Sliced Minimum Aberration Designs for Four-platform Experiments

Soheil Sadeghi
Department of Statistics at University of Wisconsin-Madison, sadeghi2@wisc.edu

Peter Z. G. Qian
Department of Statistics at University of Wisconsin-Madison, peterq@stat.wisc.edu

Neeraj Arora
Wisconsin School of Business at University of Wisconsin-Madison, neeraj.arora@wisc.edu

April 3, 2017

Multivariate testing is a popular method to improve the layout of digital products such as a website and an application. Fractional factorial designs are usually used to perform online testing with large number of attributes. However, digital spaces present a new design challenge that doest not exist in the traditional experimental design literature: online testing is conducted across multiple platforms including desktops, tablets, smart-phones, and smart-watches. The existing experimental design literature does not offer precise guidance for such a multi-platform context. Sadeghi et al. (2016) introduced a statistical design framework to address the multi-platform feature of digital experiments. Sadeghi et al. (2016) also introduced a novel "sliced effect hierarchy" and formally defined sliced minimum aberration designs for two-platform experiments. In this paper, we extend the sliced minimum aberration designs for four-platform experiments. We define and provide guidance to construct sliced minimum aberration designs for a four-platform experiment using the concepts provided in Wu and Zhang (1993) and the underlying structure of multi-platform experiments provided by Sadeghi et al. (2016). We also tabulate sliced minimum aberration designs with 16, 32, and 64 runs for four-platform experiments.

Key words: Multivariate testing; factorial designs; sliced designs; aberration; blocking; digital experiments

1. Multi-Platform Experiments

Online testing is a popular method to improve the layout of digital products such as a website and an application. It is usually conducted for the purpose of increasing the engagement metrics, e.g. page visitors and click-through rate. In its general form, online testing includes multiple attributes of a digital product and the effects of these attributes are studied on a response variable simultaneously. Fractional factorial designs are used to perform online testing with large number of attributes. One can use the rich literature of factorial designs (Wu and Hamada 2011, Montgomery 2008, Box et al. 1978) to perform an online testing. However, digital spaces present a new design challenge that doest not exist in the traditional experimental design literature: online testing is conducted across multiple platforms including desktops, tablets, smart-phones, and smart-watches. A customer can interact with an application on one of these platforms, and a different set of attribute combinations may optimize her engagement metric for each platform. For example, although presence of multiple images may work the best for an application on a tablet, a series of links might be the best for the same application on a smart-watch. Sadeghi et al. (2016) introduced a statistical design framework to address the multi-platform feature of digital experiments. Let us first provide a formal definition of a multi-platform experiment, borrowed from Sadeghi et al. (2016).

Definition 1 **(Multi-platform Experiment)**. *Consider a multi-platform experiment for studying k two-level design factors, denoted by $1,\ldots,k$, on s platforms P_1, ..., P_s. The complete design set d of the experiment consists of s sub designs, $d_1,\ldots,d_s$, 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 jth level of S is associated with P_j.*

Sadeghi et al. (2016) introduced two guiding properties to construct the designs for the experiment in Definition 1:

Property 1. For $j = 1,\ldots,s$, 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.

As a result of Property 1, each sub design d_j estimates the effects of design factors on platform P_j, and according to effect hierarchy (Wu and Hamada 2011, p. 172), the focus of estimation is on the lower-order effects - main effects and two-way interactions. On the other hand, Property 2 suggests for the complete design d to focus on the estimation of the slice factor S and its two-way interactions with the design factors. This requires a different ordering of effects than the effect hierarchy for the complete design d in which S is more likely to be important than main effects of the design factors, and two-way interaction effects of S with the design factors are more likely to be important than two-way interaction effects of the design factors. Sadeghi et al. (2016) proposed the *sliced effect hierarchy* for the complete design d in order to accommodate Property 2. To formally define this ordering of effects, for the design d in Definition 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, Sadeghi et al. (2016) defined the sliced effect hierarchy as follows:

Sliced Effect Hierarchy

(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.

In a multi-platform experiment, the slice factor is different from the design factors in two ways. First, a multi-platform experiment aims to detect what level of the design factors should be chosen for each platform, and is not trying to select between platforms. Second, according to the sliced effect hierarchy, 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 the design factors. Next we use a simple example to illustrate the slice effect hierarchy and how it differs from the well-known effect hierarchy in the literature.

EXAMPLE 1. For the experiment in Definition 1, let $k=3$ and $s=4$ for four platforms P_1, P_2, P_3 and P_4. The slice factor S is a categorical factor with four levels such that its jth level is associated with P_j. Sub designs $d_1, \ldots, d_4$ are factorial designs with three factors such that each includes seven factorial effects that are ranked in Table 1 following the effect hierarchy. All seven factorial effects have one degree of freedom. The complete design d is a factorial design with three design factors

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

Rank	Effects
(i)	1, 2, 3
(ii)	12, 13, 23
(iii)	123

and the slice factor S, and includes fifteen factorial effects (see Table 2). The two sets E_I and E_S are $\{1,2,3,12,13,23,123\}$ and $\{S,1S,2S,3S,12S,13S,23S,123S\}$, respectively. All effects of the set E_I have one degree of freedom although the effects of the set E_S have three degrees of freedom. Following the sliced effect hierarchy defined above, Table 2 ranks all fifteen factorial effects.

Table 2 Sliced Effect Hierarchy for the Complete Design of the Experiment in Example 1

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

According to Properties 1 and 2, existing methods in the literature fail to conform with the sliced effect hierarchy in order to construct for the experiment in Definition 1. First, although it is simple to implement a random splitting approach in which a complete design d is constructed and split randomly into sub designs d_j's, there is no guarantee each sub design d_j satisfies Property 1. Second, independent construction of sub designs d_j's and combining them together to form d cannot guarantee the complete design d satisfies Property 2. Third, blocked designs that are used to form blocks of homogeneous units with the aim of reducing estimation variance are ill-suited for multi-platform digital experiments. Blocking scheme suggests using the slice factor as a block factor to construct s blocks $d_1, \ldots, d_s$ in order to form d. This assumes little to no change of the design factors effects from a platform to another which means the interaction of the slice factor S with the design factors are negligible. This assumption contradicts the ordering effect of the sliced effect hierarchy and fails to satisfy Property 2.

In view of the drawbacks of the aforementioned methods, Sadeghi et al. (2016) proposed new designs, called *sliced factorial designs*, for the multi-platform experiment in Definition 1. Their basic idea for a sliced factorial design is that each sub design d_j follows the effect hierarchy and the complete design d follows the sliced effect hierarchy. Sliced factorial designs are constructed by extending aberration based criterion to accommodate the sliced effect hierarchy. As each sub design d_j follows the effect hierarchy, minimum aberration criterion can be used to judge the goodness of sub designs. However, for the complete design d, an extension of the minimum aberration criterion is needed to accommodate the sliced effect hierarchy. Sadeghi et al. (2016) formally defined *sliced minimum aberration* criterion for the experiment in Definition 1 with two platforms. They proved a theorem to help construct sliced minimum aberration designs for two-platform experiments and generalized the same idea to propose an algorithm for a general multi-platform experiment. Although Sadeghi et al. (2016)'s proposed algorithm to construct sliced factorial designs works for the experiment in Definition 1, their mathematical extension of aberration based criterion is only for the case with two platforms. In this paper, we formally define and extend the sliced minimum aberration designs for the experiment in Definition 1 with four platforms. Specifically, we propose to extend the method of replacement, originally proposed by Addelman (1963), to construct sliced factorial designs for four-platform experiments. Wu and Zhang (1993) proposed an aberration based criterion for a combination of two-level and four-level factorial designs using the method of replacement. We define a sliced minimum aberration criterion for a four-platform experiment using the concepts provided in Wu and Zhang (1993) and the underlying structure of multi-platform experiments provided by Sadeghi et al. (2016). Our extension is consistent with the general algorithm of Sadeghi et al. (2016). We also tabulate sliced minimum aberration designs with 16, 32, and 64 runs for four-platform experiments.

2. Construction of 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 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 3.

Table 3 **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	$\longrightarrow$	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 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 $s_3=s_1s_2$, and 2-level columns are represented by $1,\dots,k$. For a two-platform experiment, Sadeghi et al. (2016) 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 of Sadeghi et al. (2016) 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 W in the defining relation of d appears in three aliasing relations for the slice factor S as s_1W, s_2W, and s_3W. We extend Sadeghi et al. (2016)'s definition of the sliced wordlength pattern to be defined over the minimum length of s_1W, s_2W, and s_3W. This extension follows the main idea of the sliced minimum aberration criterion, proposed by Sadeghi et al. (2016), which is minimizing the number of shortest length of a sliced wordlength pattern. Defining the sliced wordlength pattern over the minimum length of s_1W, s_2W, and s_3W 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 $1, \ldots, k$, which is called type 0, and the second involves one of the s_j's and some of the design factors $1, \ldots, k$, 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

$$W(d) = ([A_{i0}(d), A_{i1}(d)])_{i\geq 3}, \tag{1}$$

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 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) and Sadeghi et al. (2016) to formally define the sliced wordlength pattern of a design set d for a four-platform experiment as follows:

Definition 2 **(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 1 with $s = 4$, the sliced wordlength pattern is the vector $SW(d) = ([SA_{i0}(d), SA_{i1}(d)])_{i\geq 2}$, where*

- $SA_{i0}(d) = A_{(i+1)1}(d)$ *where* $i \geq 2$
- $SA_{i1}(d) = A_{(i-1)0}(d)$ *where* $i \geq 4$
- $SA_{i1}(d) = 0$ *where* $i = 2, 3$,

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

A type 0 word W 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 $SA_{i1}(d) = A_{(i-1)0}(d)$. On the other hand, a type 1 word W 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 $SA_{i0}(d) = A_{(i+1)1}(d)$ 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 $SA_{i0}(d)$ and $SA_{i1}(d)$ 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 $SA_{i1}(d)$ than a smaller $SA_{i0}(d)$ for the same i. We thus define sliced minimum aberration designs for a four-platform experiment as follows:

Definition 3 **(Sliced Minimum Aberration Designs)**. *Suppose that, for the experiment in Definition 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 $[SA_{r0}(d^{(1)}), SA_{r1}(d^{(1)})] \neq [SA_{r0}(d^{(2)}), SA_{r1}(d^{(2)})]$. *If* $SA_{r1}(d^{(1)}) < SA_{r1}(d^{(2)})$, *or* $SA_{r1}(d^{(1)}) = SA_{r1}(d^{(2)})$ *but* $SA_{r0}(d^{(1)}) < SA_{r0}(d^{(2)})$, *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 two Definitions 2 and 3, we consider the following example:

Example 2. For the experiment in Definition 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:

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

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

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:

$$s_1 = 124s_1 = 135s_1 = 2345s_1$$

$$s_2 = 124s_2 = 135s_2 = 2345s_2$$

$$s_3 = 124s_3 = 135s_3 = 2345s_3.$$

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 $SA_{31}(d^{(1)}) = SA_{31}(d^{(2)}) = 0$ and $SA_{30}(d^{(1)}) = 0 < 2 = SA_{30}(d^{(2)})$. 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 1 below guides the construction of the sliced minimum aberration designs using readily available minimum aberration designs of fewer number of factors.

THEOREM 1. *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 1, 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 1 to provide sliced minimum aberration designs with 16, 32, and 64 runs in Tables 4, 5, and 6 receptively.

Table 4 Sliced minimum aberration designs with 16 runs, $S=(s_1, s_2, s_1 s_2)$

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

Table 5 **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 6 **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	$S,1,2,3,4,123,124,134,234$	$([0,0]_4,[0,14]_5,[0,0]_6,[0,0]_7,\ldots)$
9	$S,1,2,3,4,123,124,134,234,1234$	$([0,4]_4,[0,14]_5,[0,8]_6,[0,0]_7,\ldots)$
10	$S,1,2,3,4,123,124,134,234,1234,34$	$([0,8]_4,[0,18]_5,[0,16]_6,[0,8]_7,\ldots)$
11	$S,1,2,3,4,123,124,134,234,1234,34,24$	$([0,12]_4,[0,26]_5,[0,28]_6,[0,24]_7,\ldots)$
12	$S,1,2,3,4,123,124,134,234,1234,34,24,14$	$([0,16]_4,[0,39]_5,[0,48]_6,[0,48]_7,\ldots)$
13	$S,1,2,3,4,123,124,134,234,1234,34,24,14,23$	$([0,22]_4,[0,55]_5,[0,72]_6,[0,96]_7,\ldots)$
14	$S,1,2,3,4,123,124,134,234,1234,34,24,14,23,13$	$([0,28]_4,[0,77]_5,[0,112]_6,[0,168]_7,\ldots)$
15	$S,1,2,3,4,123,124,134,234,1234,34,24,14,23,13,12$	$([0,33]_4,[0,105]_5,[0,168]_6,[0,280]_7,\ldots)$

3. Conclusion

A unique aspect of multivariate testing in the online space is that testing needs to be conducted across multiple platforms. 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 by developing over the factorial design literature that allows us to uncover effects for each platform and compare test results across different platforms. We begin by a novel sliced effect hierarchy that generalizes the widely used effect hierarchy to the multi-platform context. Based on the sliced effect hierarchy, we develop sliced factorial designs for four-platform experiments. Specifically, we propose to extend the method of replacement, originally proposed by Addelman (1963), to construct sliced factorial designs for four-platform experiments. We define and provide guidance to construct sliced minimum aberration designs for a four-platform experiment using the concepts provided in Wu and Zhang (1993) and the underlying structure of multi-platform experiments provided by Sadeghi et al. (2016). We also tabulate sliced minimum aberration designs with 16, 32, and 64 runs for four-platform experiments.

References

Addelman, S. (1963). Techniques for constructing fractional replicate plans. *Journal of the American Statistical Association*, 58(301):45–71.

Box, G. E., Hunter, W. G., Hunter, J. S., et al. (1978). *Statistics for experimenters*. John Wiley & Sons.

Montgomery, D. C. (2008). *Design and analysis of experiments*. John Wiley & Sons.

Sadeghi, S., Qian, P. Z., and Arora, N. (2016). Sliced designs for multi-platform online experiments.

Wu, C. J. and Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*, volume 552. John Wiley & Sons.

Wu, C. J. and Zhang, R. (1993). Minimum aberration designs with two-level and four-level factors. *Biometrika*, pages 203–209.