Ann Inst Stat Math (2018) 70:249–268
https://doi.org/10.1007/s10463-018-0646-0

INVITED ARTICLE: AKAIKE MEMORIAL LECTURE

A fresh look at effect aliasing and interactions: some new wine in old bottles

C. F. Jeff Wu[1]

Received: 1 March 2017 / Revised: 14 June 2017 / Published online: 9 February 2018

Abstract Interactions and effect aliasing are among the fundamental concepts in experimental design. In this paper, some new insights and approaches are provided on these subjects. In the literature, the "de-aliasing" of aliased effects is deemed to be impossible. We argue that this "impossibility" can indeed be resolved by employing a new approach which consists of reparametrization of effects and exploitation of effect non-orthogonality. This approach is successfully applied to three classes of designs: regular and nonregular two-level fractional factorial designs, and three-level fractional factorial designs. For reparametrization, the notion of conditional main effects (cme's) is employed for two-level regular designs, while the linear-quadratic system is used for three-level designs. For nonregular two-level designs, reparametrization is not needed because the partial aliasing of their effects already induces non-orthogonality. The approach can be extended to general observational data by using a new bi-level variable selection technique based on the cme's. A historical recollection is given on how these ideas were discovered.

Keywords Conditional main effects · Fractional factorial designs · Nonregular designs · Orthogonal arrays

---

The Related Articles are https://doi.org/10.1007/s10463-017-0639-4; https://doi.org/10.1007/s10463-017-0640-y; https://doi.org/10.1007/s10463-017-0641-x.

✉ C. F. Jeff Wu
jeff.wu@isye.gatech.edu

1 H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, 755 Ferst Drive, NW, Atlanta, GA 30332, USA

1 Introduction

When it is expensive or unaffordable to run a full factorial experiment, a fractional factorial design is used instead. Since there is no free lunch for getting run size economy, a price to pay for using fractional factorial design is the aliasing of effects. Effect aliasing can be handled in different ways. Background knowledge may suggest that one effect in the aliased set is insignificant, thus making the other aliased effect estimable in the analysis. Alternatively, a follow-up experiment may be conducted, specifically to de-alias the set of aliased effects. Details on these strategies can be found in design texts like Box et al. (2005) or Wu and Hamada (2009). Another problem with effect aliasing is the difficulty in interpreting the significance of aliased effects in data analysis.

Ever since the pioneering work of Finney (1945) on fractional factorial designs and effect aliasing, it has been taken for granted that aliased effects can only be de-aliased by adding more runs. The main purpose of this paper is to show that, for three classes of factorial designs, there are strategies that can be used to de-alias aliased effects *without* the need to conduct additional runs. Each of the three cases has been studied in prior publications, but this paper is the first one to examine this class of problems with a fresh new look and in a unified framework. It also contains some additional results and insights. When discussing effect aliasing, it is unavoidable to bring up the major role interactions play in the factorial setting. Because main effects are not allowed to be aliased with other main effects (in order to keep each factor meaningful), at least one effect in the aliased set is an interaction. A key concept in the de-aliasing strategy in this paper is to *reparametrize* the interactions in a certain way to create *non-orthogonality* among some effects so that effect estimability becomes possible. The reparametrization scheme depends on the nature of designs under consideration. It will be developed in Sects. 2 and 4.

The first class of designs being considered is the two-level fractional 2^{k-q} designs, which are the simplest of the three. The main concept here is to use the conditional main effects (cme's) to reparametrize the three-dimensional space generated by the two main effects, say A and B, and their two-factor interaction AB. It will be shown in Sect. 2 that aliased effects involving one or two factors can be de-aliased by running some analysis in the space generated by the main effects and the cme's. This strategy is called the CME analysis. While this was originally suggested in the context of designed experiments, the cme's can be viewed as a new class of basis functions in variable selection involving factors with two levels. A more general strategy than CME analysis is discussed in Sect. 2.2 that can handle bi-level variable selection for general observational data. A prominent example is genetics, where each gene can be viewed as a factor with two levels, i.e., gene present or absent.

Then we move in Sect. 3 to the next level of complexity in designs, namely, two-level designs not of the 2^{k-q} type. They include many commonly used orthogonal arrays (OAs). To distinguish them from the 2^{k-q} designs in Sect. 2, we call the former *nonregular* and the latter *regular*. A rigorous discussion on their distinction and some historical notes are given in Sect. 3.1. Several key concepts are considered: full versus partial aliasing, and complex aliasing. Here we do not need to reparametrize the interactions because the "nonregular" nature of designs already endows some degrees of non-orthogonality among effects. Because the effects are not fully aliased, a data analysis strategy can be used to estimate some interactions and main effects.

In Sect. 4 we consider three-level fractional 3^{k-q} designs. These designs are constructed by using the same algebraic tools as the 2^{k-q} designs. Therefore one may assume that they are of the regular type. It is not necessarily so. Because each factor has two degrees of freedom, there are different ways to parametrize the factorial effects. Two are considered: orthogonal components system and linear-quadratic system. A 3^{k-q} design endowed with the orthogonal components system is of the regular type, while that endowed with the linear-quadratic system is of the nonregular type. Therefore the linear-quadratic system can be viewed as a reparametrization that allows aliased effects to become estimable. In Sect. 5 a historical recollection is given on how I and/or coauthors discovered these ideas. Like in most scientific discoveries, they did not come about in a straight and logical order. Section 6 contains some concluding remarks.

2 De-aliasing aliased effects in two-level fractional factorial experiments

First we review the key concepts of conditional main effects (cme's) and two-factor interactions (Wu and Hamada 2009, Chapter 4). Suppose A and B are two factors in a two-level factorial experiment. Denote the two levels by + and $-$. Define the *conditional main effect* of A given B at the + level as:

$$\text{CME}(A|B+) = \bar{y}(A+|B+) - \bar{y}(A-|B+), \tag{1}$$

where $\bar{y}(A+|B+)$ and $\bar{y}(A-|B+)$ are the averages of the response y at the level settings $A+B+$ and $A-B+$, respectively.

Similarly we can define the conditional main effect of A given B at the $-$ level as:

$$\text{CME}(A|B-) = \bar{y}(A+|B-) - \bar{y}(A-|B-), \tag{2}$$

where $\bar{y}(A+|B-)$ and $\bar{y}(A-|B-)$ are similarly defined. It is easy to see that the average of the two equals the *main effect* of A, i.e.,

$$\frac{1}{2}\{\text{CME}(A|B+) + \text{CME}(A|B-)\} = \bar{y}(A+) - \bar{y}(A-) = ME(A),$$

where $\bar{y}(A+)$ and $\bar{y}(A-)$ are the averages of y at level settings $A+$ and $A-$.

We can use the difference between the two cme's in (1) and (2) to define the two-factor interaction between A and B, i.e.,

$$\text{INT}(A, B) = \frac{1}{2}\{\text{CME}(A|B+) - \text{CME}(A|B-)\}. \tag{3}$$

By reversing the role of A and B, we have the following expression:

$$\text{INT}(A, B) = \frac{1}{2}\{\text{CME}(B|A+) - \text{CME}(B|A-)\}. \tag{4}$$

By rearranging its four terms, INT(A, B) can be rewritten as:

$$\text{INT}(A, B) = \frac{1}{2}\{\bar{y}(A+|B+)+\bar{y}(A-|B-)\}-\frac{1}{2}\{\bar{y}(A+|B-)+\bar{y}(A-|B+)\}, \tag{5}$$

which is the algebraic expression used in calculating the interaction effect. However, this expression does not convey the physical meaning of being an "interaction". Using an analogy to numerical mathematics, CME$(A|B+)$ can be viewed as the finite difference of first order of the response y between two successive levels of factor A conditioned on B at level +. And CME$(A|B-)$ has a similar interpretation in terms of finite difference. From (3), the interaction INT(A, B) is the difference of two successive finite differences, which is a second-order quantity. This way of viewing interaction conveys the physical meaning of being an interaction. For example, suppose CME$(A|B+)$ is large while CME$(A|B-)$ is near zero. Then the interaction INT(A, B) is large and it means that the magnitude of the "main effect" of A depends on whether B is at level + or $-$, i.e., the two factors *jointly* affect the response values, which corresponds to our intuition about interaction. (For a comprehensive review of the concepts of interactions in broader contexts, read the excellent paper by Cox (1984).) An infinitesimal version of this interpretation is via calculus. We can view the main effect $ME(A)$ as the first-order derivative and the interaction INT(A, B) as the second-order derivative. The second derivative is the derivative of the two first-order "derivatives" CME$(A|B+)$ and CME$(A|B-)$ evaluated at $B+$ and $B-$, respectively.

The definition of interaction given in (3) and (5) has been around for a long time, dating back to Fisher's original treatise in 1935. See its seventh edition [page 98 of Fisher (1971)]. The reason that we gave it the explicit name of conditional main effects (cme's) in Wu and Hamada (2000, 2009) is that the two cme's CME$(A|B+)$ and CME$(A|B-)$ should be viewed as two *interaction components* between A and B.

2.1 De-aliasing via the CME reparametrization

Before we can explain why and how cme's can be used to de-alias aliased effects, we need to define the concept of aliasing. We use the following simple example to illustrate the general idea. Consider the 2^{4-1} design given by $\mathbf{I} = ABCD$. It is a half fraction of the 2^4 design for the four factors A, B, C and D, i.e., it has 8 runs for four factors. The two-factor interactions (abbreviated as 2fi's) AB and CD are given in Table 1. (Previously we used the long-hand notation INT(A, B) and INT(C, D).) Notice that the column vector (or called contrast) for AB and for CD in Table 1 is identical. Therefore neither can be estimated with the given design. The two 2fi's AB and CD are said to be *aliased* (Finney 1945), because they represent the *same* contrast vector in the matrix. Notationally they are denoted as $AB = CD$. Because there are not enough degrees of freedom for all the factorial effects in a fractional factorial design, the concept of aliasing is a necessary evil in order to perform data analysis for such experiments. Ever since the pioneering work of Finney, it had been taken for granted that aliased effects like AB and CD cannot be de-aliased unless further runs are taken.

Table 1 A 2^{4-1} design with $\mathbf{I} = ABCD$

A	B	C	D	AB	$= CD$
−	−	−	−	+	+
−	−	+	+	+	+
−	+	−	+	−	−
−	+	+	−	−	−
+	−	−	+	−	−
+	−	+	−	−	−
+	+	−	−	+	+
+	+	+	+	+	+

The main idea in the work of Wu (2015) and Su and Wu (2017) is to use the cme's to reparametrize the space of factorial effects. Traditionally the factorial effects are defined in terms of main effects, two-factor interactions and higher-order interactions. Consider, for now, only the main effects $ME(A)$, $ME(B)$ and their two-factor interaction INT(A, B). These three effects form a subspace of dimension three, and they are orthogonal to each other. The key to unlocking the aliasing is to *reparametrize* this subspace by using the concept of cme's. For example, the same subspace can be defined in terms of CME$(A|B+)$, CME$(A|B-)$ and $ME(B)$, the last being the main effect of the conditioning factor B in the cme's. From now on, we will use the shorthand notation $(A|B+)$, $(A|B-)$ and B, etc. The key difference is that the three effects in the latter representation are not mutually orthogonal. This lack of orthogonality allows the de-aliasing of aliased effects to be done without adding further runs to the experiment. Therefore Wu (2015) refers to this *non-orthogonality as the saving grace*. Subsequently, the use of cme's and main effects and their inter-relationships was fully developed into an easy-to-use data analysis strategy for 2^{k-q} designs of resolution III and IV by Su and Wu (2017). For brevity, we may refer to this paper as SW.

To save space, let us refer to Table 1 of SW. It has 8 rows for the 8 runs and 14 columns consisting of four main effects, A, B, C, D, two 2fi's AB and CD, and 8 cme's involving A and B or C and D. The eight cme's are only a subset of all possible cme's. Because there are many more columns than rows, clearly some of the columns are not mutually orthogonal. As shown on page 4 of Su and Wu (2017), some non-orthogonal pairs are: (i) any pair among $(A|B+)$, $(B|A+)$, $(C|D-)$, $(D|C-)$, (ii) any pair among $(A|B+)$, $(A|C+)$, $(A|D-)$. This non-orthogonality among some cme's provides the opportunity for some of them to be estimable in addition to the main effects. SW identifies five key properties concerning the cme's, from which they propose three rules for data analysis. They then develop a new data analysis strategy called *CME analysis* based on these rules. The strategy works out well for each of the three real experiments they considered. Note that each example uses a resolution IV design. In the traditional wisdom, 2fi's in resolution IV designs that are aliased with other 2fi's or main effects cannot be estimated. Standard practice as recommended in design text books is to expand the experiment to either a resolution V design or, to save run size, to use an optimal design algorithm to add enough runs to ensure estimability of the aliased 2fi's under consideration [see Chapter 5 of Wu and Hamada (2009)].

The CME analysis is powerful in that it replaces the estimation of aliased 2fi's by the estimation of some cme's associated with the given 2fi's.

For illustration, consider the filtration experiment in SW, where a 2^{4-1}_{IV} design is run with the defining contrast $\mathbf{I} = ABCD$. The traditional analysis using a half-normal plot reveals five significant effects: A, AD, DB, D and C. Following Rule 1 in SW, the effects A and AD are replaced with the cme $(A|D+)$, and the effects D and DB are replaced with the cme $(D|B-)$, which leaves the cme model with three significant effects: $(A|D+)$, $(D|B-)$ and C. While these two models enjoy similar R^2 values, the latter has two notable advantages. First, the effects for the cme model have sizably lower p-values than that for the traditional model. Second, the cme model has much better engineering interpretability. For example, the selected cme $(D|B-)$ indicates the effectiveness of stirring rate *only* at a low concentration of formaldehyde, whereas the interaction effect DB is difficult to interpret.

2.2 Beyond designed experiments: bi-level variable selection in observational data

Because cme's are highly interpretable for a wide range of applications beyond designed experiments, a natural question is whether the CME approach in Sect. 2.1 can be expanded to observational data. Since observational studies are in greater abundance than designed experiments (witness the explosion of data in the internet), such an extension is potentially of great value. Two key distinctions are noted, however. First, the orthogonal framework in Su and Wu (2017), which motivated the three important cme groups of twin, sibling and family effects, is not applicable to observational data, because orthogonality of effects rarely occurs outside of designed experiments like 2^{k-q} designs. In a recent work, Mak and Wu (2017) proposed a new framework for this more general setting. Specifically, they identified four effect groups: siblings, cousins, parent-child pairs and uncle-nephew pairs, which capture the *correlation* structure of cme's in the non-orthogonal setting. The second difference is in the goal of study. While the disentangling of aliased effects is the primary goal of the CME analysis for designed experiments, the separation of active effects from correlated groups of inert effects is the primary interest in the Mak–Wu framework for observation studies. To this end, they proposed a new method called `cmenet`, which can identify both active cme groups and active effects within such groups. This so-called bi-level variable selection is achieved using a penalty function with two layers: the outer layer controlling between-group selection, and the inner layer controlling within-group selection.

It is worth noting that the Mak–Wu framework not only expands CME analysis to the non-orthogonal setting, but also extends important selection principles for designed experiments to observation data. For example, the penalization approach in Mak and Wu (2017) highlights two selection principles called cme coupling and cme reduction. The first allows a cme, say $(A|B+)$, to more easily enter the model when effects in its sibling or cousin group (say, $\{(A|C+), (A|D+), \ldots\}$ or $\{(C|B+), (D|B+), \ldots\}$) have already been selected. The second allows a main effect, say A, to more easily enter the model when many of its sibling or cousin cme's (say, $\{(A|B+), (A|C+), \ldots\}$ or $\{(B|A+), (C|A+), \ldots\}$) have already been selected. These two features are quite intuitive, and parallel the principles of effect heredity and effect hierarchy (Wu and Hamada 2009), which are used to guide model selection in designed experiments.

This extension also opens up new and exciting directions in genomics and social sciences. Consider, for illustration, the gene association example in Mak and Wu (2017), where the goal is to choose important genes affecting the wing shape of the common fruit fly. Here, the cme $(A|B+)$ indicates the significance of gene A *only* when gene B is active, so the selection of cme's provides valuable insight on gene activation behavior. In their paper, `cmenet` is compared with two popular variable selection techniques: the Lasso (Tibshirani 1996) and SparseNet (Mazumder et al. 2011), with the latter two performing selection on the traditional main effects and 2fi's. The new selection method has two advantages: it not only gives reduced prediction error, but also provides insight on gene activation patterns. For example, while both Lasso and SparseNet deemed $V4$ (i.e., the fourth polygene) to be active, `cmenet` instead selected the cme's $V4|V1+$, $V4|V33+$, $V10|V4+$ and $V31|V4+$, thereby providing a more nuanced conclusion that $V4$ is active only in the presence of or in activating other polygenes. This new method can therefore be used to investigate why some genes are conditionally active, and why some play a more supportive role in activating other genes.

3 Estimation of interactions in experiments with complex aliasing

The phenomenon of complex aliasing among factorial effects was first known for the 12-run Plackett–Burman design. Since this design is a special case of nonregular fractional factorial designs [see Chapter 8 of Wu and Hamada (2009)], we review in Sect. 3.1 regular and nonregular designs and discuss their algebraic and statistical properties.

3.1 A dichotomous classification of fractional factorial designs: regular vs. nonregular

Regular designs like the 2^{k-q}, 3^{k-q}, and p^{k-q} series, where p is a prime power, are the most commonly used designs in practice. In a non-mathematical sense, they are constructed as fractions of the corresponding full factorial designs by using the so-called defining contrast subgroup. This subgroup defines which fraction of the full factorial is chosen. To illustrate this for nontechnical readers, consider the 2^{6-2} design with the defining contrast subgroup $\mathbf{I} = \mathbf{125} = \mathbf{1346} = \mathbf{23456}$, where $\mathbf{I}$ is the identity element in the group. This is a quarter fraction of the 2^6 design, i.e., it has $16(= 2^4)$ runs to accommodate 6 factors. Its tabulation is given in Table 2. Its first four columns form the full factorial 2^4 design. In order to add more factors without increasing run size, define factor **5** by using the interaction column for **12**. Thus the main effect **5** is aliased with the 2fi **12**. Notationally, it is written as $\mathbf{5} = \mathbf{12}$ or $\mathbf{I} = \mathbf{125}$. Similarly define factor **6** by using the interaction column **2345**. Thus the main effect **6** is aliased with the four-factor interaction **2345**, i.e., $\mathbf{I} = \mathbf{23456}$. Each of $\mathbf{I} = \mathbf{125}$ and $\mathbf{I} = \mathbf{23456}$ is called a defining relation and defines a half fraction of the 2^4 design. With two of them, a quarter fraction is obtained. To complete the subgroup, we need to generate

Table 2 Design matrix for the 2^{6-2} design

Run	**1**	**2**	**3**	**4**	**5 = 12**	**6 = 2345**
1	−	−	−	−	+	−
2	−	−	−	+	+	+
3	−	−	+	−	+	+
4	−	−	+	+	+	−
5	−	+	−	−	−	−
6	−	+	−	+	−	+
7	−	+	+	−	−	+
8	−	+	+	+	−	−
9	+	−	−	−	−	+
10	+	−	−	+	−	−
11	+	−	+	−	−	−
12	+	−	+	+	−	+
13	+	+	−	−	+	+
14	+	+	−	+	+	−
15	+	+	+	−	+	−
16	+	+	+	+	+	+

another defining relation by multiplying the two relations: $\mathbf{125} \times \mathbf{23456} = \mathbf{1346}$, thus obtaining the third relation $\mathbf{I} = \mathbf{1346}$. Note that only two out of the three relations are independent because the third can be obtained from the other two. The subgroup $\mathbf{I} = \mathbf{125} = \mathbf{1346} = \mathbf{23456}$ has fifteen "cosets" within the whole group. Each coset corresponds to a degree of freedom in the design. For example, by multiplying **5** to the subgroup, we obtain the following coset $\mathbf{5} = \mathbf{12} = \mathbf{13456} = \mathbf{2346}$, which says that the four factorial effects **5**, **12**, **13456** and **2346** are aliased with each other. That is, they together take up one degree of freedom. As remarked in Sect. 2, they cannot be disentangled for estimation. All the fifteen cosets can be found in equation (5.3) of Wu and Hamada (2009) and they account for all the fifteen degrees of freedom of the 16-run design. For brevity, we may refer to the book as WH.

The above simple approach can be used to construct 3-level and p-level regular designs, where p is a prime but not a prime power. A more general and rigorous approach to the construction and their algebraic properties can be found in theoretical design text books [e.g., Mukerjee and Wu (2006), Cheng (2014)]. The user-friendly way of defining and constructing regular fractions as shown above was first used in the seminal paper by Box and Hunter (1961). Before their paper, construction of regular fractions was based on algebraic tools like Galois field, finite geometry, etc., and was not accessible to readers without such knowledge. The Box-Hunter approach has contributed toward the popularity of regular fractional factorial designs, especially the two-level designs.

A major statistical property of regular fractions is that

"the generalized interaction between any two factorial effects is another factorial effect". (6)

Table 3 Design matrix and lifetime data, cast fatigue experiment

Run	Factor											Lifetime logged
	A	*B*	*C*	*D*	*E*	*F*	*G*	8	9	10	11	
1	+	+	−	+	+	+	−	−	−	+	−	6.058
2	+	−	+	+	+	−	−	−	+	−	+	4.733
3	−	+	+	+	−	−	−	+	−	+	+	4.625
4	+	+	+	−	−	−	+	−	+	+	−	5.899
5	+	+	−	−	−	+	−	+	+	−	+	7.000
6	+	−	−	−	+	−	+	+	−	+	+	5.752
7	−	−	−	+	−	+	+	−	+	+	+	5.682
8	−	−	+	−	+	+	−	+	+	+	−	6.607
9	−	+	−	+	+	−	+	+	+	−	−	5.818
10	+	−	+	+	−	+	+	+	−	−	−	5.917
11	−	+	+	−	+	+	+	−	−	−	+	5.863
12	−	−	−	−	−	−	−	−	−	−	−	4.809

Take the 2^{6-2} design example again. From its algebraic construction, it is easy to show that the product of any two factorial effects involving some of the factors **1**, **2**, **3**, **4**, **5**, **6** is among the factorial effects given in the subgroup or its fifteen cosets. Mathematically, this is equivalent to saying that the absolute "correlation" between any two contrast vectors for two factorial effects is either 0 or 1. This is an easy consequence of the design construction. It will be shown later that this property has an important ramification on the estimability of effects of the design. A converse question is whether this property implies that the design must be constructed as a regular design. In order to state this question formally, we need to define the "opposite" of regular designs, namely, nonregular designs. (Note that the term "correlation" is a misnomer but has been used in the literature for convenience of reference. The correct term should be the normalized cross product between the two contrast vectors representing the two factorial effects.)

A good example of nonregular design is the cast fatigue experiment in Chapter 8 of WH. It is reproduced below as Table 3. The experiment has 12 runs and 7 factors based on the 12-run Hadamard matrix. In its full capacity, this matrix, which is called the 12-run Plackett–Burman design, can accommodate up to 11 factors by using the 11 columns in Table 3. Because its run size is not a power of 2, it cannot be a regular design. More importantly, it does not have the aliasing property of regular designs as described in (6). For example, the interaction AB between factors (i.e., columns) A and B in Table 3 cannot be found among the 11 columns. In fact, AB has a complicated aliasing relationship with other main effects. It is orthogonal to its two parent main effects A and B, but is otherwise non-orthogonal to all other main effects with absolute correlations being $1/3$. See Section 9.1 of WH for details. In general, for this design, the correlation between a given main effect and any 2fi not involving the main effect is either $1/3$ or $-1/3$. Clearly this does not follow the property given in (6).

A more rigorous formulation of the property in (6) is now in order. For simplicity, we only consider two-level designs with n factors and N runs. Let X_d be its model matrix, whose (l, i) entry $x_{li}(d)$ is the level (taking value 1 or -1) of the ith factor for the lth observation. Let the ith column of X_d be $x_i(d)$. Then any factorial effect of order k can be defined as follows. For a subset $S = \{i_1, \ldots i_k\}$ of $\{1, \ldots n\}$, let $x_S(d)$ be the entry-wise product of the k column vectors $x_{i_1}(d), \ldots, x_{i_k}(d)$. It is the column vector that corresponds to the factorial effect involving the k factors $i_1, \ldots, i_k$. For two factorial effects with the corresponding subsets denoted as S and T, it is easy to show that their correlation is equal to $[x_S(d)]^T x_T(d)/N$. For regular two-level designs, its absolute value is either 0 or 1. See Section 15.2 of Cheng (2014). These two effects are said to be orthogonal if the value is 0 and fully aliased if it is 1. An interesting theoretical question is whether the converse is true. Using the theory of indicator functions, Ye (2004) answered this in the affirmative. That is, any two-level factorial design in which any two factorial effects have absolute correlation to be 0 or 1 (and no other value) must be a regular 2^{k-q} design or replicates of a regular 2^{k-q} design.

The equivalence result by Ye (2004) is quite significant because it provides a mathematically rigorous definition for nonregular designs. As noted before, regular designs can be defined in terms of how they are constructed. There is no corresponding definition for nonregular designs since they can be constructed in many different ways [see Hedayat et al. (1999)]. Using the correlation properties between factorial effects provides a unified and rigorous way to classify and define regular and nonregular designs. The terms "regular designs" and "regular fractions" have been used in the design literature. See Chapter 9 of Raktoe et al. (1981). Earlier references include Addelman (1962). The term "irregular" was used in the literature to mean fractions of full factorials that are *not* of the regular type as explained in the discussions on regular designs. Some prominent examples of irregular factions are the $3/2^n$ replicates in Addelman (1961) and the three-quarter replicates and related constructions in Sections 8.10-8.13 of John (1971). Later Raktoe et al. (1981, page 123) defined "irregular designs" as those that are not regular. Because the term "irregular designs" may convey a negative connotation, Wu and Hamada (2000) used the term "nonregular designs" to indicate fractional factorial designs whose factorial effects can have absolute correlations between 0 and 1. In fact the theoretical work by Ye and Cheng-Ye was motivated by the attempt to justify this classification. This classification of designs was used in arranging the chapters in Wu and Hamada (2009). Among the chapters on factorials, Chapters 4–5, 7, and Sections 6.1–6.5 of Chapter 6 are of the regular type, while Chapters 8–9 and Sections 6.6–6.7 of Chapter 6 are of the nonregular type. Further remarks on this aspect will be given in Sect. 5 on three-level designs. The first reference on "nonregular designs" was Sun and Wu (1993). The terms "full aliasing" and "partial aliasing" were first used in an informal way on pages 131 and 132 of Hamada and Wu (1992). A more rigorous description of partial aliasing in terms of the correlation was given in Sun and Wu (1993).

3.2 Nonregular designs and their practical use

A collection of nonregular designs can be found in Chapter 9 of WH under the heading of orthogonal arrays (OAs). Three of them will be highlighted here. First is OA$(12, 2^{11})$, which is the same as the 12-run matrix in Table 3. It is also called the 12-run Plackett–Burman (1946) design because these authors were the first to propose the use of Hadamard matrices of order $4m$ for running two-level experiments with run size $4m$ and up to $4m - 1$ factors. The original motivation was to save experimental cost during WWII. See page 87 of Barker and Milivojevich (2016). This can be seen as follows. Suppose k factors, $8 \leq k \leq 11$, are to be studied. If a regular 2^{k-q} design is used, it would require 16 runs. Use of OA(12) can save four runs. Second is OA$(18, 2^1 3^7)$, which can be used to study seven three-level factors and one two-level factor. It is economical because it can accommodate factors with 15 $(= 1 + 2 \times 7)$ degrees of freedom for their main effects. A related one is OA$(18, 6^1 3^6)$, which can accommodate 17 $(= 5 + 2 \times 6)$ degrees of freedom for the main effects and is called a saturated design. Imagine that k three-level factors, $5 \leq k \leq 7$, are to be studied. Use of a regular 3^{k-q} design would require 27 runs like a 3^{k-q} design, $k - q = 3$ and $5 \leq k \leq 7$. Thus there is a saving of 9 runs. The third one is OA$(36, 2^{11} 3^{12})$, which can study up to 11 two-level factors and up to 12 three-level factors. In the maximum case, it is a saturated design because it can accommodate 35 $(= 11 + 2 \times 12)$ degrees of freedom for the main effects. Clearly the main rationale was *run size economy*. The second rationale as explained in Chapter 9 of WH is the *flexibility* in the level-combinations. From reading the OAs in the appendix of Chapter 9, different combinations of 2, 3, 4, 5 and 6 levels can be accommodated with run size ranging from 12 to 54. These three arrays are also called L_{12}, L_{18} and L_{36}, a notation used by G. Taguchi (1987) and much earlier in his Japanese books.

The designs OA$(18, 2^1 3^7)$ and OA$(36, 2^{11} 3^{12})$ were not used in practical experiments in the west until the mid 80's, when G. Taguchi introduced them in the US and Europe for running quality engineering experiments. I personally think Taguchi's main motivation was the run size economy referred to above, but I cannot pinpoint any reference in his writings. An indirect evidence can be seen from how he got involved in the construction of these arrays. Take, for example, OA$(18, 2^1 3^7)$. The first construction was by Masuyama (1957), which constructed the OA$(18, 3^7)$ portion by using the method of difference sets. Seeing that an extra degree of freedom can be accommodated, Taguchi (1987) (and earlier references) added the two-level factor as its first column (see Table 8C.2 of WH). A more dramatic example is how OA$(36, 2^{11} 3^{12})$ was obtained. Seiden (1954) constructed the OA$(36, 3^{12})$ portion, but her motivation was purely theoretical. Recognizing that this array uses up only 24 $(= 2 \times 12)$ degrees of freedom for the main effects, Taguchi (1987) (and earlier references) added three replicates of OA$(12, 2^{11})$ to make up the first 11 columns of OA$(36, 2^{11} 3^{12})$ (see Table 8C.6 of WH). It should be clear from these two examples of adding columns that his interest was to find arrays to accommodate as many factors as possible, thus the rationale of run size economy. Other examples can be found in the appendix of Chapter 9 of WH. In particular, the concept and construction of *nearly orthogonal arrays* (Wang and Wu 1992) was an inspired attempt to adding even more factors by slightly sacrificing the orthogonality requirement.

3.3 Exploitation of partial aliasing for the estimation of interactions

Two factorial effects are said to be *partially aliased* if their absolute correlation is between 0 and 1 and *fully aliased* if the value is 1. The adverb “fully” was first used in Hamada and Wu (1992) to distinguish it from partial aliasing. These two terms are crucial in the classification of designs into regular and nonregular as discussed in Sect. 3.1. Since then, the terminology has been adopted in Wu and Hamada (2000) and Wu and Hamada (2009). In this section, I will explain how partial aliasing can be exploited for estimating interactions in nonregular designs like OA(12, 2^{11}), which was not deemed feasible.

Returning to the cast fatigue experiment in Table 3, the analysis in the original paper by Hunter et al. (1982) followed the prevailing practice and analyzed only the main effects. They found factors F and D to be significant but noted some discrepancies. First, the sign of the effect estimate of D (heat treat) was reversed. They further suggested that the cause could be due to the interaction DE and claimed that the design did not generate enough information to determine this. The aversion from estimating interactions in experiments based on this and other nonregular designs was caused by the complex aliasing property of these designs. As pointed out in Sect. 3.1 and fully described in Section 9.1 of WH, any main effect in this design has a correlation $1/3$ or $-1/3$ with any 2fi’s that do not involve the main effect. In the worst case of 11 factors, each main effect has 45 such partial aliases. As shown on page 429 of WH, the total number of models that consist of the 11 main effects and some partially aliased 2fi’s can be close to a million. The number of similar models for the 20-run Plackett–Burman design is much higher, i.e., in the millions or tens of millions. Therefore Hamada and Wu (1992) referred to this phenomenon as *complex aliasing*. This complexity was viewed negatively as “hazards” (Daniel 1976). Before the 1992 paper, experiments with complex aliasing were used mainly for the screening purpose, i.e., for estimating main effects only.

Hamada and Wu (1992) employed the principles of effect hierarchy, sparsity and heredity to justify their new analysis method. (See Section 4.6 of WH and some historical notes in Wu (2015).) Effect hierarchy and sparsity suggest that only a few main effects and even fewer 2fi’s are relatively important. Invoking them leads to a smaller number of effect terms in the models and a much smaller number of models in the model search. Thus the model complexity is reduced. A simple example is given in Section 9.3 of WH to illustrate how the model search and estimation is greatly simplified. The (frequentist) data analysis strategy of Hamada and Wu (1992) used effect heredity in the model search. It allows the search to rule out many incompatible models. Later a Bayesian version was developed by Chipman et al. (1997). Both versions can be found in Chapter 9 of WH. This strategy as applied to the cast fatigue data identified F and FG as significant. The R^2 value is increased from 0.45 (for F only) to 0.9. More importantly, the model consisting of F and FG and the effect estimates provide a better explanation or resolution of the two problems discussed above. See pages 430–431 of WH. Since the 1992 work, many practical experiments based on the 12-, 18- and 36-run designs perform data analysis by considering interactions and obtain successful results.

Underlying the success of using these nonregular designs to estimate interactions is the partial aliasing property of these designs. Noting that partially aliased effects are non-orthogonal to each other, the success of the analysis strategy is due to the exploitation of non-orthogonality. Unlike the CME analysis in Sect. 2.1 which uses reparametrization and then exploits the non-orthogonality of the reparametrized effects, the analysis here does not need to perform reparamterization because non-orthogonality is inherent in the nonregular designs.

This line of research is related to some later work in design theory. The success of the Hamada-Wu analysis strategy suggested that designs like OA$(12, 2^{11})$, OA$(20, 2^{19})$, OA$(18, 2^1 3^7)$ and OA$(36, 2^{11} 3^{12})$ possess some projective properties not studied before. Wang and Wu (1995) coined the term "hidden projection" to describe such properties. See also the work by Lin and Draper (1992), and Box and Tyssedal (1996). These papers collectively have inspired the subsequent work to extend the minimum aberration criterion from regular to nonregular designs. The *minimum aberration criterion* has been the most commonly used and powerful tool for the optimal selections of regular 2^{k-q} and 3^{k-q} designs. The entire Chapters 5–6 of WH are devoted to the study and tabulations of minimum aberration designs. A nearly complete list of two-level minimum aberration designs can be found in Chapter 7 of "JMP 12 Design of Experiments Guide" (SAS Institute 2015). Encouraged by the increasing use of nonregular designs in practice, design researchers started to look for extensions of the minimum aberration criterion to nonregular designs. Among them, two major criteria are the *generalized minimum aberration* criterion (Tang and Deng 1999; Deng and Tang 1999; Xu and Wu 2001) and the minimum moment criterion (Xu 2003). A survey of these advances can be found in Cheng (2014).

4 3^{k-q} designs: design classification and analysis

In this section I will use the same framework to reexamine the class of 3^{k-q} designs. Recall that these designs (and the more general p^{k-q} designs for a prime power p) are defined by their defining contrast subgroup and that their construction and algebraic properties are based on standard tools like Galois field and finite geometry. They should thus be treated as regular designs. While this is the standard approach to such designs as depicted in design texts, the same designs with a new reparametrization can be treated as nonregular, which is novel and somewhat surprising. I will address these two aspects of the 3^{k-q} designs in the following subsections.

4.1 Regular designs: orthogonal components system and ANOVA

To explain the regular nature of the 3^{k-q} designs, we use the following simple example for illustration. Consider the 3^{4-1} design used in the seat-belt experiment in Table 6.2 of WH. It has 27 runs to study four factors A, B, C, D, each with three levels. It is a $1/3$ fraction of the full factorial 3^4 design. The fraction is defined by defining column D in the table as the sum of the first three columns representing A, B, C. Let x_i, $i = 1, 2, 3, 4$, represent the four factors, which take values 0, 1, 2 modulus 3. Then the relationship among the four factors (i.e., columns) is defined by $x_4 = x_1 + x_2 + x_3$.

Notationally this relationship is denoted by $D = ABC$ or equivalently its defining contrast subgroup is given by $I = ABCD^2$. This subgroup has 13 cosets, which can be found in (6.9) of WH. Only one is shown here: $A = BCD^2 = AB^2C^2D$, which says that the three factorial effects in the identity equation are fully aliased. The 13 cosets account for all the 26 degrees of freedom for the 27-run 3^{4-1} design because each coset has two degrees of freedom. In the finite geometry setting, each coset is a two-dimensional subspace and the 13 such subspaces are mutually orthogonal to each other due to the nature of its algebraic construction. Furthermore, the product of any two factorial effects in the 13 cosets can be found among the cosets. In this sense the designs are called regular but neither is this nor the one in Sect. 3.1 for the 2^{k-q} designs a rigorous definition.

Using finite geometry Raktoe et al. (1981) rigorously define what constitutes regular p^{k-q} designs for any prime power p. Briefly, it should satisfy the following condition: the collection of treatment combinations in a p^{k-q} design should be a subspace or coset of subspace in the finite geometry over $GF(p)$.They call the design "irregular" if the collection does not form a subspace or coset of subspace. Detailed discussions can be found in Section 9.5 of the book. See also Mukerjee and Wu (2006) and especially Chapter 9 of Cheng (2014).

Experiments based on regular 3^{k-q} designs are typically analyzed by using the analysis of variance (ANOVA). For the 3^{4-1} design the ANOVA decomposition is based on the 13 cosets, i.e., its total SS (sum of squares) is decomposed into 13 SS terms, each representing the factorial effects in one coset. These 13 terms are uncorrelated with each other because their corresponding subspaces are orthogonal to each other. Therefore Wu and Hamada (2000, 2009) refer to this parametrization as the *orthogonal components system*. For the seat-belt experiment, the ANOVA table is given in Table 6.6 of WH. Notice that the interaction between factors A and B, denoted as $A \times B$, has four degrees of freedom and is decomposed into two interaction components denoted as AB and AB^2 and displayed in Table 6.6. The ANOVA approach is the most commonly used analysis method for 3^{k-q} designs because it corresponds to how the degrees of freedom are allocated among the cosets.

However, the ANOVA approach has two distinct disadvantages. First, for fractional factorial designs, some of the SS terms are (fully) aliased. For example, the interaction component AB in Table 6.6 of WH, which is significant, is aliased with the interaction component CD^2. Another significant interaction component AC in Table 6.6 is aliased with BD^2. These aliased terms cannot be de-aliased. Second, while the interaction $A \times B$ (with 4 df's) has a ready interpretation, its interaction component AB or AB^2 (each with 2 df's) does not usually render a meaningful interpretation. See the discussion around Table 6.4 of WH on this point. Therefore if only one of the two interaction components is significant in the ANOVA, e.g., AB but not AB^2 in Table 6.6 of WH, it can cause problem in interpreting the analysis result.

4.2 Nonregular designs: linear-quadratic system and variable selection analysis

Because of the two difficulties associated with the regular 3^{k-q} designs, Wu and Hamada (2000) first proposed a new parametrization of effects, called the *linear-*

quadratic system, to circumvent such problems. The basic idea is very simple. For a factor A with three levels denoted by 0, 1, 2, denote the observations at these three levels as y_0, y_1, y_2. Define its linear effect as $y_2 - y_0$ and its quadratic effect as $(y_2 + y_0)/2 - y_1$. Their corresponding (unnormalized) contrast vectors are denoted as $A_l = (-1, 0, 1)$ and $A_q = (1, -2, 1)$, respectively. For two factors A and B, their linear-by-linear interaction, denoted as $(AB)_{ll}$, is defined as the component-wide product of the two vectors A_l and B_l. The linear-by-quadratic, quadratic-by-linear and quadratic-by-quadratic interactions can be similarly defined and denoted as $(AB)_{lq}$, $(AB)_{ql}$, $(AB)_{qq}$. Then the four degrees of freedom for the $A \times B$ interaction can be represented by these four terms, each represented by a contrast vector. This parametrization can be extended to any number of quantitative factors with three levels. Detailed development and extension to qualitative factors can be found in Section 6.6 of WH.

As in the case of the cme reparametrization for two-level designs, the linear-quadratic (abbreviated as l-q) reparametrization creates non-orthogonality among effects. And non-orthogonality provides the opportunity for aliased effects to be de-aliased. For illustration, consider the 3^{3-1} design for factors A, B, C with $C = AB$. The three main effects have six degrees of freedom: $A_l, A_q, B_l, B_q, C_l, C_q$, which are mutually orthogonal. Using the l-q system, we can define 12 ($= 4 \times 3$) interaction effect terms among the three factors. But there are only two degrees of freedom left for interactions. Between main effect and interaction or between two interactions, there ought to be many non-orthogonal pairs. For example, suppose only $(AB)_{ll}$ and $(AB)_{lq}$ are considered for the remaining two degrees of freedom. It can be shown that each of the two interaction effects is correlated with the main effects C_l and C_q. See (6.15) of WH. For a general treatment of the full or partial aliasing relationships between main effects and interaction effects in the l-q system, see Sabbaghi et al. (2014), which uses the tool of indicator functions to tackle this problem.

We take the liberty of calling a 3^{k-q} design equipped with the l-q system a "non-regular" design, but this would require a broader definition of regular or irregular design than the one given in Sect. 4.1. Recall that the definition given there is based on the finite geometry structure. Since each effect in this geometry has two degrees of freedom, it is not applicable to the l-q system, which consists of effects of one degree of freedom. Our inspiration or justification comes from the approach taken by Wu and Hamada (2000) as discussed near the end of Sect. 3.1. They define "regular" in the case of two-level designs as equivalent to the fact that any pair of factorial effects have absolute correlation either 0 or 1. An extension of this approach to 3^{k-q} designs would not be trivial. For example, one may consider extending the work of Ye (2004) to cover 3^{k-q} designs.

Next we discuss the implications in estimation. First, the non-orthogonality among some effects in the l-q system provides the basis for effect estimability in the system. Similar in spirit (but not in details) to the CME analysis in Sect. 2, a variable selection strategy was developed in Section 6.6 of WH. Briefly, it works as follows. First, list the candidate set of main effects and two-factor interaction effects in the l-q system. Note that each such effect has only one degree of freedom, which is amenable to the use of any reasonable variable selection method. In WH, only the stepwise regression or subset selection procedure is used but more modern variable selection methods (Hastie et al. 2009) like Lasso can also be applied. In selecting models, the effect heredity principle is invoked to rule out incompatible models. This was similarly done for nonregular 2^{k-q} designs in Sect. 3.3. When this strategy was applied to the seat-belt data as in Section 6.6 of WH, it led to very good results. The final model includes the following main effects: A, D (each with two degrees of freedom) and one component B_l and C_l of factors B and C. More interestingly, it identifies the following interaction effects: $(AB)_{ql}$, $(AC)_{ll}$, and $(CD)_{l,12}$. Note that factor D in the experiment is a qualitative factor. By comparison, the ANOVA analysis in Table 6.6 of WH identifies A, B, D and the AC and AB interaction components (in the orthogonal components system) as significant. These findings have the two shortcomings of the ANOVA approach as described in Sect. 4.1. First AC is (fully) aliased with BD^2, and AB is aliased with CD^2. These aliased effects cannot be disentangled. Second, even if one of the aliased effects is chosen into the model, it does not render a good interpretation for users of the model. Comparisons of the two approaches on the seat-belt data clearly show the advantages of the l-q parametrization system.

Finally, the success of the l-q system sheds some new light on the choice of designs. In the traditional practice as seen in design texts, use of 3^{k-q} designs with resolution III or IV is discouraged because some interactions or interaction components in these designs are aliased and thus are not estimable. Instead 3^{k-q} designs with resolution V are recommended because their two-factor interactions are estimable. These designs are, however, quite large and expensive. The results given above show that variable selection based on the l-q system can be applied to designs with resolution IV or even III, which are smaller and more economical than V designs. Therefore this traditional wisdom is conservative and somewhat misguided. It also provides another illustration that the resolution criterion is too coarse for classifying and ranking the capabilities of regular fractional factorial designs.

5 A historical perspective

This paper is structured according to the nature or complexity of designs. We start with regular 2^{k-q} designs, move to nonregular 2^{k-q} designs and end with 3^{k-q} designs. But the ideas were discovered not necessarily in the same order. In this section I will give a historical perspective on how I and/or coauthors came up with the ideas.

In the mid 80's, Taguchi introduced to the west nonregular two-level designs like L_{18} and L_{36} for parameter design experiments. Since then, I had become interested in and intrigued by these designs. First these designs are economical in run size and can accommodate a flexible combination of factor levels. A natural question to ask was whether interactions should or can be entertained in these designs. Researchers in the traditional camp said no by referring to the complex aliasing of the L_{12} design (see the discussion in Sect. 3.3). Taguchi also said no because he did not advocate the inclusion of interactions in his analysis. His rationale was that a good robust parameter design experiment would have no need to estimate interactions. Robustness would take care of the effects of interactions. See Taguchi (1987) and much earlier references in Japanese. So I was confused by the various opinions at that time. In 1986 a group of researchers from U. of Wisconsin and AT&T Bell Labs organized a delegation headed by George Box to visit Japan to understand its practice in quality engineering [see Box et al. (1988)]. Then the revelation moment came. I recall it was a hot summer afternoon in Nagoya when we were attending presentations made by members of the Central Japan Quality Association. All case studies employed Taguchi's idiosyncratic L_{18} or L_{36}. The case studies were successful in getting new insights and achieving variation reduction. Yet their analyses did not consider interactions as they were taught so by Taguchi. I suddenly realized that these designs must have some intrinsic theoretical properties that relate to the estimability of interactions. After I returned to Madison, I mentioned this thought to Mike Hamada, who was writing his thesis. At that time I already had the notion of full and partial aliasing and was aware of the complex aliasing of L_{12}, the 12-run Plackett–Burman design. So we made some progress but were stymied by the many models that came from the analysis. So we did not pursue this further. Mike defended his thesis on another topic in 1987. Then we both moved to U. of Waterloo to work as a team on quality engineering and design of experiments. Sometime in 1990, we hit the subject again and Mike reminded me about the many models he found in the analysis. Then we suddenly hit the notion of effect heredity and their use in reducing the number of models in model search. Mike said he would try the analysis using this new concept. When I met him the next day, I saw a big smile in his face. The analysis results became much more clean and definitive. It was our eureka moment. The manuscript was written in 1991 and eventually appeared as Hamada and Wu (1992).

The discovery of CME followed a longer and more winding path. Soon after I started teaching design of experiments in the early 80's in Madison, using the now classic text by Box et al. (1978), I realized that two-factor interaction can be defined through the concept of conditional main effects (cme's) but it did not go very far. Design was not my major research field at that time but the notion of cme's was on the back of my mind. In 1988 I moved to Waterloo to assume a chair professorship on quality improvement. Design of experiments became my research focus. During my first term there, I used data from a "car marriage station simulation experiment" at GM Canada to test the idea of using cme's in analyzing data from a resolution IV design. I got some encouraging preliminary results but my idea at that time was still primitive. I guess I did not push any further because the time was not ripe and related supporting ideas had not yet been developed. Even if I wanted to publish such work in the 1980's, it would have been rejected because the concept of effect aliasing was well entrenched. Any attempt to de-alias fully aliased effects without adding runs would have been viewed as lunatic. So this idea had remained dormant and been shelved for about 21 years until 2010 when I was invited to give the Fisher Lecture at the Joint Statistical Meetings. The timing was ripe because the methods for estimating interactions in nonregular two-level designs and in three-level designs using the linear-quadratic (l-q) system have been laid out in the two editions of our book Wu and Hamada (2000, 2009). I also knew that the new materials on cme's would not be rejected by JASA because it would be part of the Fisher Lecture. This led to the paper Wu (2015), which contained the Canadian data on car marriage station simulation experiment.

The discovery of the l-q system and its use in making a 3^{k-q} design into a nonregular design followed a more logical and natural path. After the Hamada-Wu success in analyzing interactions in nonregular two-level designs, it was the natural next question to ask whether interactions can be entertained in a three-level design. Because each three-level factor has two degrees of freedom instead of one (for the two-level case), there is more room to maneuver. Use of linear and quadratic effects for three-level factors has been known in ANOVA and regression analysis. However, their use in the context of de-aliasing aliased effects in 3^{k-q} designs was novel. When Mike Hamada and I realized this possibility, we were already in the middle of finishing Wu and Hamada (2000). We were busy getting these new materials into the chapter on three-level designs. There was no time to write it up as a separate paper for journal publication. Therefore these materials remain unknown even to researchers in design unless they would delve deeply into the book.

To summarize my experience in this line of research, novel ideas are rarely developed in a logical order and a straight manner. Simple looking ideas may come at a later time than more complex ones. A good example is the development of the CME analysis for analyzing very simple 2^{k-q} designs with resolution IV or III. Despite the simplicity of the design, it was developed much later than its two cousins. Discovery may depend on luck, serendipity, and the applicational environment of the time. One can never know but must forge ahead.

6 Concluding remarks

Our approach to de-aliasing consists of two key concepts: first we *reparametrize* an appropriate space of effects, which induces non-orthogonality among effects; then we exploit this *non-orthogonality* to enable the estimation of effects not considered possible before. For regular 2^{k-q} designs, we use the conditional main effects for reparametrization; for 3^{k-q} designs we use the linear-quadratic system for reparametrization. In the case of nonregular two-level designs, there is no need of reparametrization because non-orthogonality is inherent in these designs.

In Sects. 3 and 4, we employ the principles of hierarchy, sparsity, and heredity to motivate or justify the analysis strategies. Note that these principles govern the relationships among factorial effects, which were first summarized in Wu and Hamada (2000, 2009) for factorial designs. Some historical notes on these principles can be found in Wu (2015). However, these principles are not applicable to the CME analysis in Sect. 2 because the conditional main effects (cme's) do not fit into the framework for these principles, which deal with the traditional factorial effects like main effects and interactions of various orders. One exception to this statement is the analogy of cme coupling and cme reduction to effect heredity and effect hierarchy as discussed in Sect. 2.2. A challenge is to develop a design-theoretic framework for the cme's that plays a similar role to the minimum aberration criterion for regular p^{k-q} designs for prime power p. The only relevant work to my knowledge is the paper by Mukerjee et al. (2017), which tries to develop one such theory, albeit in some limited situations. Discussions on the minimum aberration criterion can be found in Mukerjee and Wu (2006) and Cheng (2014).

Although our work was originally motivated by the attempt to de-alias aliased effects in designed experiments, it has applications in broader settings. In Sect. 2.2, we outline some ongoing work that extends the CME analysis to general observational data with input factors at two levels. This bi-level variable selection strategy uses the cme's as the basis functions in variable search. It should find broader applications than the original CME analysis. Another example, not reported here, is the extension of the effect heredity principle to general variable selection. It enables a more efficient search for best models in variable selection through the use of optimization techniques. See Yuan et al. (2009) for details. Finally, one may also argue that the collection of work as reported here can serve as a transition from orthogonal experiments to non-orthogonal experiments or studies such as optimal designs and observational studies. How this will pan out is a big unknown.

Acknowledgements Research supported by ARO W911NF-17-1-0007. The author is grateful to the referees for helpful comments.

References

Addelman, S. (1961). Irregular fractions of the 2^n factorial experiments. *Technometrics*, *3*(4), 479–496.
Addelman, S. (1962). Augmenting factorial plans to accommodate additional two-level factors. *Biometrics*, *18*(3), 308–322.
Barker, T. B., Milivojevich, A. (2016). *Quality by experimental design* (4th ed.). New York: CRC Press.
Box, G. E. P., Hunter, J. S. (1961). The 2^{k-p} fractional factorial designs Part I. *Technometrics*, *3*(3), 311–351.
Box, G. E. P., Tyssedal, J. (1996). Projective properties of certain orthogonal arrays. *Biometrika*, *83*(4), 950–955.
Box, G. E. P., Hunter, W. G., Hunter, J. S. (1978). *Statistics for experimenters: An introduction to design, data analysis, and model building*. New York: Wiley.
Box, G. E. P., Kackar, R., Nair, V., Phadke, M., Shoemaker, A. C., Wu, C. F. J. (1988). Quality practices in Japan. *Quality Progress*, *21*(3), 37–41.
Box, G. E. P., Hunter, J. S., Hunter, W. G. (2005). *Statistics for experimenters: Design, innovation, and discovery* (2nd ed.). New York: Wiley.
Cheng, C.-S. (2014). *Theory of factorial design: Single- and multi-stratum experiments*. New York: CRC Press.
Chipman, H., Hamada, M., Wu, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics*, *39*(4), 372–381.
Cox, D. R. (1984). Interaction. *International Statistical Review*, *52*(1), 1–24.
Daniel, C. (1976). *Applications of statistics to industrial experiments*. New York: Wiley.
Deng, L.-Y., Tang, B. (1999). Generalized resolution and minimum aberration criteria for Plackett–Burman and other nonregular factorial designs. *Statistica Sinica*, *9*(4), 1071–1082.
Finney, D. (1945). The fractional replication of factorial arrangements. *Annals of Eugenics*, *12*, 291–303.
Fisher, R. A. (1971). *The design of experiments* (7th ed.). London: Oliver and Boyd.
Hamada, M., Wu, C. F. J. (1992). Analysis of designed experiments with complex aliasing. *Journal of Quality Technology*, *24*(3), 130–137.
Hastie, T., Tibshirani, R., Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York: Springer.
Hedayat, A. S., Sloane, N. J. A., Stufken, J. (1999). *Orthogonal arrays: Theory and applications*. New York: Springer.
Hunter, G. B., Hodi, F. S., Eagar, T. W. (1982). High-cycle fatigue of weld repaired cast Ti–6AI–4V. *Metallurgical Transactions A*, *13*(9), 1589–1594.
John, P. W. M. (1971). *Statistical design and analysis of experiments*. New York: Macmillan.
Lin, D. K. J., Draper, N. R. (1992). Projection properties of Plackett and Burman designs. *Technometrics*, *34*(4), 423–428.
Mak, S., Wu, C. F. J. (2017). `cmenet`: A new method for bi-level variable selection of conditional main effects. *Journal of the American Statistical Association*. https://arxiv.org/abs/1701.05547 **(submitted)**.
Masuyama, M. (1957). On difference sets for constructing orthogonal arrays of index two and of strength two. *Report on Statistical Application and Research, JUSE*, *5*, 27–34.

Mazumder, R., Friedman, J. H., Hastie, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, *106*(495), 1125–1138.

Mukerjee, R., Wu, C. F. J. (2006). *A modern theory of factorial design*. New York: Springer.

Mukerjee, R., Wu, C. F. J., Chang, M.-C. (2017). Two-level minimum aberration designs under a conditional model with a pair of conditional and conditioned factors. *Statistica Sinica*, *27*(3), 997–1016.

Plackett, R. L., Burman, J. P. (1946). The design of optimum multifactorial experiments. *Biometrika*, *33*(4), 305–325.

Raktoe, B., Hedayat, A., Federer, W. T. (1981). *Factorial designs*. New York: Wiley.

Sabbaghi, A., Dasgupta, T., Wu, C. F. J. (2014). Indicator functions and the algebra of the linear-quadratic parametrization. *Biometrika*, *101*(2), 351–363.

SAS Institute Inc. (2015). *JMP®12 design of experiments guide*. Cary, NC: SAS Institute Inc.

Seiden, E. (1954). On the problem of construction of orthogonal arrays. *The Annals of Mathematical Statistics*, *25*(1), 151–156.

Su, H., Wu, C. F. J. (2017). CME analysis: A new method for unraveling aliased effects in two-level fractional factorial experiments. *Journal of Quality Technology*, *49*(1), 1–10.

Sun, D. X., Wu, C. F. J. (1993). Statistical properties of hadamard matrices of order 16. In W. Kuo (Ed.), *Quality through engineering design* (pp. 169–179). New York: Elsevier.

Taguchi, G. (1987). *System of experimental design*. Whilt Plains, NY: Unipub/Kraus International Publications.

Tang, B., Deng, L.-Y. (1999). Minimum g_2-aberration for nonregular fractional factorial designs. *Annals of Statistics*, *27*(6), 1914–1926.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, *58*(1), 267–288.

Wang, J.-C., Wu, C. F. J. (1992). Nearly orthogonal arrays with mixed levels and small runs. *Technometrics*, *34*(4), 409–422.

Wang, J. C., Wu, C. F. J. (1995). A hidden projection property of Plackett–Burman and related designs. *Statistica Sinica*, *5*(1), 235–250.

Wu, C. F. J. (2015). Post-Fisherian experimentation: From physical to virtual. *Journal of the American Statistical Association*, *110*(510), 612–620.

Wu, C. F. J., Hamada, M. S. (2000). *Experiments: Planning, analysis and parameter design optimization*. New York: Wiley.

Wu, C. F. J., Hamada, M. S. (2009). *Experiments: Planning, analysis, and optimization* (2nd ed.). New York: Wiley.

Xu, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statistica Sinica*, *13*(3), 691–708.

Xu, H., Wu, C. F. J. (2001). Generalized minimum aberration for asymmetrical fractional factorial designs. *Annals of Statistics*, *29*(2), 549–560.

Ye, K. Q. (2004). A note on regular fractional factorial designs. *Statistica Sinica*, *14*(4), 1069–1074.

Yuan, M., Joseph, V. R., Zou, H. (2009). Structured variable selection and estimation. *The Annals of Applied Statistics*, *3*(4), 1738–1757.