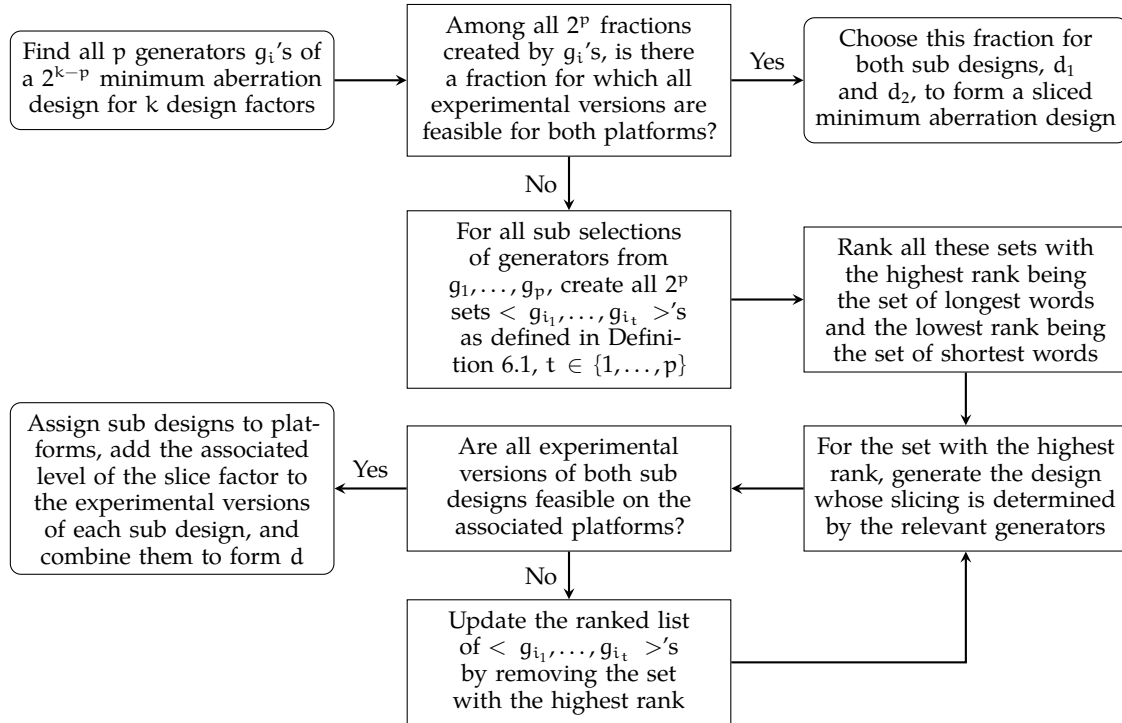


Figure 6.1: Flow Chart of Algorithm 2.

*Example 3 (Continued).* Let us assume that version 8 should be in sub design  $d_1$  and the combination 24568 cannot be used in the experimental versions of sub design  $d_2$ . Therefore, the sliced minimum aberration design cannot be obtained. The design whose slicing is determined by  $g_1$  provides the experimental versions {NULL, 1678, 267, 128, 368, 137, 2378, 1236, 478, 146, 2468, 1247, 3467, 1348, 234, 1234678, 58, 1567, 25678, 125, 356, 13578, 2357, 123568, 457, 14568, 2456, 124578, 345678, 1345, 23458, 1234567} for  $d_2$ , where NULL represents the version with all eight design factors at  $-$ . The design whose slicing is determined by  $g_3$  can also

be used that provides the experimental versions {68, 17, 278, 126, 3, 13678, 2367, 1238, 467, 148, 24, 124678, 3478, 1346, 23468, 12347, 56, 1578, 257, 12568, 358, 13567, 235678, 1235, 45678, 145, 2458, 124567, 3457, 134568, 23456, 1234578} for  $d_2$ . The design whose slicing is determined by  $g_2$  results in the experimental versions {87, 16, 268, 127, 367, 138, 23, 123678, 4, 14678, 2467, 1248, 3468, 1347, 23478, 12346, 57, 1568, 256, 12578, 35678, 135, 2358, 123567, 458, 14567, 245678, 1245, 3456, 134578, 23457, 1234568} for  $d_2$ . Therefore, the design whose slicing is determined by  $g_2$  cannot be used as the experimental version 245678 includes the combination 24568.

To compare the property of the design whose slicing is determined by  $g_1$  with the design whose slicing is determined by  $g_3$ , compare the set  $\langle g_1 \rangle = \{13458, 24568, 23578, 15678\}$  with  $\langle g_3 \rangle = \{1236, 3467, 24568, 15678\}$ . The set  $\langle g_3 \rangle$  has two words of length four which are shorter than all words in  $\langle g_1 \rangle$ . The set  $\langle g_3 \rangle$  changes the sliced wordlength pattern  $(5^3, 6^4)$  of the sliced minimum aberration design to  $(4^2, 5^3, 6^2)$  for the design whose slicing is determined by  $g_3$ . The design whose slicing is determined by  $g_1$  is hence better since its sliced wordlength pattern is  $(5^7)$ .

Table 6.1 includes all different designs whose slicing is determined by different combinations of three generators  $g_1$ ,  $g_2$ , and  $g_3$ . All eight sliced factorial designs considered in Table 6.1 are the designs that best satisfy Property 1 because all their sub designs are minimum aberration designs consisting of design factors only. However, these sliced factorial designs perform differently in terms of Property 2. Among all these sliced factorial designs that are orthogonal and balanced in main effects, the sliced minimum aberration design is the best.

Table 6.1: Rank of  $2^{9-3}$  Orthogonal Sliced Factorial Designs for Example 3 with Three Generators  $g_1$ : **13458**,  $g_2$ : **1247**, and  $g_3$ : **1236**

Rank	Sliced Factorial Design	The Set < determining generators >	Sliced Wordlength Pattern	Comment
1	Sliced minimum aberration	Empty set	$(5^3, 6^4)$	
2	Slicing determined by $g_1$	$\langle g_1 \rangle = \{13458, 24568, 23578, 15678\}$	$(5^7)$	
3	Slicing determined by $g_2$	$\langle g_2 \rangle = \{1247, 3467, 23578, 15678\}$	$(4^2, 5^3, 6^2)$	MA* design
3	Slicing determined by $g_1, g_2$	$\langle g_1, g_2 \rangle = \{1247, 3467, 13458, 24568\}$	$(4^2, 5^3, 6^2)$	MA design
3	Slicing determined by $g_3$	$\langle g_3 \rangle = \{1236, 3467, 24568, 15678\}$	$(4^2, 5^3, 6^2)$	MA design
3	Slicing determined by $g_1, g_3$	$\langle g_1, g_3 \rangle = \{1236, 3467, 13458, 23578\}$	$(4^2, 5^3, 6^2)$	MA design
3	Slicing determined by $g_2, g_3$	$\langle g_2, g_3 \rangle = \{1236, 1247, 24568, 23578\}$	$(4^2, 5^3, 6^2)$	MA design
3	Slicing determined by $g_1, g_2, g_3$	$\langle g_1, g_2, g_3 \rangle = \{1236, 1247, 13458, 15678\}$	$(4^2, 5^3, 6^2)$	MA design

\* MA: minimum aberration

## 7 EMPIRICAL APPLICATION

---

The purpose of this study was to improve the email design for a digital magazine, named Update, published by Wisconsin school of business. This magazine has both print and digital versions. The business school sends an email to its alumni base twice a year inviting them to read a variety of news in the magazine. The invitation email contains links to the landing page and multiple articles that relate to the business school. Version one in Table 7.2 shows the top half of the most recent issue of the magazine prior to this study.

For this study, the team managing the digital magazine identified six binary design factors for the multivariate test for two platforms  $P_1$  and  $P_2$ . Platform  $P_1$  refers to a tablet or cell phone and  $P_2$  refers to a desktop or laptop computer. We decided to combine the two devices for each platform in order to have enough observations/platform. For contexts where the expected sample size is large, this aggregation would be unnecessary. The slice factor  $S$  is defined as a binary for which the level  $-$  represents platform  $P_1$  and the level  $+$  represents platform  $P_2$ . For a full design we would have to create  $2^6$  versions for each platform  $P_1$  and  $P_2$ . Instead, using the design criteria developed in Section 4, we create  $2^3$  versions to perform the multivariate testing. The design we use is a  $2^{7-3}$  sliced factorial design. Table 7.1 includes six binary design factors that were identified for this study. These factors are 1: full width banner, 2: stories with minimal teaser, 3: class note stories, 4: cover image at top, 5: full width logo, and 6: call to action button. For each design factor, the  $+$  level is the change from the control version of magazine. The control version was identical to the most recently published magazine prior to this study (version one in Table 7.2).

Each additional version in a multivariate design has an associated incremental cost. In our empirical application, because of the extensive programming related resource constraints, the maximum number of versions was restricted to eight. In another multi-platform experiment context, such as website redesign, constraints include number of websites that one could realistically test without disrupting business. The eight versions in our setup included the control version, the format used in the most recent magazine prior to this study. We used Algorithm 1 to generate a sliced factorial design for this study. Following Step 1 of Algorithm 1, three generators **124**, **135**, and **236** provide the  $2^{6-3}$  minimum aberration design of the design factors (Wu and Hamada, 2011, p. 253). Using the same design for both platforms results in the defining relation of  $\mathbf{I} = \mathbf{124} = \mathbf{135} = \mathbf{236} = \mathbf{456} = \mathbf{2345} = \mathbf{1346} = \mathbf{1256}$  for both  $d_1$  and  $d_2$ . Both sub designs  $d_1$  and  $d_2$  are  $2^{6-3}$  minimum aberration designs with resolution III. The resulting experimental versions for both  $d_1$  and  $d_2$  are {NULL, 145, 246, 1256, 356, 1346, 2345, 123}, where NULL is the control version and 145, for instance, represents the version that has three factors 1, 4, and 5 at + and the other three factors 2, 3, and 6 at -. Adding S to all experimental versions of  $d_2$  and combining it with  $d_1$  provides the  $2^{7-3}$  sliced minimum aberration design with the sliced defining relation  $\mathbf{S} = \mathbf{124S} = \mathbf{135S} = \mathbf{236S} = \mathbf{456S} = \mathbf{2345S} = \mathbf{1346S} = \mathbf{1256S}$  that has the sliced wordlength pattern  $(4^4, 5^3)$ . Algorithm 1 generated a sliced minimum aberration design that required eight versions. In contrast, a blocked design would have required sixteen versions thus needing greater resources compared to our methodology.

The first version of our design is the control version which is partially presented

in Table 7.2. Version two has factors 1, 4, and 5 that are at + levels and the remaining three factors are at – levels. Similarly, version three has factors 2, 4, and 6 at + levels although the other three factors are at – levels. Version four has factors 1, 2, 5, and 6 at + levels and the remaining factors are at – levels. Version five has factors 3, 5, and 6 at + levels and the rest at – levels. Version six has factors 1, 3, 4, and 6 at + levels and the rest at – levels. Version seven has factors 2, 3, 4, and 5 at + levels and the rest at – levels. Finally for version eight, three factors 1, 2, and 3 are at + levels and the remaining factors are at – levels. Table 7.2 and Table 7.3 include the top part of all eight versions used for our empirical application. For versions five, six, seven, and eight class notes are displayed as a set of 3 example notes at the bottom.

The response variable for this study is the number of page views that are obtained from Google Analytics. Google Analytics records page views for each version of the study. These data are aggregated across users that are exposed to each version. How page views vary within a version is not known to us. This is very typical of Google Analytics data. Therefore, a popular method called Lenth's test (Lenth, 1989) is used to identify statistically significant factors for this study. The Lenth's test is specifically designed for testing effects in experiments for which variance estimates are not available. It is simple to use and performs well according to Hamada and Balakrishnan (1998) who report an extensive review of different methods one could use to analyze an unreplicated factorial experiment. The Lenth's test is also well studied by Ye and Hamada (2000).

To use Lenth's test, a robust estimator of the standard deviation of estimated

effects called the pseudo standard error (PSE) is considered (Wu and Hamada, 2011). To calculate PSE, an initial standard error is defined to be 1.5 times the median of the absolute value of estimated effects, where 1.5 is the scaling factor. This initial standard error is a consistent estimator of the standard deviation of estimated effects when the effects are zero and the underlying error distribution is normal. Lenth's method then trims the estimated effects by considering the ones within 2.5 times the initial standard error. PSE is defined to be 1.5 times the median of the absolute value of the trimmed estimated effects, and is a robust estimator of the standard deviation of estimated effects. The term "robust" here means that PSE's performance is not sensitive to the nonzero effects. A t-like statistic is then defined by dividing the estimated effects by PSE. This statistic is compared with critical values provided by Lenth (1989) to assess if an effect is significant. Table 7.4 includes some descriptive statistics for this study. The total number of recipients is 25,693 and they are divided into roughly equal eight sets, each set of recipients received a version of the email.

There are eight versions for each platform and as a result seven effects of design factors can be estimated per platform. Table 7.5 includes the confounded effects within each platform of our empirical study. These confounded effects are identified by the sliced factorial design we generated by using Algorithm 1. Such confounds are implicit in any fractional factorial design that one uses in marketing and we make them explicit to offer greater clarity. For convenience, we label each set of confounded effects.

In the sliced factorial design framework, slices are used for analyses on both  $P_1$  and  $P_2$ . Tables 7.6 and 7.7 include the effects of design factors that are estimated using sub designs  $d_1$  on platform  $P_1$  and  $d_2$  on platform  $P_2$ , respectively. Lenth's method is used to test the significance of effects and to report the p-values. Following Property 1, the analyses for each slice are used to estimate the effect of design factors within each platform.

Comparing the results of Tables 7.6 and 7.7 suggests that effect **C** is significant on  $P_2$  (desktops and laptops) although two effects **F** and **G** are significant on  $P_1$  (smartphone, tablet). Table 7.5 reveals that effect **C** is the sum of the following confounded effects **3, 15, 26, 245, 146, 1234, 3456, 12356**. As sub designs follow effect hierarchy principle, **C** can be viewed to represent effect **3** by assuming that all higher-order confounded effects are negligible. The main takeaway for  $P_2$  from Table 7.7 is that displaying class notes as a set of three (see factor 3 in Table 7.1) will improve the number of page views and that other factors are unlikely to have a positive effects on page views.

Looking at the results for  $P_1$  in Tables 7.6, the two significant effects are **F** and **G**. Table 7.5 shows that effect **F** is the sum of the following confounded effects **6, 23, 45, 134, 125, 1246, 1356, 23456**. As sub designs follow effect hierarchy principle, **F** can be viewed to represent effect **6** by assuming that all higher-order confounded effects are negligible. Further, effect **G** is the sum of the following confounded effects **16, 34, 25, 145, 246, 356, 123, 123456**. Once again, the effect hierarchy principle suggests that **G** represents sum of three effects **16, 34, and 25**. Next, by applying the effect heredity principle, **G** can be viewed to represent effect **16** as its parent **6**

is significant. The effects **34** and **25** can be safely assumed to be zero because their parent effects (**2**, **3**, **4**, **5**) are statistically insignificant. The effect **6** is positive and the effect **16** is negative. Therefore, the main takeaway for  $P_1$  is that adding the call to action button on top (factor 6 in Table 7.1) and refraining from the full width banner format (see factor 1 in Table 7.1) will improve number of page views. The remaining factors do not impact page views in any way so should not be considered for adoption the next digital magazine.

Following Property 2, to compare the results of two platforms, the complete design  $d$  is then used to estimate the slice factor and its interaction with the platform specific significant effects. Table 7.8 includes four effects **S**, **3S**, **6S**, and **16S**. The effect **S** is positive and significant implying that  $P_2$  brings significantly more page views compared to  $P_1$ . Further, the magnitude of the effect **S** is around three times larger than the effects of design factors. This is important to note because it is consistent with the sliced effect hierarchy principle. The effects **3S**, **6S**, and **16S** uncover the way the effects **3**, **6**, and **16** interact with the slice factor, respectively, meaning how these effects differentially affects page views from  $P_1$  to  $P_2$ . We note that these interactions are not statistically significant implying that the differential effects of **3**, **6**, and **16** on page views from  $P_1$  to  $P_2$  are not significant. As a result, one version that adopts the changes of displaying “class notes as a set of three”, adding the “call to action button”, and refraining from the “full width banner format” will improve number of page views for both platforms  $P_1$  and  $P_2$ .

Next, we calculate the expected incremental gain in page views because of the design changes the multi-platform multivariate test reveals. For  $P_2$ , average

pageview can be estimated using

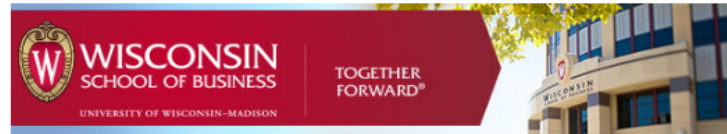
$$\text{Average Pageview} = 1.864 + 0.139C. \quad (7.1)$$

When compared to the control group, this suggests an expected 16% gain in page views when the “class notes story” ( $C = +$  vs.  $C = -$ ) is included in the magazine design. Similarly for  $P_1$ , average pageview can be estimated by using

$$\text{Average Pageview} = 1.193 + 0.019F - 0.022AF. \quad (7.2)$$

When compared to the control group, this provides a 7% gain in page views when the “call to action” button is added on the top and the “full width banner format” ( $F = +$  and  $A = -$  vs.  $F = -$  and  $A = -$ ) is excluded in the magazine design.

In sum, the sliced factorial design methodology suggests different set of designs factors would increase page views for  $P_2$  and  $P_1$ . Displaying the class notes story will improve pageview on  $P_2$  and adding the call to action button on top while avoiding the full width banner format will improve page views on  $P_1$ . As a result of this multivariate test, the next issue of the digital magazine (see Figure 7.1) adopted the following changes to increase page views: (i) display the class notes story, (ii) add the call to action button on top, and (iii) not use the full width banner format.



## [See the Latest Alumni News in the Fall 2015 Issue of Update](#)

Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.

[Read Update Magazine »](#)



### [Alumni Legend: Ab Nicholas](#)

Ab Nicholas (B.S. '52, MBA '55) was a record-breaking basketball player for UW-Madison. After earning his bachelor's, he had a choice: play professional basketball or pursue an MBA at the Wisconsin School of Business. [Learn why the Wisconsin MBA was the best play.](#)

## Class Notes

Business Badgers from around the world are earning promotions, accepting new positions, starting new businesses, and marking major life events. [See what's happening](#) in Class Notes.

**Richard Himes (B.S. '66, MBA '69)** passed away this spring. In his name, his daughter, Katherine Himes (MBA '01), launched the Richard Himes Scholarship....  
[Read more »](#)

**Clarke Caywood (BBA '69, Ph.D. '85)** received the first educator's award of the national Black Public Relations Society. The award is named after Ofield Dukes...  
[Read more »](#)

**Marisa Menzel (BBA '00)** was recognized as one of Madison's "40 Under 40" by In Business magazine in its March 2015 issue. In addition to helping clients in...  
[Read more »](#)

Figure 7.1: Next Issue of Update After the Study- Fall 2015.

Table 7.1: Six Binary Design Factors for Update Email Optimization









Factor	+	-
1	<p>Selected articles in a full width banner format</p>  <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p>	<p>Keep as is in the control version</p>  <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p>
2	<p>Stories with very minimal teaser copy</p>  <p><b>What Inspires You?</b> <a href="#">Find out</a> what inspires the co-owners of Splash Studio in Milwaukee, Wis.</p>	<p>Keep as is in the control version</p>  <p><b>What Inspires You?</b> Marla Poytinger (MBA '09, Arts Administration) and David Poytinger (MBA '10, Supply Chain Management) share insight into why teamwork matters. <a href="#">Find out what inspires the Poytingers.</a></p>
3	<p>The class notes story as a set of 3 example notes</p> <p><b>Class Notes</b> Business Badgers from around the world are earning promotions, accepting new positions, starting new businesses, and marking major life events. <a href="#">Read the latest alumni news in Class Notes.</a></p> <div style="display: flex; justify-content: space-between;"> <div style="width: 30%;"> <p><b>Jim Mottern, MBA '74,</b> works with multinational corporations on complicated CFO and CIO issues such as cost reduction and operations improvement ... <a href="#">Read more »</a></p> </div> <div style="width: 30%;"> <p><b>Lisa Harris, BBA '97</b> is a regional director for Sanrio, a former Target buyer, current member of the Wisconsin Business Alumni Board, and a contemporary ... <a href="#">Read more »</a></p> </div> <div style="width: 30%;"> <p><b>Megan Ramey, MBA '07,</b> founded Bikabout.com with her husband Kyle Ramey (MBA '07) to inspire two-wheeled tourism in North America's best biking cities. The website ... <a href="#">Read more »</a></p> </div> </div>	<p>Keep as is in the control version</p>  <p><b>Class Notes</b> Business Badgers from around the world are earning promotions, accepting new positions, starting new businesses, and marking major life events. <a href="#">See what's happening in Class Notes.</a></p>
4	<p>An image of the cover at top</p> <p><a href="#">See the Latest Alumni News in the Spring 2015 Issue of Update</a></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> 	<p>Keep as is in the control version</p>
5	<p>The full width Logo on top</p> 	<p>Keep as is in the control version</p> 
6	<p>A call to action button on top</p> <p><a href="#">Read the Spring 2015 Issue »</a></p>	<p>Keep as is in the control version</p>

Table 7.2: Versions One, Two, Three, and Four - Update Spring 2015


Version one: control version	Version two: 145
1 : -, 2 : -, 3 : -, 4 : -, 5 : -, 6 : -	1 : +, 2 : -, 3 : -, 4 : +, 5 : +, 6 : -
 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p> <p><b>What Inspires You?</b> Marla Poytinger (MBA '09, Arts Administration) and David Poytinger (MBA '10, Supply Chain Management) share insight into why teamwork matters. <a href="#">Find out what inspires the Poytingers.</a></p> <p><b>WSB Research: Now You Know</b> When is a good time to start talking to children about money? How does the long-term value of American-made Toyotas compare to those made in Japan? What makes for a kinder, gentler workplace? <a href="#">Get answers to these intriguing questions.</a></p>	 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p>  <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p>  <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p> <p><b>What Inspires You?</b> Marla Poytinger (MBA '09, Arts Administration) and David Poytinger (MBA '10, Supply Chain Management) share insight into why teamwork matters. <a href="#">Find out what inspires the Poytingers.</a></p>
Version three: 246	Version four: 1256
1 : -, 2 : +, 3 : -, 4 : +, 5 : -, 6 : +	1 : +, 2 : +, 3 : -, 4 : -, 5 : +, 6 : +
 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p>  <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> <p><a href="#">Read the Spring 2015 Issue »</a></p> <p><b>Alumni Inspiration: 8 to Watch</b> Meet eight WSB alumni who are making an impact.</p> <p><b>What Inspires You?</b> Find out what inspires the co-owners of Splash Studio in Milwaukee, Wis.</p> <p><b>WSB Research: Now You Know</b> Get answers that you can trust to important business questions.</p>	 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> <p><a href="#">Read the Spring 2015 Issue »</a></p>  <p><b>Alumni Inspiration: 8 to Watch</b> Meet eight WSB alumni who are making an impact.</p> <p><b>What Inspires You?</b> Find out what inspires the co-owners of Splash Studio in Milwaukee, Wis.</p> <p><b>WSB Research: Now You Know</b> Get answers that you can trust to important business questions.</p>

Table 7.3: Versions Five, Six, Seven, and Eight - Update Spring 2015

Version five: 356 1 : -, 2 : -, 3 : +, 4 : -, 5 : +, 6 : +	Version six: 1346 1 : +, 2 : -, 3 : +, 4 : +, 5 : -, 6 : +
 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> <p><a href="#">Read the Spring 2015 Issue &gt;</a></p>  <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p>	 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p>  <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p> <p><a href="#">Read the Spring 2015 Issue &gt;</a></p>  <p><b>Alumni Inspiration: 8 to Watch</b> The impact of WSB alumni is far-reaching. Meet eight Business Badgers whose influence will be felt in their industries and beyond for years to come. <a href="#">Read about these amazing alumni.</a></p>
Version seven: 2345 1 : -, 2 : +, 3 : +, 4 : +, 5 : +, 6 : -	Version eight: 123 1 : +, 2 : +, 3 : +, 4 : -, 5 : -, 6 : -
 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p>  <p><b>Alumni Inspiration: 8 to Watch</b> Meet eight WSB alumni who are making an impact.</p> 	 <p><b>See the Latest Alumni News in the Spring 2015 Issue of Update</b></p> <p>Find out what's happening with your fellow alumni, and read the inspiring stories coming out of the Wisconsin School of Business today.</p>  <p><b>Alumni Inspiration: 8 to Watch</b> Meet eight WSB alumni who are making an impact.</p> <p><b>What Inspires You?</b> Find out what inspires the co-owners of Splash Studio in Milwaukee, Wis.</p> 

Table 7.4: Descriptive Statistics for Empirical Application

	V1	V2	V3	V4	V5	V6	V7	V8
Recipients	3225	3224	3205	3205	3211	3215	3209	3199
page views	674	327	402	350	587	304	580	337

Table 7.5: Confounded Effects

Labels	Confounded Effects
<b>A</b>	<b>1 = 24 = 35 = 346 = 256 = 1236 = 1456 = 12345</b>
<b>B</b>	<b>2 = 14 = 36 = 345 = 156 = 2456 = 1235 = 12346</b>
<b>C</b>	<b>3 = 15 = 26 = 245 = 146 = 1234 = 3456 = 12356</b>
<b>D</b>	<b>4 = 12 = 56 = 235 = 136 = 1345 = 2346 = 12456</b>
<b>E</b>	<b>5 = 13 = 46 = 126 = 234 = 1245 = 2356 = 13456</b>
<b>F</b>	<b>6 = 23 = 45 = 134 = 125 = 1246 = 1356 = 23456</b>
<b>G</b>	<b>16 = 34 = 25 = 145 = 246 = 356 = 123 = 123456</b>

Table 7.6: Results for  $P_1$ 

Effect	Estimate	P-value
<b>A</b>	0.010	> 0.2
<b>B</b>	0.006	> 0.2
<b>C</b>	0.011	> 0.2
<b>D</b>	0.001	> 0.2
<b>E</b>	-0.025	0.12
<b>F</b>	0.038	0.04 *
<b>G</b>	-0.044	0.03 *

Table 7.7: Results for  $P_2$ 

Effect	Estimate	P-value
<b>A</b>	0.181	> 0.2
<b>B</b>	0.182	> 0.2
<b>C</b>	0.278	0.08 *
<b>D</b>	0.062	> 0.2
<b>E</b>	-0.068	> 0.2
<b>F</b>	-0.055	> 0.2
<b>G</b>	-0.102	> 0.2

Table 7.8: Slice Factor Behavior

Effect	Estimate	P-value
<b>S</b>	0.671	< 0.01 *
<b>3S</b>	0.134	0.14
<b>6S</b>	-0.046	> 0.2
<b>16S</b>	-0.029	> 0.2

## 8 CONCLUSION

---

A unique aspect of multivariate testing in the online space is that testing needs to be conducted across multiple platforms including laptops, tablets, smartphones, and smartwatches. The existing design literature does not offer precise guidance for such a multi-platform context. The primary focus of this paper is to fill this void in the literature. Our primary contribution is that we develop a multi-platform, multivariate design that allows us to uncover effects for each platform and compare test results across different platforms by studying the interactions of design factors and platforms. We develop a new factorial design, called the sliced factorial design, that is used to perform a multivariate multi-platform test.

We propose a novel sliced effect hierarchy principle that generalizes the widely used effect hierarchy principle to the multi-platform context. We show that widely used resolution and aberration criteria fail to satisfy the sliced effect hierarchy principle. To address this problem, we propose sliced resolution and sliced aberration as the design criteria for our multi-platform context. It is well known that minimum aberration designs are popular among practitioners and tables to construct them are readily available in software and textbooks. In an effort to build upon the rich results of the aberration literature, we prove two theorems that connect sliced factorial designs to minimum aberration designs. These theorems help construct sliced factorial designs that we propose. Using the novel design criteria and the theorems, we develop two algorithms to construct sliced factorial designs. The first algorithm is for the case where all combinations of design factors are feasible for all platforms. This algorithm works for both symmetric and asymmetric designs. The second algorithm is usable in situations where some combinations of design

factors may be infeasible for some platforms.

We find that it is desirable to have the sliced factorial design be divided into homogeneous slices: experimental versions made for a slice should be as similar as possible to the ones for other slices. In terms of resources, the multivariate design we propose requires a smaller number of versions compared to other designs. From the standpoint of cost, this is highly desirable aspect of the designs we propose. We illustrate that although slices are used to uncover factor effects within each platform, the sliced factorial design can be used to compare the results across different platforms.

Finally, we illustrate our novel design framework in the context of an empirical email optimization application intended to improve engagement for a digital magazine. The dependent variable of interest was page views and involved over 25000 users. Algorithm 1 was used to generate a sliced factorial design for this study. Google Analytics was used to record the number of page views for each version. Lenth's test, which is well suited for such data, revealed interesting insight about how different factors were effective for different platforms. Based on the results for this multi-platform experiment, the expected gain in page views for the two platforms was 16% and 7%.

In closing, we hope that the proposed design framework will become a useful way for practitioners to design and analyze multi-platform experiments online.

REFERENCES

---

- Addelman, Sidney. 1963. Techniques for constructing fractional replicate plans. *Journal of the American Statistical Association* 58(301):45–71.
- Arora, Neeraj, and Joel Huber. 2001. Improving parameter estimates and model prediction by aggregate customization in choice experiments. *Journal of Consumer Research* 28(2):273–283.
- Atkinson, Anthony, Alexander Donev, and Randall Tobias. 2007. *Optimum experimental designs, with sas*, vol. 34. Oxford University Press.
- Box, George EP, and Norman R Draper. 1959. A basis for the selection of a response surface design. *Journal of the American Statistical Association* 54(287):622–654.
- . 1963. The choice of a second order rotatable design. *Biometrika* 335–352.
- Box, George EP, and J Stuart Hunter. 1961. The  $2^{k-p}$  fractional factorial designs. *Technometrics* 3(3):311–351.
- Box, George EP, William Gordon Hunter, J Stuart Hunter, et al. 1978. *Statistics for experimenters*. John Wiley & Sons.
- Bursztyn, Dizza, and David M Steinberg. 2006. Comparison of designs for computer experiments. *Journal of Statistical Planning and Inference* 136(3):1103–1119.
- Butler, Neil A. 2003. Some theory for constructing minimum aberration fractional factorial designs. *Biometrika* 90(1):233–238.
- Chaffey, Dave. 2016. Mobile marketing statistics compilation, available at <http://www.smartinsights.com/mobile-marketing/mobile-marketing-analytics/mobile-marketing-statistics/>.
- Cheng, C-S, David M Steinberg, and Don X Sun. 1999. Minimum aberration and model robustness for two-level fractional factorial designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61(1):85–93.

- Cheng, Ching-Shui, and Boxin Tang. 2005. A general theory of minimum aberration and its applications. *Annals of Statistics* 944–958.
- Draper, Norman R, and Irwin Guttman. 1992. Treating bias as variance for experimental design purposes. *Annals of the Institute of Statistical Mathematics* 44(4): 659–671.
- Elrod, Terry, Jordan J Louviere, and Krishnakumar S Davey. 1992. An empirical comparison of ratings-based and choice-based conjoint models. *Journal of Marketing research* 29(3):368.
- Fang, Kai-Tai, and Rahul Mukerjee. 2000. Miscellanea. a connection between uniformity and aberration in regular fractions of two-level factorials. *Biometrika* 87(1):193–198.
- Fries, Arthur, and William G Hunter. 1980. Minimum aberration  $2^{k-p}$  designs. *Technometrics* 22(4):601–608.
- Green, Paul E. 1974. On the design of choice experiments involving multifactor alternatives. *Journal of Consumer Research* 1(2):61–68.
- Green, Paul E, and Vithala R Rao. 1971. Conjoint measurement for quantifying judgmental data. *Journal of Marketing research* 355–363.
- Green, Paul E, and Venkat Srinivasan. 1990. Conjoint analysis in marketing: new developments with implications for research and practice. *The Journal of Marketing* 3–19.
- Green, Paul E, and Venkatachary Srinivasan. 1978. Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research* 5(2):103–123.
- Hamada, Michael, and N Balakrishnan. 1998. Analyzing unreplicated factorial experiments: a review with some new proposals. *Statistica Sinica* 1–28.

- Heredia-Langner, Alejandro, Douglas C Montgomery, W Matthew Carlyle, and Connie M Borrer. 2004. Model-robust optimal designs: a genetic algorithm approach. *Journal of Quality Technology* 36(3):263–279.
- Huber, Joel, and Klaus Zwerina. 1996. The importance of utility balance in efficient choice designs. *Journal of Marketing research* 307–317.
- Jacroux, Mike. 2004. A modified minimum aberration criterion for selecting regular  $2^{m-k}$  fractional factorial designs. *Journal of statistical planning and inference* 126(1): 325–336.
- Jones, Bradley, and Christopher J Nachtsheim. 2011. Efficient designs with minimal aliasing. *Technometrics* 53(1):62–71.
- Kessels, Roselinde, Peter Goos, and Martina Vandebroek. 2006. A comparison of criteria to design efficient choice experiments. *Journal of Marketing Research* 43(3): 409–419.
- Kessels, Roselinde, Bradley Jones, Peter Goos, and Martina Vandebroek. 2009. An efficient algorithm for constructing bayesian optimal choice designs. *Journal of Business & Economic Statistics* 27(2):279–291.
- Läuter, Elisabeth. 1974. Experimental design in a class of models. *Statistics: A Journal of Theoretical and Applied Statistics* 5(4-5):379–398.
- Ledolter, Johannes, and Arthur J Swersey. 2007. *Testing 1-2-3: Experimental design with applications in marketing and service operations*. Stanford University Press.
- Lenth, Russell V. 1989. Quick and easy analysis of unreplicated factorials. *Technometrics* 31(4):469–473.
- Liu, Qing, and Neeraj Arora. 2011. Efficient choice designs for a consider-then-choose model. *Marketing Science* 30(2):321–338.

- Louviere, Jordan J, and George Woodworth. 1983. Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of marketing research* 350–367.
- Montepiedra, Grace, and Valery V Fedorov. 1997. Minimum bias designs with constraints. *Journal of Statistical Planning and Inference* 63(1):97–111.
- Montgomery, Douglas C. 2008. *Design and analysis of experiments*. John Wiley & Sons.
- Mukerjee, Rahul, and CF Jeff Wu. 2007. *A modern theory of factorial design*. Springer Science & Business Media.
- Qian, Peter Z G. 2012. Sliced latin hypercube designs. *Journal of the American Statistical Association* 107(497):393–399.
- Sándor, Zsolt, and Michel Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* 38(4):430–444.
- . 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21(4):455–475.
- . 2005. Heterogeneous conjoint choice designs. *Journal of Marketing Research* 42(2):210–218.
- Stanhope, Joe. 2013. The forrester wave™: Online testing platforms, q1 2013. Tech. Rep., Forrester Research.
- Toubia, Olivier, and John R Hauser. 2007. Research note-on managerially efficient experimental designs. *Marketing Science* 26(6):851–858.
- Wu, CF Jeff, and Michael S Hamada. 2011. *Experiments: planning, analysis, and optimization*, vol. 552. John Wiley & Sons.
- Wu, CF Jeff, and Runchu Zhang. 1993. Minimum aberration designs with two-level and four-level factors. *Biometrika* 203–209.

Xu, Xu, BEN Haaland, and Peter Z G Qian. 2011. Sudoku-based space-filling designs. *Biometrika* 98(3):711–720.

Ye, Kenny Q, and Michael Hamada. 2000. Critical values of the lenth method for unreplicated factorial designs. *Journal of Quality Technology* 32(1):57–66.

Yu, Jie, Peter Goos, and Martina Vandebroek. 2009. Efficient conjoint choice designs in the presence of respondent heterogeneity. *Marketing Science* 28(1):122–135.

Zhou, Xiaojie, Lawrence Joseph, David B Wolfson, and Patrick Bélisle. 2003. A bayesian a-optimal and model robust design criterion. *Biometrics* 59(4):1082–1088.