

Design and Analysis of Two-Level Factorial Experiments With Partial Replication

Chen-Tuo LIAO

Division of Biometry
Institute of Agronomy
National Taiwan University
Taipei, Taiwan
(ctliao@ntu.edu.tw)

Feng-Shun CHAI

Institute of Statistical Science
Academia Sinica
Taipei, Taiwan
(fschai@stat.sinica.edu.tw)

In a two-level factorial experiment, we consider construction of parallel-flats designs with two identical parallel flats that allow estimation of a set of specified possibly active effects and the pure error variance. A set of sufficient conditions is presented for the designs to be D-optimal for the specified effects, assuming that the other effects are negligible, over the class of competing parallel-flats designs. In addition, an algorithm is developed to generate the D-optimal designs with a choice of flexible degrees of freedom for the pure error variance. Because the proposed partially replicated designs are highly efficient in estimating the possibly active effects and provide a replication-based estimate of the error variance, they provide a practical compromise between the power in identifying truly active effects and the number of runs in experiments. This property is verified through a simulation study.

KEY WORDS: Active effect; Parallel-flats design; Pure error; Unreplicated design.

1. INTRODUCTION

During the initial stages of experimentation, unreplicated two-level factorial designs are commonly used to identify important or active effects (Box and Meyer 1986), due mainly to their cost-effectiveness. But statistical inference based on such unreplicated experiments typically is a challenge, because there are no pure replicates for estimating the experimental error variance. Although several useful methods of identifying active effects using unreplicated data are available (Hamada and Balakrishnan 1998), this would be more reliable if the variance of experimental error could be estimated based on pure replicates. Obviously, the fully replicated design is a simple resolution to provide pure replicates, but it often leads to a costly experiment. Thus designs with partial duplication may provide promising alternatives, not only offering pure replicates, but also providing cost savings. We use a real-data experiment to motivate the current study.

Snee (1985) presented a 2^{7-3} fractional factorial design to identify the important effects in the viscosity measurement of a high-volume product. The 16-run regular design was determined by the *generators* $5 = 234$, $6 = 134$, and $7 = 123$. More interestingly, four additional runs of the design were duplicated to estimate the experimental error variance. (See Snee 1985 for details regarding the experimental factors, their levels, treatment combinations of the design and experimental data.) To illustrate the goals of our study, we first analyze the unreplicated 16-run data using the method of Lenth (1989). The pseudo-standard error estimate (PSE) is given by 57.57. Then the active effects are determined using a margin error (ME) of $ME = 2.156 \times PSE$, where 2.156 is the critical value of Lenth's method for controlling the individual error rate (IER) at 0.05 (Ye and Hamada 2000). According to the rule that the active effects are those with the absolute value of effect estimates larger than the ME, we can identify five active effects: main effects for factors 2, 3, 4, and 6 and a two-factor interaction between factors 3 and 6. Because of the imbalance of the partially replicated data, we analyze these using least squares

regression according to the saturated effects model, including the constant term μ and all 15 possible estimable effects (actually linear combinations of aliased effects). The Student *t*-test with 4 degrees of freedom (df) based on the estimate of pure error variance is used to determine active effects. Note that the estimated effect standard errors for the partially replicated design are all equal to 26.21, which is smaller than the PSE. At a significance level of 0.05, we identify the five active effects declared by Lenth's method and find one more active main effect, for factor 5. These findings indicate that the identified active effects can differ when different designs are used.

This example raises several interesting issues. As Snee (1985) pointed out, no particular criteria were used in selecting the duplicated points other than the desire to replicate uniformly over the design space. Thus the first goal of the present study was to develop a suitable criterion for determining the partial replication. The second goal was to investigate the situations in which the partially replicated designs are useful in practice. We return to this example in Section 3 when discussing the D-optimality of the partially replicated design.

Constructing the most powerful designs for identifying active effects can be very difficult, because the power of a design may depend on several unknown factors, such as the effect sizes of truly active effects, the number of truly active effects, the true value of experimental error variance, and others. For industrial applications, minimum aberration two-level fractional factorial designs often are used to search for active effects in a screening experiment, based mainly on some practical principles, such as the hierarchical ordering principle and effect sparsity principle (Wu and Hamada 2000). The power of such designs in identification of truly active effects has not been thoroughly investigated, however. In some practical experiments, some information on which effects are likely to be active may be available,

allowing one to take advantage of this information to design an appropriate plan, not necessarily a minimum aberration design, to ensure estimation of the specified possibly active effects with high efficiency (see, e.g., Franklin and Bailey 1977; Wu and Chen 1992; Liao, Iyer, and Vecchia 1996). Instead of finding the most powerful designs, we proceed with construction of the partially replicated designs so as to meet the requirements that the error variance be estimated using pure replicates and the possibly active effects be estimated with high efficiency. It would be reasonable for the desired designs to be useful in identifying truly active effects. In addition, as discussed previously by Youden and Hunter (1955) and Pigeon and McAllister (1989), the partially replicated designs can be used to test those effects that are believed to be negligible against the pure error, providing a lack-of-fit test for the specified model.

The design issues involving factorial experiments with partial replication are rarely explored in the literature. Some scattered results have been given by Dykstra (1959), Snee (1985), and Pigeon and McAllister (1989). Most recently, Liao and Chai (2004) recognized the partially replicated two-level factorial design presented by Pigeon and McAllister (1989) as one composed of three different fractions, belonging to a family of regular 2^{n-k} designs with the same defining relations, and a duplicate of one of the three fractions. In other words, it is a parallel-flats design (PFD) with two identical flats. Liao and Chai focused on only three-flat and four-flat designs and presented two classes of designs with run sizes $N = 12, 16, 24,$ and 32 and 4 or 8 df for the pure error variance. In this article we discuss a more general situation in which users can have a choice of various df, which are 1 or a power of 2 up to one-half of the run size of a regular 2^{n-k} design. Most interestingly, we show that the resulting designs are D-optimal among all possible competing PFDs with two identical flats. We evaluate the proposed designs for identifying truly active effects through a simulation study.

The rest of this article is organized as follows. In Section 2 we discuss sufficient conditions for a design to be D-optimal for a set of specified possibly active effects over the class of PFDs with two identical flats. Based on this, in Section 3 we present an algorithm for generating the desired designs. In Section 4 we report a simulation study conducted to investigate the performance of the proposed partially replicated designs, unreplicated designs, and fully replicated designs in identifying truly active effects. Finally, we give some concluding remarks in Section 5.

2. PARALLEL-FLATS DESIGNS WITH PARTIAL REPLICATION

Let β denote the vector of the possibly active effects that are not assumed to be zero. The corresponding linear model for the observations from an experiment using a design d may be written in the form

$$y = X\beta + \epsilon, \tag{1}$$

where the model matrix X depends on β and d . The vector ϵ consists of noise random variables, which are assumed to be pairwise uncorrelated with common mean 0 and variance σ^2 .

A full two-level factorial design with n factors contains all possible 2^n level combinations, and a regular 2^{n-k} design contains a subset of these runs such that they satisfy k defining

relations, expressed as independent linear equations over *Galois Field* of order 2, GF[2]. Using finite geometry terminology, such designs have been called *single-flat designs*. Thus a single-flat 2^{n-k} design d_i , defined by *alias matrix* A and *coset indicator vector* c_i , consists of all treatment combinations $t = [t_1, t_2, \dots, t_n]'$ satisfying the equation $At = c_i$ over GF[2], where A is a $k \times n$ matrix of rank k and c_i is any $k \times 1$ vector over GF[2]. Meanwhile, if we consider the union of several single-flat designs d_1, d_2, \dots, d_f , which are defined by the same alias matrix A and coset indicator vectors c_1, c_2, \dots, c_f . This forms an f parallel-flats design, abbreviated as f -PFD. Clearly, an f -PFD consists of $N = f \times 2^{n-k}$ runs. Let C be the $k \times f$ *coset indicator matrix* with columns c_1, c_2, \dots, c_f . Thus an f -PFD is determined by the pair of matrixes (A, C) . Specifically, if there are exactly two identical c_i among the f coset indicator vectors, these parallel-flats designs with partial replication are designated f -PFDRs by Liao and Chai (2004). Thus an f -PFDR contains f flats, one of which is replicated twice.

Suppose that d is an f -PFDR determined by the two matrixes (A, C) and that the linear model in (1) holds. The *defining vector* of a factorial effect is defined by an n -tuple whose entries are 1 or 0 over GF[2] depending on whether or not the corresponding factors involve the factorial effect. Again, from Liao and Chai (2004), let e_1 and e_2 be the two defining vectors of two different elements in β . Then their corresponding element in the information matrix $M = X'X$ using d is given by

$$m(e_1, e_2) = \left[\sum_{i=1}^f (-1)^{(e_1+e_2)z_i} \right] \left[\sum_v (-1)^{(e_1+e_2)Bv} \right], \tag{2}$$

where the exponents of -1 in both factors are computed using arithmetic modulo 2, but the sums of the powers of -1 are not performed modulo 2. In addition, B is an $n \times (n - k)$ matrix of rank $n - k$ such that $AB = 0$, where 0 is a matrix with all elements equal to 0 over GF[2]; z_i is a solution to the equation $At = c_i$, for $i = 1, 2, \dots, f$; and v ranges over all possible vectors of length $(n - k)$ over GF[2]. Moreover, the information matrix can always be expressed as a block diagonal matrix given by

$$M = \begin{bmatrix} M_0 & 0 & \cdots & 0 \\ 0 & M_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_g \end{bmatrix}, \tag{3}$$

where each submatrix M_j corresponds to an alias set, $j = 0, 1, 2, \dots, g$, and all of the off-diagonal submatrixes are zero matrixes. Here M_0 corresponds to the alias set containing the constant term μ and $g + 1$ is the number of distinct alias sets defined by β and A . Liao and Chai (2004) have provided more details and an illustrative example involving (2) and (3).

In the current study, we considered the problem of finding the D-optimal design for β over the class of all possible f -PFDRs. By using the D-optimal design, the generalized variance of the best linear unbiased estimators (BLUEs) of β attains the minimum among the competing designs. Obviously, 2-PFDRs are fully replicated regular 2^{n-k} designs and have 2^{n-k} df for estimating the pure error variance. Moreover, a 2-PFDR is an orthogonal design, causing it to be D-optimal among all possible two-level designs with $N = 2 \times 2^{n-k} = 2^{n-k+1}$. Based on (2) and (3), we present the following theorem for an f -PFDR with $f > 2$.

Theorem 1. For given a set of specified possibly active effects β with v elements, let \mathcal{D} be the collection of nonsingular f -PFDRs for β . In addition, let design $d^* \in \mathcal{D}$ determined by the two matrixes $(\mathbf{A}^*, \mathbf{C}^*)$, where $\mathbf{C}^* = [\mathbf{C}_0^*, \mathbf{c}^*]$ and \mathbf{C}_0^* consists of the distinct $f - 1$ coset indicator vectors and \mathbf{c}^* is any column of \mathbf{C}_0^* . Then d^* is D-optimal for β over \mathcal{D} if the following two conditions are satisfied:

(a) The $(f - 1)$ -PFD determined by the pair of matrixes $(\mathbf{A}^*, \mathbf{C}_0^*)$ is orthogonal for β .

(b) Let v_j^* denote the number of elements in alias set j determined by \mathbf{A}^* , for $j = 0, 1, \dots, m$, where $m = 2^{n-k} - 1$. Note that $v_j^* = 0$ is allowable. Then $v_0^*, v_1^*, \dots, v_m^*$ attain the maximum of $\prod_{j=0}^m (f - 1 + v_j)$ subject to $\sum_{j=0}^m v_j = v$.

A proof of this theorem is given in the [Appendix](#). Note that a theorem similar to Theorem 1 was presented by Li (1991), but her theorem seems to be incomplete, because it excludes condition (b). Furthermore, among the v_j 's, if there are only two distinct values with a difference of 1, then condition (b) of Theorem 1 can be satisfied (see Cheng 1978, lemma 3.1). Thus we have the following corollary.

Corollary 1. Under the same assumption as in Theorem 1, d^* is D-optimal for β over \mathcal{D} if it satisfies condition (a) and the condition that there are $v - 2^{n-k} \lfloor v/2^{n-k} \rfloor$ of the v_j^* 's equal to $\lfloor v/2^{n-k} \rfloor + 1$, and the other v_j^* 's equal to $\lfloor v/2^{n-k} \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x .

Note that we can verify that the 20-run partially replicated design presented by Snee (1985) is actually a 5-PFDR determined by

$$\mathbf{A}(5 \times 7) = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{C}(5 \times 5) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

By Corollary 1, this 5-PFDR can be easily verified to be D-optimal for the saturated effects model over the competing parallel-flats designs. This does not hold for some other β , however; for example, it is not D-optimal for $\beta = \{\mu, 1, 2, 3, 4, 5, 6, 7, 12, 13\}$ or $\beta = \{\mu, 1, 2, 3, 4, 5, 6, 7, 12, 34, 56\}$. But D-optimal partially replicated designs exist for these two β , as given in Tables 1 and 2 of Section 3.

3. AN ALGORITHM

Before presenting an algorithm for generating the D-optimal designs described in Corollary 1, we give the following example to illustrate some connections between alias sets determined by \mathbf{A} and the specification of β .

Example 1. Suppose that $\beta = \{\mu, 1, 2, 3, 4, 5, 6, 12, 13, 14, 15, 16\}$ and $v = 12$. The regular 2^{6-2} design determined by the

defining relations $I = 1236 = 1345 = 2456$ can be used to estimate β , assuming that the other effects not in the span are negligible. The alias sets determined by the design are given by

- $G_0: I = 1236 = 1345 = 2456,$
- $G_1: 1 = 236 = 345 = 12456,$
- $G_2: 2 = 136 = 12345 = 456,$
- $G_3: 3 = 126 = 145 = 23456,$
- $G_4: 4 = 12346 = 135 = 256,$
- $G_5: 5 = 12356 = 134 = 246,$
- $G_6: 6 = 123 = 13456 = 245,$
- $G_7: 12 = 36 = 2345 = 1456,$
- $G_8: 13 = 26 = 45 = 123456,$
- $G_9: 14 = 2346 = 35 = 1256,$
- $G_{10}: 15 = 2356 = 34 = 1246,$
- $G_{11}: 16 = 23 = 3456 = 1245,$
- $G_{12}: 24 = 1346 = 1235 = 56,$
- $G_{13}: 25 = 1356 = 1234 = 46,$
- $G_{14}: 124 = 346 = 235 = 156,$
- $G_{15}: 125 = 356 = 234 = 146.$

A total of 16 alias sets are defined, each consisting of 4 factorial effects. Each of the first 12 alias sets contains exactly 1 element of β , but none of the remaining 4 alias sets contains any element of β . A single-flat design is orthogonal for β if and only if each of the 2^{n-k} alias sets determined by \mathbf{A} contains at most one element of β . Another important property is that the 2^{n-k} alias sets together form an Abelian group under the binary product of all elements of one alias set by one element of another alias set (Raktoe, Hedayat, and Federer 1981, pp. 194–203); for instance, $G_i * G_0 = G_i$ for all i , $G_1 * G_2 = G_7$ and $G_3 * G_4 = G_{10}$ in the foregoing example, where $*$ denotes the operation and G_0 is the identity element.

Now Corollary 1 can be applied to develop the following forward selection procedures for generating D-optimal designs for any specified β with 2^p df for the pure error variance, $p = 0, 1, 2, \dots, n - k - 1$:

Step 1. For given β with v possibly active effects, obtain a regular 2^{n-k} design (single-flat design) that is orthogonal for β using Franklin–Bailey (1977) algorithm. The alias matrix of the resulting design is denoted by \mathbf{A} . Note that Franklin–Bailey algorithm is capable of generating all possible such orthogonal single-flat designs for β .

Step 2. Initialize $i = 0$ and $\mathbf{C}_0 = \mathbf{0}_k$.

Step 3. For the current alias matrix \mathbf{A} of size $(k+i) \times n$, spell out all 2^{n-k-i} alias sets: $G_0, G_1, \dots, G_{2^{n-k-i}-1}$. Let Ω be the set consisting of all of the 2^{n-k-i} alias sets. Also let Ω_1 and Ω_2 be the two exclusive subsets of Ω , consisting of the alias sets that contain exactly i and $i + 1$ elements of β . Moreover, let $\Omega_3 = \Omega_2$ if the number of the elements in Ω_2 is less than or equal to that in Ω_1 ; otherwise, $\Omega_3 = \Omega_1$.

Then define $\Omega_4 = \{G_j * G_{j'} | G_j \text{ and } G_{j'} \in \Omega_3 \text{ for } j \neq j'\} \cup \{G_0\}$. Let $\Omega_5 = \Omega \setminus \Omega_4$; here Ω_5 is the complement of Ω_4 . If Ω_5 is empty, then return to step 1 and restart, using a different orthogonal single-flat design for β . If no such design is available for step 1, then go to step 8.

Step 4. Increase i by 1, that is, $i \leftarrow i + 1$. Extend the current \mathbf{A} by adding the defining vector of an effect (arbitrarily choosing the first one) of some $G_i \in \Omega_5$ to be its $(k + i)$ th row.

Step 5. Augment the current \mathbf{C}_0 by one copy of itself, letting $\mathbf{C}_0 = [\mathbf{C}_0, \mathbf{C}_0]$. Then add the row vector $[0'_i, 1'_i]$, where $\mathbf{1}_i$ denotes the vector of order i with all elements equal to 1 over GF[2] as the $(k + i)$ th row of the current \mathbf{C}_0 .

Step 6. Let $\mathbf{c} = \mathbf{0}_{k+i}$. The f -PFDR determined by the current (\mathbf{A}, \mathbf{C}) , where $f = 2^i + 1$ and $\mathbf{C} = [\mathbf{C}_0, \mathbf{c}]$, is the desired partially replicated design with 2^{n-k-i} df for the pure error variance.

Step 7. Return to step 3 until $i = n - k$.

Step 8. Stop.

By Corollary 1, it can be verified that the designs generated from this algorithm are D-optimal for the specified β over the class of competing f -PFDRs. In step 4, when the alias matrix \mathbf{A} is extended by one defining vector, the number of alias sets is automatically reduced in half by merging two of the alias sets into a single alias set. The choice of the defining vector from Ω_5 of step 3 avoids the merging together of any pair of Ω_3 s. This also makes all of the v_j 's equal to two distinct values with a difference of 1. Thus the second condition of Corollary 1 is fulfilled. In step 5, it can be easily checked that for defining vectors \mathbf{e}_1 and \mathbf{e}_2 with a vector \mathbf{w} satisfying $\mathbf{e}_1 + \mathbf{e}_2 = \mathbf{w}'\mathbf{A}$ (i.e., the two effects corresponding to \mathbf{e}_1 and \mathbf{e}_2 are in the same alias set), $\mathbf{w}'\mathbf{C}_0$ always includes equal occurrences of 0 and 1. By Equations (2) and (3), the modification of \mathbf{C}_0 in this step ensures that the PFD determined by the current $(\mathbf{A}, \mathbf{C}_0)$ is orthogonal for β , so condition (a) of Corollary 1 is satisfied. In step 6, the \mathbf{c} is just a copy of $\mathbf{0}$ (the first coset indicator vector of \mathbf{C}_0), which is added with the \mathbf{C}_0 to duplicate the single-flat design determined by the current $(\mathbf{A}, \mathbf{0})$.

The algorithm applied to the design of Example 1 might run as follows:

Step 1. For $\beta = \{\mu, 1, 2, 3, 4, 5, 6, 12, 13, 14, 15, 16\}$, by the Franklin–Bailey algorithm, we have an orthogonal single-flat 2^{6-2} design for β determined by the two independent defining relations, $I = 1236$ and $I = 1345$, whose alias matrix \mathbf{A} can be written as

$$\mathbf{A}(2 \times 6) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Step 2. $i = 0$ and

$$\mathbf{C}_0(2 \times 1) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Step 3. The $2^{6-2} = 16$ alias sets are as given in Example 1: $\Omega_1 = \{G_{12}, G_{13}, G_{14}, G_{15}\}$; all of these alias sets contain none element of β ;

$$\Omega_2 = \{G_0, G_1, G_2, G_3, G_4, G_5,$$

$$G_6, G_7, G_8, G_9, G_{10}, G_{11}\};$$

each of these alias sets contains exactly one element of β . Thus $\Omega_3 = \Omega_1, \Omega_4 = \{G_0, G_1, G_3, G_8\}$, and

$$\Omega_5 = \{G_2, G_4, G_5, G_6, G_7, G_9, G_{10},$$

$$G_{11}, G_{12}, G_{13}, G_{14}, G_{15}\}.$$

Step 4. For $i = 1$, defining vector $[0, 1, 0, 0, 0, 0]$ corresponding to word “2” in G_2 of Ω_5 is added with the current \mathbf{A} , yielding

$$\mathbf{A}(3 \times 6) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Step 5. The current \mathbf{C}_0 is modified to be

$$\mathbf{C}_0(3 \times 2) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Step 6. The f -PFDR with $f = 2^i + 1 = 3$, determined by the current (\mathbf{A}, \mathbf{C}) , where \mathbf{C} , is given by

$$\mathbf{C}(3 \times 3) = [\mathbf{C}_0, \mathbf{0}] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

is the desired design with $N = 24$ and 8 df for the pure error variance. In other words, this 3-PFDR comprises the originally chosen 2^{6-2} design and its 1/2 fraction satisfying $I = 2$.

Step 7. For $i = 2$, perform steps 3–6 as follows:

Step 3. The new $2^{6-2-1} = 8$ alias sets are given by

$$G_0: \quad I = 1236 = 1345 = 2456 = 2 \\ = 136 = 12345 = 456,$$

$$G_1: \quad 1 = 236 = 345 = 12456 = 12 \\ = 36 = 2345 = 1456,$$

$$G_2: \quad 3 = 126 = 145 = 23456 = 23 \\ = 16 = 1245 = 3456,$$

$$G_3: \quad 6 = 123 = 13456 = 245 = 26 \\ = 13 = 123456 = 45,$$

$$G_4: \quad 4 = 12346 = 135 = 256 = 24 \\ = 1346 = 1235 = 56,$$

$$G_5: \quad 5 = 12356 = 134 = 246 = 25 \\ = 1356 = 1234 = 46,$$

$$G_6: \quad 14 = 2346 = 35 = 1256 = 124 \\ = 346 = 235 = 156,$$

$$G_7: \quad 15 = 2356 = 34 = 1246 = 125 \\ = 356 = 234 = 146.$$

Now $\Omega_1 = \{G_4, G_5, G_6, G_7\}$ and $\Omega_2 = \{G_0, G_1, G_2, G_3\}$. Let $\Omega_3 = \Omega_2$. So $\Omega_4 = \{G_0, G_1, G_2, G_3\}$ and $\Omega_5 = \{G_4, G_5, G_6, G_7\}$.

Step 4. For $i = 2$, defining vector $[0, 0, 0, 1, 0, 0]$ corresponding to word “4” in G_4 of Ω_5 is added with the current \mathbf{A} , yielding

$$\mathbf{A}(4 \times 6) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Step 5. Modify the current \mathbf{C}_0 to be

$$\mathbf{C}_0(4 \times 4) = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Step 6. The f -PFDR with $f = 2^i + 1 = 5$, determined by the current (\mathbf{A}, \mathbf{C}) , where \mathbf{C}_0 , is given by

$$\mathbf{C}(4 \times 5) = [\mathbf{C}_0, \mathbf{0}] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix},$$

is the required design with $N = 20$ and 4 df for the pure error variance. This 5-PFDR comprises the original 2^{6-2} design and its $1/4$ fraction satisfying $I = 2$ and $I = 4$.

Similarly, for $i = 3$, perform steps 3–6 to yield the desired 9-PFDR that comprises the original 2^{6-2} design and its $1/8$ fraction satisfying $I = 2, I = 4$, and $I = 1$. This 18-run design has 2 df for the pure error variance. Finally, for $i = 4$, perform steps 3–6 to get a 17-PFDR that comprises the originally chosen

2^{6-2} design and its $1/16$ fraction satisfying $I = 2, I = 4, I = 1$, and $I = 3$. There is precisely one run in each flat of the design, and this 17-PFDR has only one df for the pure error variance.

Obviously, the algorithm generates a sequence of D-optimal f -PFDRs for β . And the resulting designs are those of the original 2^{6-2} design, determined by the defining relations $I = 1236$ and $I = 1345$, augmented with its $1/2^q$ fractions, for $q = 1, 2, 3, 4$. The defining relations of these $1/2^q$ fractions can be obtained by combining those of the 2^{6-2} design with the first q defining relations of the set $\{I = 2, I = 4, I = 1, I = 3\}$. We denote these D-optimal f -PFDRs simply as $\{1236, 1345\} \cup \{2, 4, 1, 3\}$ (all “ $I =$ ” are omitted in the defining relations).

The algorithm has been implemented in Fortran code. For the purpose of illustration, we use the computer program to generate the D-optimal f -PFDRs for two classes of specified possibly active effects. The cases of class 1, labeled as interactions within subgroups, always consist of the constant term, all main effects, and all two-factor interactions within each of the subgroups of factors. The cases of class 2, labeled as interactions between subgroups, always consist of the constant term, all main effects, and all two-factor interactions between subgroups of factors. The run sizes of the orthogonal single-flat designs from Franklin–Bailey algorithm for these cases are fixed at 16 (i.e., $2^{n-k} = 2^{n-(n-4)} = 16$). The results are displayed in Tables 1 and 2. In the tables, the sequence of D-optimal f -PFDRs for a specified β is represented by $S_1 \cup S_2$, where S_1 comprises the $n - 4$ independent defining relations of the chosen single-flat design for β and S_2 comprises another 4 independent defining relations used to determine the duplicated fractions. The twice-replicated 2^p treatment combinations of the f -PFDRs are those

Table 1. D-optimal f -PFDRs for the interactions within subgroups

n	v	Estimable interactions	The sequence of D-optimal designs
5	9	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{12345\} \cup \{4, 1, 2, 3\}$
5	12	$\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$	$\{12345\} \cup \{1, 234, 3, 4\}$
6	9	$\{1 \cdot 2, 3 \cdot 4\}$	$\{1235, 1246\} \cup \{134, 2, 3, 4\}$
6	10	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{1235, 1246\} \cup \{4, 1, 3, 2\}$
6	13	$\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$	$\{1235, 1246\} \cup \{3, 4, 1, 2\}$
7	9	$\{1 \cdot 2\}$	$\{1235, 1246, 1347\} \cup \{134, 2, 3, 4\}$
7	10	$\{1 \cdot 2, 3 \cdot 4\}$	$\{1235, 1246, 1347\} \cup \{2, 3, 4, 1\}$
7	11	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6\}$	$\{1235, 1346, 1247\} \cup \{2, 3, 1, 4\}$
7	11	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{1235, 1246, 1347\} \cup \{4, 1, 2, 3\}$
7	14	$\{1, 2, 3, 4\} \times \{1, 2, 3, 4\}$	$\{1235, 1246, 1347\} \cup \{2, 3, 4, 1\}$
8	10	$\{1 \cdot 2\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 3, 4, 2\}$
8	11	$\{1 \cdot 2, 3 \cdot 4\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 3, 2, 4\}$
8	12	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6\}$	$\{1235, 1346, 1247, 2348\} \cup \{1, 3, 2, 4\}$
8	12	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 4, 2, 3\}$
8	13	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6, 7 \cdot 8\}$	$\{1235, 1346, 1247, 2348\} \cup \{1, 2, 3, 4\}$
9	11	$\{1 \cdot 2\}$	$\{135, 146, 237, 248, 349\} \cup \{123, 4, 2, 3\}$
9	12	$\{1 \cdot 2, 3 \cdot 4\}$	$\{135, 146, 237, 248, 1239\} \cup \{4, 123, 2, 3\}$
9	13	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6\}$	$\{135, 246, 147, 238, 1239\} \cup \{1, 4, 2, 3\}$
9	13	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{145, 246, 347, 1238, 1249\} \cup \{3, 4, 1, 2\}$
9	14	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6, 7 \cdot 8\}$	$\{135, 246, 147, 1238, 239\} \cup \{1, 2, 4, 3\}$
10	12	$\{1 \cdot 2\}$	$\{135, 146, 237, 248, 349, 1237\} \cup \{4, 123, 2, 3\}$
10	13	$\{1 \cdot 2, 3 \cdot 4\}$	$\{135, 146, 237, 248, 1239, 1247\} \cup \{3, 4, 1, 2\}$
10	14	$\{1 \cdot 2, 3 \cdot 4, 5 \cdot 6\}$	$\{135, 246, 147, 238, 1239, 1247\} \cup \{1, 3, 4, 2\}$
10	14	$\{1, 2, 3\} \times \{1, 2, 3\}$	$\{145, 246, 347, 1238, 1249, 1347\} \cup \{2, 3, 4, 1\}$

NOTE: In Tables 1 and 2, we adopt the notation that $A \times B$ represents all two-factor interactions obtained by selecting one element of A and one element of B . Note that the specified effects always contain the constant term and all main effects. In the tables, “ T ” denotes 10.

Table 2. D-optimal f -PFDRs for the interactions between subgroups

n	v	Estimable interactions	The sequence of D-optimal designs
5	9	$\{1\} \times \{2, 3, 4\}$	$\{12345\} \cup \{234, 3, 4, 1\}$
5	10	$\{1\} \times \{2, 3, 4, 5\}$	$\{12345\} \cup \{2, 3, 4, 1\}$
5	12	$\{1, 2\} \times \{3, 4, 5\}$	$\{12345\} \cup \{1, 2, 3, 4\}$
6	9	$\{1\} \times \{2, 3\}$	$\{1235, 1246\} \cup \{134, 2, 3, 4\}$
6	10	$\{1\} \times \{2, 3, 4\}$	$\{1235, 1246\} \cup \{134, 2, 3, 4\}$
6	11	$\{1\} \times \{2, 3, 4, 5\}$	$\{1235, 1246\} \cup \{4, 3, 1, 2\}$
6	11	$\{1, 2\} \times \{3, 4\}$	$\{1235, 1246\} \cup \{3, 4, 1, 2\}$
6	12	$\{1\} \times \{2, 3, 4, 5, 6\}$	$\{1235, 1246\} \cup \{3, 4, 1, 2\}$
6	13	$\{1, 2\} \times \{3, 4, 5\}$	$\{1345, 1236\} \cup \{1, 2, 4, 3\}$
7	10	$\{1\} \times \{2, 3\}$	$\{1235, 1246, 1347\} \cup \{124, 2, 3, 4\}$
7	11	$\{1\} \times \{2, 3, 4\}$	$\{1235, 1246, 1347\} \cup \{123, 4, 2, 3\}$
7	12	$\{1\} \times \{2, 3, 4, 5\}$	$\{1235, 1246, 1347\} \cup \{4, 123, 2, 3\}$
7	12	$\{1, 2\} \times \{3, 4\}$	$\{1235, 1246, 1347\} \cup \{3, 24, 1, 4\}$
7	13	$\{1\} \times \{2, 3, 4, 5, 6\}$	$\{1235, 1246, 1347\} \cup \{3, 4, 2, 1\}$
7	14	$\{1\} \times \{2, 3, 4, 5, 6, 7\}$	$\{1235, 1246, 1347\} \cup \{2, 3, 4, 1\}$
7	14	$\{1, 2\} \times \{3, 4, 5\}$	$\{1345, 1236, 1247\} \cup \{1, 2, 3, 4\}$
8	11	$\{1\} \times \{2, 3\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 4, 2, 3\}$
8	12	$\{1\} \times \{2, 3, 4\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 234, 3, 4\}$
8	13	$\{1\} \times \{2, 3, 4, 5\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 4, 2, 3\}$
8	13	$\{1, 2\} \times \{3, 4\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 3, 2, 4\}$
8	14	$\{1\} \times \{2, 3, 4, 5, 6\}$	$\{1235, 1246, 1347, 2348\} \cup \{1, 3, 4, 2\}$
9	12	$\{1\} \times \{2, 3\}$	$\{145, 236, 247, 348, 1239\} \cup \{4, 123, 2, 3\}$
9	13	$\{1\} \times \{2, 3, 4\}$	$\{235, 246, 347, 1238, 1249\} \cup \{3, 4, 1, 2\}$
9	14	$\{1\} \times \{2, 3, 4, 5\}$	$\{235, 246, 347, 1248, 1349\} \cup \{2, 3, 4, 1\}$
9	14	$\{1, 2\} \times \{3, 4\}$	$\{125, 346, 1237, 1248, 1349\} \cup \{2, 3, 4, 1\}$
10	13	$\{1\} \times \{2, 3\}$	$\{145, 236, 247, 348, 1239, 1247\} \cup \{3, 4, 1, 2\}$
10	14	$\{1\} \times \{2, 3, 4\}$	$\{235, 246, 347, 1238, 1249, 1347\} \cup \{2, 3, 4, 1\}$

NOTE: Notes are the same as in Table 1.

of the original single-flat design satisfying the first q defining relations of S_2 , where $q = 4 - p$ and $p = 0, 1, 2, 3$.

The proposed algorithm is able to generate D-optimal designs for any specified β over the class of f -PFDRs with the same run size. However, it may be of interest to investigate the D-optimality of the designs obtained compared with all possible two-level designs. Because the original design in the algorithm is orthogonal for β , and because an orthogonal design is known to be D-optimal among all possible two-level designs, we only check whether or not the determination of duplicated runs in the algorithm is the best choice. For each case given in Tables 1 and 2, we augment the originally chosen single-flat design with a specific number of duplicated runs from its 16 treatment combinations. There are $\binom{16}{8}$ possible choices for partially replicated designs with 8 additional runs, $\binom{16}{4}$ possible choices for these designs with 4 additional runs, and $\binom{16}{2}$ possible choices for these designs with 2 additional runs. Then we calculate the determinants of all of the partially replicated designs and compare them with those of the proposed f -PFDRs. Interestingly, all of the proposed f -PFDRs attain the maximum as some other choices, meaning that the best choice is not unique. Finally, note that the 17-PFDRs are 16-run orthogonal designs plus one run plans, so D-optimality holds among all possible two-level designs (Mukerjee 1999).

4. A SIMULATION STUDY

To justify the use of partially replicated designs, we conduct a simulation study. We base the study on the 2^{6-2} design in

Example 1, designated the 16-run unreplicated design (URD). We also consider the replicated designs derived from the 16-run URD, including the fully replicated design (FRD) with 16 df for the pure error variance and the 3-PFDR (8 df), 5-PFDR (4 df), 9-PFDR (2 df), and 17-PFDR(1 df) presented in Section 3. We include the 2^{6-1} design of resolution VI, designated the 32-run URD, as well.

To explore the power of the designs in identification of truly active effects, we specify six scenarios in the simulation.

Note that, from Corollary 1, it can be verified that the 3-PFDR, 5-PFDR, 9-PFDR, and 17-PFDR discussed here are all D-optimal for the effects of any β_T in Table 3 if they are specified for seeking the D-optimal designs.

For given β_T and a design under consideration, a simulated data set \mathbf{y} is realized according to the model $\mathbf{y} = \mathbf{X}\beta_T + \epsilon$. We set $\beta_T = \theta \mathbf{1}$, where $\mathbf{1}$ denotes the vector with all entries equal

Table 3. The truly active effects of the model for simulation study. Let β_T denote the set of these effects

Scenario	β_T
1	$\{\mu, 1\}$
2	$\{\mu, 1, 3, 13\}$
3	$\{\mu, 1, 3, 4, 13, 14\}$
4	$\{\mu, 1, 3, 4, 5, 13, 14, 15\}$
5	$\{\mu, 1, 2, 3, 4, 5, 12, 13, 14, 15\}$
6	$\{\mu, 1, 2, 3, 4, 5, 6, 12, 13, 14, 15, 16\}$

to 1 of the same order as β_T . θ is a specified value representing the effect size of truly active effects. Thus \mathbf{y} is generated independently of the normal distribution $N(\theta(\mathbf{X}\mathbf{1}), \sigma^2\mathbf{I})$, where \mathbf{I} denotes the identity matrix. Let $\sigma^2 = 16$ for the 16-run URD and its derived partially replicated designs and $\sigma^2 = 32$ for the 32-run URD. This forces the least squares estimators of the truly active effects specified in β_T to follow $N(\theta, 1)$ and those of the truly inert effects to follow $N(0, 1)$. The effect size θ is fixed at 1, 2, 3, and 4 in the simulation. For each combination of scenario, design, and effect size θ , data are repeatedly generated 10,000 times.

The unreplicated designs are analyzed by Lenth's method with a critical value of 2.156 for the 16-run URD and 2.064 for the 32-run URD, both of which control the IER at 0.05 (Ye and

Hamada 2000). The partially and fully replicated designs are analyzed by least squares regression under the saturated effects model. Student's t -test, under a significance level of 0.05, based on the estimates of pure error variance, is used to determine active effects. We first verify the IER for each design, and find that all the simulated IERs are very close to the nominal value of 0.05. The simulated power of each design for each scenario is defined as the average proportion of the truly active effects that are determined to be active or significant. The results are displayed in Table 4.

From Table 4, the 16-run URD with Lenth's method appears to be reliable only when the number of active effects is not large, that is, scenarios 1, 2, and 3. This is reasonable, because PSE is highly related to the number of inert effects. Interest-

Table 4. The simulated power of the designs under the six scenarios given in Table 3

Design	Scenario	$\theta = 1$	$\theta = 2$	$\theta = 3$	$\theta = 4$
16-run URD	1	0.1320	0.3799	0.6950	0.9026
	2	0.1121	0.2919	0.5850	0.8446
	3	0.0950	0.1883	0.3809	0.6663
	4	0.0781	0.0915	0.1246	0.2355
	5	0.0604	0.0331	0.0070	0.0005
	6	0.0470	0.0092	0.0003	0
17-PFDR	1	0.0737	0.1261	0.1890	0.2506
	2	0.0741	0.1269	0.1894	0.2501
	3	0.0753	0.1287	0.1902	0.2478
	4	0.0768	0.1307	0.1896	0.2456
	5	0.0743	0.1266	0.1886	0.2514
	6	0.0742	0.1266	0.1885	0.2512
9-PFDR	1	0.1005	0.2287	0.4050	0.5881
	2	0.1003	0.2288	0.4050	0.5870
	3	0.0990	0.2291	0.4030	0.5840
	4	0.0998	0.2328	0.4046	0.5878
	5	0.0984	0.2292	0.4045	0.5862
	6	0.0984	0.2291	0.4041	0.5861
5-PFDR	1	0.1320	0.3750	0.6766	0.8850
	2	0.1333	0.3752	0.6746	0.8851
	3	0.1332	0.3752	0.6731	0.8854
	4	0.1331	0.3776	0.6697	0.8877
	5	0.1321	0.3735	0.6732	0.8857
	6	0.1322	0.3734	0.6730	0.8855
3-PFDR	1	0.1758	0.5278	0.8576	0.9802
	2	0.1761	0.5285	0.8574	0.9802
	3	0.1761	0.5287	0.8570	0.9801
	4	0.1751	0.5281	0.8536	0.9806
	5	0.1751	0.5269	0.8568	0.9805
	6	0.1749	0.5269	0.8566	0.9805
FRD	1	0.2654	0.7564	0.9780	0.9995
	2	0.2657	0.7579	0.9779	0.9996
	3	0.2649	0.7566	0.9780	0.9996
	4	0.2637	0.7554	0.9770	0.9995
	5	0.2645	0.7563	0.9780	0.9995
	6	0.2645	0.7564	0.9780	0.9995
32-run URD	1	0.1476	0.4446	0.7771	0.9524
	2	0.1371	0.4004	0.7446	0.9436
	3	0.1283	0.3525	0.6930	0.9284
	4	0.1168	0.2965	0.6175	0.8980
	5	0.1070	0.2363	0.5076	0.8327
	6	0.0968	0.1760	0.3620	0.6964

ingly, the 17-PFDR or 9-PFDR with only 1 or 2 df for the pure error variance is not as powerful as the 16-run URD in these scenarios of effect sparsity. The 5-PFDR or 3-PFDR with 4 or 8 replicated runs appears to be highly competitive, because both require much fewer runs than the fully replicated design and the 32-run URD. But these two partially replicated designs not only have higher power than the 16-run URD, but also perform as well as or even better than the 32-run URD with Lenth's method. Another advantage of the partially replicated designs over the unreplicated designs is that their powers do not degrade with an increasing number of active effects. This is because the error variance estimate using pure replicates is model-independent.

A potential improved method for analyzing the partially replicated designs without sufficient pure error df is to combine Lenth's method with least squares regression using the pure error estimate; that is, first use Lenth's method to analyze the unreplicated design of them and then apply least squares regression on the whole data. We thus conduct the simulation study on the 17-PFDR and 9-PFDR using the combined analysis method. Similarly, we first check the simulated IERs for this study, which are 0.0938 for the 17-PFDR and 0.0905 for the 9-PFDR. These two simulated IERs are larger than the nominal value of 0.05. The simulated powers are displayed in Table 5.

As can be seen in Table 5, the simulated power using the combined analysis method compared with the least squares regression only, is indeed improved for the scenarios of effect sparsity, but not significantly so for the scenarios with a large number of active effects (scenarios 5 and 6). They also outperform the 16-run URD with Lenth's method in terms of power. On the other hand, the improvement is at a cost of inflation in the IER. But the inflated IER may be acceptable so as to increase the power at the early stages of experimentation. Thus the combined analysis method can be recommended for the PFDRs without sufficient df for pure error.

5. CONCLUDING REMARKS

In a two-level factorial experiment, unreplicated designs provide no information on the pure error variance. This may cause the entire experiment to be less informative in identifying active effects, particularly when the number of inert effects is not

large. Based on the results of the present study, we suggest that users consider including partial replication in their experimental designs to compensate for the deficiency of unreplicated designs, mainly because the additional runs may offer a useful replication-based estimate of the error variance for finding active effects. Certainly, additional runs can be planned to estimate effects other than those of β , by including treatment combinations distinct from the original single-flat design. In practice, the specification of β should contain all of the potentially active effects; thus it should prove valuable by trading off some df of the effects (which are already deemed negligible) for a reliable estimate of the experimental error variance. In this study, we take advantage of the prior information of β and design an experiment with higher precision on β . Further study is needed to determine the partial replication if no such prior information is available.

Determining the run size of partial replication could depend on the cost of running experiments and the power requirement for identifying active effects. In this article we have proposed an algorithm to generate a sequence of D-optimal f -PFDRs for a specified β , which provides a flexible choice of the number of replicated runs. As we show in Section 4, a 16-run orthogonal design plus 4 or 8 duplicated runs plan is a workable choice for identifying truly active effects. The result of adding 4 or 8 duplicated runs can be applicable to orthogonal designs with run sizes other than 16, because the Student t -test is based on the estimate of pure error variance. The proposed algorithm is capable of generating such practical partially replicated designs for any specified β ; however, the Franklin–Bailey approach is adopted as part of this algorithm. In our experience, using the Franklin–Bailey algorithm may require a significant amount of computing time for cases with a large number of factors or the number of elements in β close to the run size of the required orthogonal single-flat design.

ACKNOWLEDGMENTS

The authors thank the editor, an associate editor, and three referees for their constructive suggestions and comments that resulted in a much improved article. The work of Liao was supported in part by the National Science Council of ROC (contract NSC 94-2118-M-002-006).

Table 5. Simulated power of the 17-PFDR and 9-PFDR using the combined analysis method under the scenarios given in Table 3

Design	Scenario	$\theta = 1$	$\theta = 2$	$\theta = 3$	$\theta = 4$
7-PFDR	1	0.1862	0.4380	0.7330	0.9174
	2	0.1687	0.3650	0.6472	0.8713
	3	0.1528	0.2777	0.4805	0.7340
	4	0.1381	0.2005	0.2799	0.4197
	5	0.1246	0.1522	0.1936	0.2518
	6	0.1139	0.1352	0.1896	0.2495
9-PFDR	1	0.1961	0.4667	0.7609	0.9317
	2	0.1804	0.4062	0.6925	0.9000
	3	0.1634	0.3382	0.5806	0.8167
	4	0.1516	0.2801	0.4544	0.6557
	5	0.1398	0.2462	0.4070	0.5864
	6	0.1303	0.2318	0.4049	0.5860

APPENDIX: PROOF OF THEOREM 1

First, we discuss the choice of indicator coset matrix \mathbf{C} . Suppose that $\boldsymbol{\beta}$ is specified and \mathbf{A} is fixed to be \mathbf{A}^* . Let \mathcal{D}_1 denote the subset of \mathcal{D} , which consists of all f -PFDRs determined by $(\mathbf{A}^*, \mathbf{C})$. All designs in \mathcal{D}_1 have the same alias sets, so the numbers of elements in alias sets v_j are fixed. We now prove that d^* is D-optimal over \mathcal{D}_1 . Recall that d^* is determined by the pair of matrixes $(\mathbf{A}^*, \mathbf{C}^*)$, where $\mathbf{C}^* = [\mathbf{C}_0^*, \mathbf{c}^*]$ and the $(f-1)$ -PFD determined by $(\mathbf{A}^*, \mathbf{C}_0^*)$ is orthogonal for $\boldsymbol{\beta}$. Let \mathbf{e}_1 and \mathbf{e}_2 be two defining vectors corresponding to two distinct effects in alias set j . By Equation (2), we have

$$\begin{aligned} m(\mathbf{e}_1, \mathbf{e}_2) &= \left[\sum_{i=1}^f (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{z}_i} \right] \left[\sum_{\mathbf{v}} (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{B}\mathbf{v}} \right] \\ &= (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{z}^*} \left[\sum_{\mathbf{v}} (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{B}\mathbf{v}} \right] \\ &= 2^{n-k} \times (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{z}^*}, \end{aligned}$$

where \mathbf{z}^* is a solution to the equation $\mathbf{A}^*\mathbf{t} = \mathbf{c}^*$. Thus, from Equation (3), the submatrix \mathbf{M}_j^* , corresponding to alias set j on the diagonal of information matrix \mathbf{M}^* using d^* , is given by

$$\mathbf{M}_j^* = N_0 \mathbf{I}_{v_j} + 2^{n-k} \mathbf{h}_j \mathbf{h}_j',$$

where $N_0 = (f-1) \times 2^{n-k}$ and \mathbf{h}_j is the vector of order v_j with elements equal to $+1$ or -1 , depending on whether $(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{z}^*$ is equal to 0 or 1. Furthermore, the characteristic equation of \mathbf{M}_j^* is given by

$$\begin{aligned} |\mathbf{M}_j^* - \lambda \mathbf{I}_{v_j}| &= |N_0 \mathbf{I}_{v_j} + 2^{n-k} \mathbf{h}_j \mathbf{h}_j' - \lambda \mathbf{I}_{v_j}| \\ &= |N_0 \mathbf{I}_{v_j} + 2^{n-k} \mathbf{1}_{v_j} \mathbf{1}_{v_j}' - \lambda \mathbf{I}_{v_j}| \\ &= |N_0 \mathbf{I}_{v_j} + 2^{n-k} \mathbf{J}_{v_j} - \lambda \mathbf{I}_{v_j}|, \end{aligned}$$

where \mathbf{J}_{v_j} denotes the square matrix of order v_j with all entries equal to 1. Thus \mathbf{M}_j^* has the same eigenvalues as the completely symmetric matrix $N_0 \mathbf{I}_{v_j} + 2^{n-k} \mathbf{J}_{v_j}$, which are $\lambda_1 = N_0$ with multiplicity $v_j - 1$ and $\lambda_2 = N_0 + v_j \times 2^{n-k}$ with multiplicity 1.

For any design $d \in \mathcal{D}_1$, let \mathbf{M}_j be its submatrix corresponding to \mathbf{M}_j^* . The trace of \mathbf{M}_j , denoted by $\text{Tr}(\mathbf{M}_j)$, is a constant. Again, by Equation (2), the existence of an orthogonal $(f-1)$ -PFD for $\boldsymbol{\beta}$ implies that f must be an odd number for $f > 2$. The smallest absolute value of $\sum_{i=1}^f (-1)^{(\mathbf{e}_1 + \mathbf{e}_2)\mathbf{z}_i}$ is equal to 1 for odd f , implying that $\text{Tr}(\mathbf{M}_j^*)^2$ attains the minimum over \mathcal{D}_1 . By theorem 2.3 of Cheng (1978), we have that $|\mathbf{M}_j^*| \geq |\mathbf{M}_j|$, leading to

$$|\mathbf{M}^*| = \prod_{j=0}^{2^{n-k}-1} |\mathbf{M}_j^*| \geq \prod_{j=0}^{2^{n-k}-1} |\mathbf{M}_j| = |\mathbf{M}|,$$

where $|\mathbf{M}_j^*|$ and $|\mathbf{M}_j|$ take the value 1 if $v_j = 0$ (empty alias set). Thus d^* is D-optimal over \mathcal{D}_1 .

Now we argue that \mathbf{A}^* is the best choice among all possible \mathbf{A} , implying that d^* is D-optimal over \mathcal{D} . The determinant of \mathbf{M}^* is equal to the product of its eigenvalues and is given by

$$\begin{aligned} |\mathbf{M}^*| &= \prod_{j=0}^{2^{n-k}-1} |\mathbf{M}_j^*| \\ &= \prod_{j=0}^{2^{n-k}-1} (N_0)^{v_j-1} (N_0 + v_j \times 2^{n-k}) \\ &= \frac{(N_0)^v}{(f-1)^{2^{n-k}}} \prod_{j=0}^{2^{n-k}-1} [(f-1) + v_j]. \end{aligned}$$

Therefore, \mathbf{A}^* must be chosen to make the v_j^* , $j = 0, 1, \dots, 2^{n-k} - 1$, attain the maximum of $\prod_{j=0}^{2^{n-k}-1} (f-1 + v_j)$ subject to $\sum_{j=0}^{2^{n-k}-1} v_j = v$. This completes the proof.

[Received January 2007. Revised May 2008.]

REFERENCES

- Box, G. E. P., and Meyer, R. D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.
- Cheng, C. S. (1978), "Optimality of Certain Asymmetrical Experimental Designs," *The Annals of Statistics*, 6, 1239–1261.
- Dykstra Jr., O. (1959), "Partial Duplication of Factorial Experiments," *Technometrics*, 1, 63–75.
- Franklin, M. F., and Bailey, R. A. (1977), "Selecting Defining Contrasts and Confounded Effects in Two-Level Experiments," *Applied Statistics*, 26, 321–326.
- Hamada, H., and Balakrishnan, N. (1998), "Analyzing Unreplicated Factorial Experiments: A Review With Some New Proposals," *Statistica Sinica*, 8, 1–41.
- Lenth, R. V. (1989), "Quick and Easy Analysis of Unreplicated Factorials," *Technometrics*, 31, 469–473.
- Li, F. (1991), "Sequential and Optimal Single Stage Factorial Designs With Industrial Applications," Ph.D. dissertation, Colorado State University, Fort Collins, CO.
- Liao, C. T., and Chai, F. S. (2004), "Partially Replicated Two-Level Fractional Factorial Designs," *The Canadian Journal of Statistics*, 32, 421–438.
- Liao, C. T., Iyer, H. K., and Vecchia, D. F. (1996), "Orthogonal Designs of Two-Level User-Specified Resolution Where $N \neq 2^k$," *Technometrics*, 38, 342–353.
- Mukerjee, R. (1999), "On the Optimality of Orthogonal Array Plus One Run Plan," *The Annals of Statistics*, 27, 82–93.
- Pigeon, J. G., and McAllister, P. R. (1989), "A Note on Partially Replicated Orthogonal Main-Effect Plan," *Technometrics*, 31, 249–251.
- Raktoe, B. L., Hedayat, A., and Federer, W. T. (1981), *Factorial Designs*, New York: Wiley.
- Snee, R. D. (1985), "Experimenting With a Large Number of Variables," in *Experiments in Industry: Design, Analysis and Interpretation of Results*, Milwaukee, WI: American Society for Quality Control, pp. 25–35.
- Wu, C. F. J., and Chen, Y. (1992), "A Graph-Aided Method for Planning Two-Level Experiments When Certain Interactions Are Important," *Technometrics*, 34, 162–174.
- Wu, C. F. J., and Hamada, M. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York: Wiley.
- Ye, K. Q., and Hamada, M. (2000), "Critical Values of the Lenth Method for Unreplicated Factorial Designs," *Journal of Quality Technology*, 32, 57–66.
- Youden, W. J., and Hunter, J. S. (1955), "Partially Replicated Latin Squares," *Biometrics*, 11, 399–405.