Statistica Sinica **21** (2011), 1571-1589
doi:http://dx.doi.org/10.5705/ss.2008.336

A THEORY ON CONSTRUCTING 2^{n-m} DESIGNS WITH GENERAL MINIMUM LOWER ORDER CONFOUNDING

Pengfei Li[1], Shengli Zhao[2] and Runchu Zhang[3,4]

[1] *University of Alberta,* [2] *Qufu Normal University,*
[3] *Northeast Normal University and* [4] *Nankai University*

Abstract: When designing an experiment, it is important to choose a design that is optimal under model uncertainty. The general minimum lower-order confounding (GMC) criterion can be used to control aliasing among lower-order factorial effects. A characterization of GMC via complementary sets was considered in Zhang and Mukerjee (2009a); however, the problem of constructing GMC designs is only partially solved. We provide a solution for two-level factorial designs with n factors and $N = 2^{n-m}$ runs subject to a restriction on (n, N): $5N/16 + 1 \leq n \leq N - 1$. The construction is quite simple: every GMC design, up to isomorphism, consists of the last n columns of the saturated $2^{(N-1)-(N-1-n+m)}$ design with *Yates order*. In addition, we prove that GMC designs differ from minimum aberration designs when (n, N) satisfies either of the following conditions: (i) $5N/16 + 1 \leq n \leq N/2 - 4$, or (ii) $n \geq N/2$, $4 \leq n + 2^r - N \leq 2^{r-1} - 4$ with $r \geq 4$.

Key words and phrases: Aliased effect-number pattern, effect hierarchy principle, fractional factorial design, minimum aberration, resolution, wordlength pattern, Yates order.

1. Introduction

Regular two-level fractional factorial designs are common in practice. In the past three decades, many statisticians have investigated optimality criteria for selecting good designs; see Wu and Hamada (2000) and Mukerjee and Wu (2006) for detailed reviews. Minimum aberration (MA) has become the most popular criterion. Following the landmark paper of Fries and Hunter (1980), many investigators have contributed theory and methodology concerning MA; e.g., Franklin (1984), Chen and Wu (1991), Chen, Sun and Wu (1993), Chen and Hedayat (1996), Tang and Wu (1996), Zhang and Shao (2001), Butler (2003), Cheng and Tang (2005), Chen and Cheng (2006), and Xu and Cheng (2008).

Wu and Chen (1992) proposed the clear effects (CE) criterion, motivated by their observation that MA can fail to produce a design maximizing the number of clear two-factor interactions (2fis). A design is CE-optimal if it maximizes the number of clear main effects and 2fis. Wu and Hamada (2000) and Li et al.

(2006) discovered more examples of designs that are CE-optimal but not MA. For additional developments on CE, see Chen and Hedayat (1998), Tang et al. (2002), Wu and Wu (2002), Ai and Zhang (2004), Chen et al. (2006), Yang et al. (2006), and Zhao and Zhang (2008).

Sun (1993) introduced the maximum estimation capacity (MEC) criterion. A design is called MEC-optimal if it allows one to estimate the maximum number of models involving all main effects and some 2fis. For details, see Cheng and Mukerjee (1998) and Cheng, Steinberg, and Sun (1999).

Zhang et al. (2008) recently applied the effect hierarchy principle (Wu and Hamada (2000)) to motivate a characterization of designs using the *aliased effect-number pattern* (AENP). Based on the AENP, they proposed the *general minimum lower-order confounding* (GMC) criterion. They proved that the MA, CE, MEC, and GMC criteria can each be viewed as sequentially minimizing or maximizing the components of a corresponding vector function of the AENP. The GMC criterion compares designs by treating the AENP as a set, thus providing a unified approach applicable to the other criteria. GMC is flexible, accommodating prior information about the relative importance of factors and so incorporating preferences for estimation of the most important lower-order effects. This kind of prior information is often available in practice, so GMC designs are widely applicable.

Several articles have extended the GMC theory and methodology, including work on the construction of GMC designs by Zhang and Mukerjee (2009a,b). The first of these two papers characterizes the GMC criterion via a complementary set. This approach yields a powerful tool for GMC design construction when the number of factors in the complementary design is at most 15 and factors have prime or prime power levels. Much work remains, however, before the GMC construction problem is fully solved.

The current paper considers only two-level factorial designs with n factors and $N = 2^{n-m}$ runs. The primary contribution of the paper is the solution of the GMC construction problem for pairs (n, N) satisfying $5N/16 + 1 \leq n \leq N - 1$. The construction is quite simple: up to isomorphism, every GMC design consists simply of the last n columns of the saturated $2^{(N-1)-(N-1-n+m)}$ design with the Yates order. This simplicity makes the adoption of GMC designs convenient for practitioners.

The paper is organized as follows. Section 2 reviews the MA and GMC criteria, introduces notation, and presents our key Theorem 1. Section 3 develops our solution of the GMC construction problem subject to the restriction on (n, N). Section 4 shows that, for specified subsets of (n, N), the MA and GMC criteria yield different designs. The Appendix contains an outline of the proof of Theorem 1. Details of the proof are given in a supplement, available at the journal website `http://www.stat.sinica.edu.tw/statistica`.

2. Definitions, Notation and a Key Theorem

Let D denote a 2^{n-m} regular fractional factorial design with n factors, $N = 2^{n-m}$ runs, and m independent defining words. Factors are numbered $1, 2, \ldots, n$, but are also referred to as letters. The product (juxtaposition) of any subset of the n letters is called a word. The m independent defining words generate an identical subgroup, called the defining contrast subgroup of D. Here the operation for the defining contrast subgroup is multiplication with exponents reduced modulo 2. The number of letters in a word is called its wordlength. Let $A_i(D)$ denote the number of words with length i in the defining contrast subgroup of D. The vector $A(D) = (A_1(D), A_2(D), \ldots, A_n(D))$ is called wordlength pattern of D. The smallest r satisfying $A_r > 0$ is called the resolution of D. A 2^{n-m} design with resolution r is denoted by 2_r^{n-m}. A design is said to be MA if it sequentially minimizes $(A_1, A_2, \ldots, A_n)$ among all possible regular designs for given parameters n and m.

We now review some concepts from Zhang et al. (2008) concerning the GMC criterion in the context of two-level regular designs. If an ith-order effect is aliased with k jth-order effects simultaneously, we say that the severe degree of the ith-order effect being aliased with jth-order effects is k. Let ${}_i^{\#}C_j^{(k)}$ denote the number of ith-order effects aliased with jth-order effects at degree k, and define

$${}_i^{\#}C_j = ({}_i^{\#}C_j^{(0)}, {}_i^{\#}C_j^{(1)}, \ldots, {}_i^{\#}C_j^{(K_j)}),$$

where $K_j = \binom{n}{j}$. The sequence (or the set) of numbers

$${}^{\#}C = ({}_1^{\#}C_1, {}_0^{\#}C_2, {}_1^{\#}C_2, {}_2^{\#}C_1, {}_2^{\#}C_2, {}_0^{\#}C_3, {}_1^{\#}C_3, {}_2^{\#}C_3, {}_3^{\#}C_1, {}_3^{\#}C_2, {}_3^{\#}C_3, \ldots) \quad (2.1)$$

is called an aliased effect-number pattern (AENP). In (2.1) as a sequence, the general rule is that ${}_i^{\#}C_j$ is placed ahead of ${}_s^{\#}C_t$ if $\max(i, j) < \max(s, t)$, or if $\max(i, j) = \max(s, t)$ and $i < s$, or if $\max(i, j) = \max(s, t)$, $i = s$ and $j < t$.

Zhang and Mukerjee (2009a) noted that some terms in (2.1) are uniquely determined by the terms before them, for example, ${}_j^{\#}C_1^{(1)} = \sum_{k \geq 1} k\, {}_1^{\#}C_j^{(k)}$, and so the sequence (2.1) can be simplified into the version

$${}^{\#}C = ({}_1^{\#}C_2, {}_2^{\#}C_2, {}_1^{\#}C_3, {}_2^{\#}C_3, {}_3^{\#}C_2, {}_3^{\#}C_3, \ldots). \quad (2.2)$$

The GMC criterion based on (2.2) is defined as follows.

Definition 1. Let ${}^{\#}C_l$ be the l-th component of ${}^{\#}C$, and ${}^{\#}C(D_1)$ and ${}^{\#}C(D_2)$ be the AENPs of designs D_1 and D_2, respectively. Suppose that ${}^{\#}C_t$ is the first component such that ${}^{\#}C_t(D_1)$ and ${}^{\#}C_t(D_2)$ are different. If ${}^{\#}C_t(D_1) > {}^{\#}C_t(D_2)$, then D_1 is said to have less general lower-order confounding than D_2. A design D is said to have general minimum lower-order confounding if no other design has less general lower-order confounding than D, and such a design is called a GMC design.

It is not possible for one optimality criterion to be suitable for all theoretical and practical situations. The GMC design is no exception. According to the usual interpretation of the effect hierarchy principle for factorial effects, a good design should maximize the number of main effects and 2fis that can be separately estimated. When an experimenter has prior information about the relative importance of the factors in an experiment, a 'good' design should reserve the best estimates for the most important effects. According to Zhang et al. (2008), the GMC design must have the maximum number of clear main effects and 2fis if a CE-optimal design exists. If clear effects exist, then factors can be assigned to the columns of the GMC design so that the most important effects are clear. If clear effects are not available, then the GMC criterion can still be used to rank designs, and factors can be assigned to the columns of the GMC design so that the most important effects are least aliased.

Zhang et al. (2008) observed that MA designs sequentially minimize the lower-order confounding averaging over the severe degrees. Obviously, the average estimability of all effects of the same order differs from their separate estimability. It is known that, in some situations, the MA design may not have the maximum number of clear main effects and 2fis (Wu and Hamada (2000) and Li et al. (2006)). MA designs are thus more suitable when there is no prior information about the relative importance of the factors.

To illustrate the above points, let us revisit Example 4 in Zhang et al. (2008) in which the following two 2^{9-4} designs are considered:

$$D_1 : I = 1236 = 1247 = 1258 = 13459, \qquad D_2 : I = 1236 = 1247 = 1348 = 23459.$$

D_1 and D_2 are of MA and GMC, respectively. Both have ${}_1^{\#}C_2 = (9, 0, \ldots, 0)$, but ${}_2^{\#}C_2(D_1) = (8, 24, 0, 4, 0 \ldots, 0)$ and ${}_2^{\#}C_2(D_2) = (15, 0, 21, 0, \ldots, 0)$. Therefore both can clearly estimate all the main effects. As for the 2fis, the GMC design D_2 can clearly estimate all 15 2fis involving factor 5 or factor 9 while the MA design D_1 can only clearly estimate 8 2fis involving factor 9. Hence, when the experimenter is interested in some or all 2fis involving factor 5 or factor 9, design D_2 is a better choice than D_1. According to the usual interpretation of the effect hierarchy principle, D_2 is also a better choice (Wu and Hamada (2000)).

We now introduce notation needed for a key theorem used in our construction of GMC designs. For a 2^{n-m} design, write $q = n - m$ and let $1, \ldots, q$ stand for q independent columns with 2^q components of entries 1 or -1. Denote the saturated design by $H_q = \{1, 2, 12, 3, 13, 23, 123, \ldots, 12 \cdots q\}$, which is generated by the q independent columns and has the Yates order. Let H_r be the design consisting of the first $2^r - 1$ columns of H_q that is generated by the first r independent columns, $1, \ldots, r$. We then have

$$H_1 = \{1\} \text{ and } H_r = \{H_{r-1}, r, rH_{r-1}\} \text{ for } r = 2, \ldots, q.$$

Furthermore, let $S_{qr} = H_q \backslash H_r$, i.e., the set of columns in H_q but not in H_r. Also let $F_{qr} = \{q, qH_{r-1}\}$, where $qH_{r-1} = \{qd : d \in H_{r-1}\}$, and let $T_r = F_{rr}$. For consistency, we take $F_{q1} = \{q\}$, $T_1 = \{1\}$, $qH_{1-1} = \{q\}$, and $1H_{1-1} = \{1\}$. The designs F_{qr} and T_r with $r \geq 3$ are isomorphic saturated resolution IV designs with r independent columns. Introducing both F_{qr} and T_r simplifies the presentation in Sections 3 and 4.

Usually a 2^{n-m} design D can be obtained by selecting a subset of n columns from H_q such that D has q independent columns. For example, when $q = 3$, the H_3 design with the Yates order can be written as

$$H_3 = \{1, 2, 12, 3, 13, 23, 123\},$$

where 1, 2, and 3 denote the three independent columns, and "12" denotes the componentwise product of 1 and 2. The 2^{4-1} design $D = \{1, 2, 3, 123\}$ can be considered as a subset of H_3 consisting of the independent columns 1, 2, and 3 and their product 123. Since factor levels in an experiment are allocated to runs based on the the columns of the 2^{n-m} design, we do not distinguish between factors and columns. In addition, we label the columns of H_q using the natural products of the independent columns $1, \ldots, q$. For example, "125" stands for the 19-th column of H_q with $q \geq 5$ (since 19 in the decimal system equals 10011 in the binary system).

Throughout the paper, let S denote a design, a subset of H_q, with s factors (columns). In later sections, in order to get the GMC design, we need to maximize ${}_1^{\#}C_2(S)$ among all designs with s factors. Note that ${}_1^{\#}C_2(S)$ is maximized if S has resolution at least IV, or S consists of s independent factors. Strictly speaking, the resolution of S is not well-defined if S consists of s independent factors since all the elements in the wordlength pattern of S are 0. However, for this type of design, none of the main effects are aliased with other main effects or 2fis, which is an essential property of a design of resolution IV or higher. For convenience of presentation in later sections, we treat designs consisting of s independent factors, including $s \leq 3$, as designs of resolution at least IV.

For a given design $S \subset H_q$ and a $\gamma \in H_q$, define

$$B_i(S, \gamma) = \#\{(d_1, d_2, \ldots, d_i) : d_1, d_2, \ldots, d_i \in S, d_1 d_2 \cdots d_i = \gamma\},$$

where # denotes the cardinality of a set, $d_1, \ldots, d_i$ are different columns in S, and $d_1 d_2 \cdots d_i$ is the ith order interaction of $d_1, d_2, \ldots, d_i$. By this definition, $B_i(S, \gamma)$ is the number of ith order interactions in S aliased with γ. An equivalent definition can be found in (2.5) of Zhang and Mukerjee (2009a). For example, consider $q = 3$, $S = \{1, 2, 3, 12, 23\}$, $i = 2$, $\gamma_1 = 12 \in S$, and $\gamma_2 = 123 \in H_q \backslash S$. Among the 10 2fis in S, there is one 2fi (between 1 and 2) and two 2fis (between

1 and 23, and between 3 and 12) that are aliased with γ_1 and γ_2, respectively. Hence, $B_2(S,\gamma_1) = 1$ and $B_2(S,\gamma_2) = 2$.

It is also useful to take

$$\bar{g}(S) = \#\{\gamma : \gamma \in H_q \backslash S, B_2(S,\gamma) > 0\}, \tag{2.3}$$

the number of main effects in $H_q\backslash S$ aliased with at least one 2fi of S. Note that the $g(S)$ defined in Zhang and Mukerjee (2009a) is represented as $\bar{g}(H_q\backslash S)$ here. Zhang and Mukerjee (2009a) observed that minimizing $g(H_q\backslash S)$ is an important step in the search for a GMC design. The following theorem describes the structure of designs that minimize $\bar{g}(S)$ and plays a key role in the construction of GMC designs.

Theorem 1. *Let $S \subset H_q$ be a design with s factors (columns). Under isomorphism, we have*

(a) *if $2^{r-1} \le s \le 2^r - 1$ for some $r \le q$ and $\bar{g}(S)$ is minimized among all the designs with s factors, then S has r independent factors and $S \subset H_r$;*

(b) *if $2^{r-2} + 1 \le s \le 2^{r-1}$ for some $r \le q$ and $\bar{g}(S)$ is minimized among all the designs with s factors and resolution at least IV, then S has r independent factors and $S \subset F_{qr}$ (or T_r);*

(c) *if $2^{r-2} + 1 \le s \le 2^{r-1}$ for some $r \le q$, then S sequentially maximizes the components of*

$$\{-\bar{g}(S),\ {}_2^{\#}C_2(S)\} \tag{2.4}$$

among all the designs with s factors and resolution at least IV if and only if S is any one of four isomorphic designs: that consisting of the first s columns of F_{qr}; that consisting of the last s columns of F_{qr}; that consisting of the first s columns of T_r; that consisting of the last s columns of T_r. Here F_{qr} and T_r have the Yates order.

An outline of the proof for Theorem 1 is given in the Appendix. A full proof of the theorem is available as a supplement at the journal website `http://www.stat.sinica.edu.tw/statistica`.

In the sequel it is helpful to employ several abbreviations. The statement "a design sequentially maximizes the components of the sequence" is shortened to "a design maximizes the sequence". And "$\bar{g}(S)$ is minimized among all the designs with s factors" is reduced to "$\bar{g}(S)$ is minimized". We also suppress the phrase "up to isomorphism", since all the isomorphic designs are viewed as equivalent.

3. A Theory on Constructing GMC 2^{n-m} Designs

In this section, let D be a 2^{n-m} regular fractional factorial design. It is convenient to develop the theory separately for two subsets of pairs (n, N): first $5N/16 + 1 \leq n < N/2$, then $n \geq N/2$.

3.1. GMC 2^{n-m} designs with $5N/16 + 1 \leq n \leq N/2$

Theorem 1 in Zhang et al. (2008) has shown that a GMC design has maximum resolution. Since for $5N/16 + 1 \leq n \leq N/2$, the maximum resolution of 2^{n-m} designs is IV, the resolution of GMC designs here is also IV. Moreover, by the results of Bruen, Haddad and Wehlau (1998) and Butler (2007), any 2^{n-m}_{IV} design D with $5N/16 + 1 \leq n \leq N/2$ must be taken from F_{qq}, i.e., $D \subset F_{qq}$. In this case the number of factors in $F_{qq} \backslash D$, which is $N/2 - n$, is less than that of D, which is n.

To study the construction of GMC designs, let us first investigate the relationships between the AENP of D and that of $F_{qq} \backslash D$. We have the following result.

Lemma 1. *Let $D \subset F_{qq}$ be a 2^{n-m} design with $q \geq 4$ and $n > N/4$. Then*

(a) $$B_2(D, \gamma) = \begin{cases} 0, & \text{if } \gamma \in F_{qq}, \\ B_2(F_{qq} \backslash D, \gamma) + n - \frac{N}{4}, & \text{if } \gamma \in H_{q-1}, \end{cases}$$

(b) $$ {}_1^{\#}C_2^{(k)}(D) = \begin{cases} n, & \text{if } k = 0, \\ 0, & \text{if } k \geq 1, \end{cases}$$

(c) $$ {}_2^{\#}C_2^{(k)}(D) = \begin{cases} 0, & \text{if } k < n - \frac{N}{4} - 1, \\ -(k+1)\bar{g}(F_{qq} \backslash D) + (k+1)(\frac{N}{2} - 1), & \text{if } k = n - \frac{N}{4} - 1, \\ \frac{k+1}{k+1-(n-N/4)} \, {}_2^{\#}C_2^{(k-n+N/4)}(F_{qq} \backslash D), & \text{if } k \geq n - \frac{N}{4}. \end{cases}$$

Proof. Recall that $B_2(D, \gamma)$ is the number of 2fis in D aliased with γ. From the structure of F_{qq}, any $\gamma \in F_{qq}$ is not aliased with the 2fis in F_{qq}. The first equality of (a) and two equalities of (b) follow.

For the second equality of (a), first note that for any $\gamma \in H_{q-1}$ there are $N/4$ pairs of factors in F_{qq} such that the interaction formed by each pair is aliased with γ. These $N/4$ pairs can be partitioned into three groups: $B_2(D, \gamma)$ with both factors from D, $B_2(F_{qq} \backslash D, \gamma)$ with both factors from $F_{qq} \backslash D$, and $n - 2B_2(D, \gamma)$ with one factor from D and the other from $F_{qq} \backslash D$. Therefore

$$B_2(D, \gamma) + B_2(F_{qq} \backslash D, \gamma) + n - 2B_2(D, \gamma) = \frac{N}{4},$$

which implies the second equality of (a).

For (c), note that part (a) and the definition of ${}_2^{\#}C_2^{(k)}(D)$ together imply

$$\begin{aligned}{}_2^{\#}C_2^{(k)}(D) &= (k+1)\#\{\gamma : \gamma \in H_q, B_2(D,\gamma) = k+1\} \\ &= (k+1)\#\{\gamma : \gamma \in H_{q-1}, B_2(F_{qq}\backslash D,\gamma) = k+1-(n-\frac{N}{4})\}.\end{aligned}$$

The first and third equalities in (c) follow directly from this and the definition of ${}_2^{\#}C_2^{(k)}(F_{qq}\backslash D)$. To get the second equality, use part (a), $k = n - N/4 - 1$, and $\#\{H_{q-1}\} = N/2 - 1$ to obtain

$$\begin{aligned}{}_2^{\#}C_2^{(k)}(D) &= (k+1)\#\{\gamma : \gamma \in H_{q-1}, B_2(F_{qq}\backslash D,\gamma) = 0\} \\ &= (k+1)(\frac{N}{2}-1) - (k+1)\#\{\gamma : \gamma \in H_{q-1}, B_2(F_{qq}\backslash D,\gamma) > 0\} \\ &= (k+1)(\frac{N}{2}-1) - (k+1)\#\{\gamma : \gamma \in H_q\backslash(F_{qq}\backslash D), B_2(F_{qq}\backslash D,\gamma) > 0\} \\ &= (k+1)(\frac{N}{2}-1) - (k+1)\bar{g}(F_{qq}\backslash D).\end{aligned}$$

Lemma 1 implies that maximizing the first two terms $\{{}_1^{\#}C_2(D), {}_2^{\#}C_2(D)\}$ of the sequence (2.2) is equivalent to maximizing the sequence $\{-\bar{g}(F_{qq}\backslash D), {}_2^{\#}C_2(F_{qq}\backslash D)\}$. We therefore have the following result.

Lemma 2. *Suppose (n, N) satisfies $5N/16 + 1 \le n \le N/2$. Consider the family of 2^{n-m} designs D with $D \subset F_{qq}$. If there is a unique design in this family that maximizes*

$$\{-\bar{g}(F_{qq}\backslash D), {}_2^{\#}C_2(F_{qq}\backslash D)\} \tag{3.1}$$

then this design has GMC.

Combining Lemma 2 and Part (c) of Theorem 1, we have the following result.

Theorem 2. *Suppose the columns in H_q and F_{qq} are written in the Yates order. For $5N/16 + 1 \le n \le N/2$, the GMC 2^{n-m} design is the design that consists of the last n columns in H_q or F_{qq}.*

Proof. Suppose $2^{r-2} + 1 \le N/2 - n \le 2^{r-1}$ for some r. By applying part (c) of Theorem 1 with $S = F_{qq}\backslash D$ and $s = N/2 - n$, we observe that the design $F_{qq}\backslash D$ consisting of the first $N/2 - n$ columns of F_{qr} uniquely maximizes the sequence (3.1). When H_q and F_{qq} are written in the Yates order, the first $N/2-n$ columns of F_{qr} are also the first $N/2 - n$ columns of F_{qq}. Consequently, the GMC design D consists of the last n columns of F_{qq}, which are the same as the last n columns of H_q.

The following example illustrates how Theorem 2 can be used to construct a GMC design.

Example 1. Suppose that we require a GMC design with $N = 32$ runs and n factors where $11{=}5/16N + 1 \leq n \leq N/2 = 16$. The design F_{55} with the Yates order can be written as

$$\begin{aligned} F_{55} &= \{5, 5H_4\} \\ &= \{5, 15, 25, 125, 35, 135, 235, 1235, 45, 145, 245, 1245, 345, 1345, 2345, 12345\}, \end{aligned}$$

where the columns of F_{55} correspond to the last 16 columns of H_5.

According to Theorem 2, to get a 2^{n-m} design with $n-m = 5$, we only need to take the last n columns (or delete the first $16 - n$ columns) from F_{55}. For example, by taking the last 13 columns of F_{55} we obtain the GMC 2^{13-8} design:

$$\begin{aligned} D_3 &= F_{55} \setminus \{5, 15, 25\} \\ &= \{125, 35, 135, 235, 1235, 45, 145, 245, 1245, 345, 1345, 2345, 12345\}, \end{aligned}$$

where these 13 columns are assigned to the main effects of the 13 factors, and the 32 rows are the factor combinations in the design. By taking the last 12 columns of F_{55}, we get the GMC 2^{12-7} design:

$$\begin{aligned} D_4 &= F_{55} \setminus \{5, 15, 25, 125\} \\ &= \{35, 135, 235, 1235, 45, 145, 245, 1245, 345, 1345, 2345, 12345\}. \end{aligned}$$

Similarly, these 12 columns are assigned to the main effects of the 12 factors, and the 32 rows are the factor combinations in the design.

3.2. GMC 2^{n-m} designs with $n \geq N/2$

In this subsection, we still use D to denote the regular 2^{n-m} fractional factorial design and $q = n - m$. Recall that $S_{qr} = H_q \backslash H_r$. Zhang and Mukerjee (2009a) showed that, if $n \geq N/2$ and D has GMC, then $\bar{g}(H_q \backslash D)$ is minimized. According to Part (a) of Theorem 1, if the number of columns $N-1-n$ in $H_q \backslash D$, satisfies $2^{r-1} \leq N - 1 - n \leq 2^r - 1$ for some r, then $H_q \backslash D$ has r independent factors. Therefore $H_q \backslash D \subset H_r$ and $S_{qr} \subset D$.

If $N - 1 - n = 2^r - 1$, which equals the number of columns in H_r, then $H_q \backslash D = H_r$ and $D = S_{qr}$ has GMC.

If $2^{r-1} \leq N-1-n < 2^r - 1$, then it is convenient to use $D \backslash S_{qr}$ to construct GMC designs because the number of columns in $D \backslash S_{qr}$ is much smaller than that in D. For example, consider the construction of a GMC 2^{10-6} design. Here $n = 10$, $m = 6$, $q = n-m = 4$, $N = 2^q = 16$, and $r = 3$ since $2^{3-1} \leq N-1-n = 5 \leq 2^3 - 1$. The GMC 2^{10-6} design D can be partitioned as $D = S_{43} \cup (D \backslash S_{43})$ where $S_{43} = H_4 \backslash H_3$ has $(16-1) - (8-1) = 8$ columns and $D \backslash S_{43}$ has just 2 columns. By choosing two columns from H_3 according to rules developed later in Lemma 5, we can construct the GMC 2^{10-6} design. We will also show that, if $D\backslash S_{43} = \{23, 123\}$, then the resulting design

$$D = S_{43} \cup \{23, 123\} = \{23, 123, 4, 14, 24, 124, 34, 134, 234, 1234\}$$

has GMC.

Next, we study a connection between $B_2(D, \gamma)$ and $B_2(D\backslash S_{qr}, \gamma)$ that sheds light on the relationship between the AENPs of D and $D\backslash S_{qr}$.

Lemma 3. *Suppose D is a 2^{n-m} design with $S_{qr} \subset D$.*

(a) *If $\gamma \in S_{qr}$, then $B_2(D, \gamma) = n - N/2$.*

(b) *If $\gamma \in H_r$, then $B_2(D, \gamma) = B_2(D\backslash S_{qr}, \gamma) + N/2 - 2^{r-1}$.*

Proof. Again note that $B_2(D, \gamma)$ is the number of 2fis in D aliased with γ.

(a) For any $\gamma = d_1d_2 \in S_{qr}$, there are two possibilities: either both d_1 and d_2 are in S_{qr}, or d_1 and d_2 are respectively in $D\backslash S_{qr}$ and S_{qr}. Therefore

$$\begin{aligned} B_2(D, \gamma) = &\#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in D\backslash S_{qr},\ d_2 \in S_{qr}\} \\ &+\#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in S_{qr},\ d_2 \in S_{qr}\}. \end{aligned}$$

For any $d_1 \in D\backslash S_{qr}$ we can uniquely determine $d_2 = d_1\gamma$ in S_{qr}, so

$$\#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in D\backslash S_{qr},\ d_2 \in S_{qr}\} = n - (N - 2^r).$$

For any $\gamma \in S_{qr}$, there are $N/2 - 1$ pairs of factors in H_q whose interaction is aliased with γ. Among these pairs, there are $2^r - 1$ with one factor from H_r and another one from S_{qr}; for the remaining $N/2 - 2^r$ pairs, both factors are from S_{qr}. Part (a) follows from

$$\begin{aligned} \#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in S_{qr},\ d_2 \in S_{qr}\} &= \frac{N}{2} - 2^r, \\ B_2(D, \gamma) = n - (N - 2^r) + \frac{N}{2} - 2^r &= n - \frac{N}{2}. \end{aligned}$$

(b) For any $\gamma = d_1d_2 \in H_r$, there are two possibilities: both d_1 and d_2 are in $D\backslash S_{qr}$ or both are in S_{qr}. Now we have

$$\begin{aligned} B_2(D, \gamma) &= \#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in D\backslash S_{qr},\ d_2 \in D\backslash S_{qr}\} \\ &\quad +\#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in S_{qr},\ d_2 \in S_{qr}\} \\ &= B_2(D\backslash S_{qr}, \gamma) + \#\{(d_1, d_2) : \gamma = d_1d_2,\ d_1 \in S_{qr},\ d_2 \in S_{qr}\}, \end{aligned}$$

where the second equality is from the definition of $B_2(D\backslash S_{qr}, \gamma)$. For any $\gamma \in H_r$, there are $N/2-1$ pairs of factors in H_q whose interaction is aliased with γ. Among these pairs, $(2^r-2)/2 = 2^{r-1}-1$ are from H_r and $N/2-2^{r-1}$ are from S_{qr}. Part (b) follows from

$$\#\{(d_1,d_2):\gamma = d_1d_2,\ d_1\in S_{qr},\ d_2\in S_{qr}\} = \frac{N}{2}-2^{r-1},$$
$$B_2(D,\gamma) = B_2(D\backslash S_{qr},\gamma)+\frac{N}{2}-2^{r-1}.$$

Zhang and Mukerjee (2009a) studied the connection between $B_2(D,\gamma)$ and $B_2(H_q\backslash D,\gamma)$ in their Theorem 1, and applied this connection to investigate the relationship between the AENPs of D and $H_q\backslash D$. Lemma 3 describes the connection between $B_2(D,\gamma)$ and $B_2(D\backslash S_{qr},\gamma)$. The relationship between the leading terms of AENPs of D and $D\backslash S_{qr}$ can be obtained via Lemma 3 and the following.

Lemma 4. *Suppose $D=\{S_{qr}, D\backslash S_{qr}\}$.*

(a) $${}_1^{\#}C_2^{(k)}(D) = \begin{cases} constant, & if\ k<\frac{N}{2}-2^{r-1},\\ {}_1^{\#}C_2^{(k-N/2+2^{r-1})}(D\backslash S_{qr})+constant, & if\ k\ge\frac{N}{2}-2^{r-1},\end{cases}$$

(b) $${}_2^{\#}C_2^{(k)}(D) = \begin{cases} constant, & if\ k<\frac{N}{2}-2^{r-1}-1,\\ -(k+1)\bar{g}(D\backslash S_{qr})+(k+1)\,{}_1^{\#}C_2^{(0)}(D\backslash S_{qr}) & \\ \ +constant, & if\ k=\frac{N}{2}-2^{r-1}-1,\\ \frac{k+1}{k-N/2+2^{r-1}+1}\,{}_2^{\#}C_2^{(k-N/2+2^{r-1})}(D\backslash S_{qr}) & \\ \ +constant, & if\ k\ge\frac{N}{2}-2^{r-1},\end{cases}$$

where the constants are non-negative values depending only on n, k, and N.

Proof. (a) From the definition of ${}_1^{\#}C_2^{(k)}(D)$, we have

$${}_1^{\#}C_2^{(k)}(D) = \#\{\gamma:\gamma\in S_{qr}, B_2(D,\gamma)=k\}+\#\{\gamma:\gamma\in D\backslash S_{qr}, B_2(D,\gamma)=k\}.$$

Part (a) follows from an application of Lemma 3:

$${}_1^{\#}C_2^{(k)}(D) = I(n-\frac{N}{2}=k)\times(N-2^r)$$
$$+\#\{\gamma:\gamma\in D\backslash S_{qr}, B_2(D\backslash S_{qr},\gamma)+\frac{N}{2}-2^{r-1}=k\},$$

where $I(\cdot)$ is the indicator function.

(b) By the definition of ${}_2^{\#}C_2^{(k)}(D)$, we have

$${}_2^{\#}C_2^{(k)}(D) = (k+1)\#\{\gamma:\gamma\in S_{qr}, B_2(D,\gamma)=k+1\}$$
$$+(k+1)\#\{\gamma:\gamma\in H_r, B_2(D,\gamma)=k+1\}.$$

Applying Lemma 3, this reduces to

$$\begin{aligned}{}_2^{\#}C_2^{(k)}(D) = I(n - \frac{N}{2} = k+1)\times(k+1)(N-2^r)\\ +(k+1)\#\{\gamma : \gamma \in H_r, B_2(D\backslash S_{qr},\gamma) = k+1-\frac{N}{2}+2^{r-1}\}.\end{aligned}$$

The first and third expressions of (b) follow directly from this and the definition of ${}_2^{\#}C_2^{(k)}(D\backslash S_{qr})$.

To derive the second expression of (b), put $k = N/2 - 2^{r-1} - 1$. We have

$$\begin{aligned}{}_2^{\#}C_2^{(k)}(D) &= (k+1)\#\{\gamma : \gamma \in H_r, B_2(D\backslash S_{qr},\gamma) = 0\}+constant\\ &= (k+1)\#\{\gamma : \gamma \in D\backslash S_{qr}, B_2(D\backslash S_{qr},\gamma) = 0\}\\ &\quad +(k+1)\#\{\gamma : \gamma \in H_q\backslash D, B_2(D\backslash S_{qr},\gamma) = 0\}+constant.\end{aligned}$$

From the definition of ${}_1^{\#}C_2^{(k)}(D\backslash S_{qr})$ and $\bar{g}(\cdot)$ in (2.3), we obtain

$$\#\{\gamma : \gamma \in D\backslash S_{qr}, B_2(D\backslash S_{qr},\gamma) = 0\} = {}_1^{\#}C_2^{(0)}(D\backslash S_{qr})$$

and

$$\begin{aligned}&\#\{\gamma : \gamma \in H_q\backslash D, B_2(D\backslash S_{qr},\gamma) = 0\}\\ &\quad = (N-1-n) - \#\{\gamma : \gamma \in H_q\backslash D, B_2(D\backslash S_{qr},\gamma) > 0\}\\ &\quad = (N-1-n) - \#\{\gamma : \gamma \in S_{qr} \cup (H_q\backslash D), B_2(D\backslash S_{qr},\gamma) > 0\}\\ &\quad = (N-1-n) - \#\{\gamma : \gamma \in H_q\backslash(D\backslash S_{qr}), B_2(D\backslash S_{qr},\gamma) > 0\}\\ &\quad = (N-1-n) - \bar{g}(D\backslash S_{qr}).\end{aligned}$$

The second equality above follows from the structure of S_{qr} and $D\backslash S_{qr}$. A rearrangement of terms yields the desired result.

The following lemma can be used to construct GMC designs when the number of factors in $D\backslash S_{qr}$ is small.

Lemma 5. *Suppose (n, N) satisfies $2^{r-1} \le N-1-n \le 2^r-1$ for some $r \le q-1$. Consider the family of 2^{n-m} designs D with $S_{qr} \subset D$. If there is a unique design in this family that maximizes*

$$\{\,{}_1^{\#}C_2(D\backslash S_{qr}),\ -\bar{g}(D\backslash S_{qr}),\ {}_2^{\#}C_2(D\backslash S_{qr})\}, \tag{3.2}$$

then it has GMC.

Proof. The result follows directly from Lemma 4.

If $2^{r-1} \le N-1-n \le 2^r-1$, then there are r independent factors in H_r and $n+2^r-N$ ($< 2^{r-1}$) factors in $D\backslash S_{qr}$. We can thus find a design with resolution at least IV, and with $n+2^r-N$ factors in H_r. Note that ${}_1^{\#}C_2(D\backslash S_{qr})$ is maximized if $D\backslash S_{qr}$ has resolution at least IV. The two terms following ${}_1^{\#}C_2(D\backslash S_{qr})$ in the AENP are $-\bar{g}(D\backslash S_{qr})$ and ${}_2^{\#}C_2(D\backslash S_{qr})$. Applying Part (c) of Theorem 1, we obtain a result similar to Theorem 2.

Theorem 3. *Suppose the columns in H_q are written in the Yates order. For $n \geq N/2$, the GMC 2^{n-m} design is the design that consists of the last n columns in H_q.*

Proof. Suppose $2^{r-1} \leq N-1-n \leq 2^r-1$ for some $r \leq q-1$, and $S_{qr} \subset D$. Let $f_r = n+2^r-N$, the number of columns in $D\backslash S_{qr}$. We then have $0 \leq f_r \leq 2^{r-1}-1$.

If $f_r = 0$ or 1, then $D = S_{qr}$ or $S_{qr} \cup \{12\cdots r\}$, and the result is obvious. Next suppose $2^{l-2}+1 \leq f_r \leq 2^{l-1}$ for some $2 \leq l \leq r$. Let $S = D\backslash S_{qr}$, $s = f_r$ and apply Part (c) of Theorem 1. If $D\backslash S_{qr}$ consists of the first f_r columns of T_l, then $D\backslash S_{qr}$ uniquely maximizes the sequence (3.2). Here T_l is defined in Section 2. When H_q is written in the Yates order, the design consisting of the first f_r columns of T_l is isomorphic to the one consisting of the first f_r columns of T_r. Part (c) of Theorem 1 shows that the design consisting of the first f_r columns of T_r is isomorphic to the one consisting of the last f_r columns of T_r. Therefore, if $D\backslash S_{qr}$ consists of the last f_r columns of T_r, then $D\backslash S_{qr}$ uniquely maximizes the sequence (3.2) under isomorphism. Combining the last f_r columns of T_r with S_{qr}, we can conclude that the design consisting of the last n columns of H_q has GMC.

Here is an example that illustrates the construction method in Theorem 3.

Example 2. Suppose we want a GMC design with 32 runs and more than 16 factors, i.e., $N = 32$ and $n > 16$. The design H_5 with the Yates order can be written as

$$H_5 = \{1, 2, 12, 3, 13, 23, 123, 4, 14, 24, 124, 34, 134, 234, 1234\} \cup F_{55}.$$

Here, F_{55} is given in Example 1.

According to Theorem 3, if we take the last n columns or delete the first $31-n$ columns from H_5, then we get the GMC 2^{n-m} design. For example, taking the last 20 columns or deleting the first 11 columns from H_5, we get the GMC 2^{20-15} design

$$D_5 = \{34,134,234,1234,5,15,25,125,35,135,235,1235,45,145,245,1245,345,1345,\\ 2345,12345\}.$$

These 20 columns are assigned to the main effects of the 20 factors, and the 32 rows are the factor combinations in the design.

4. When Do the MA and GMC Designs Differ?

Since the MA and GMC criteria differ, it is natural to ask: under what circumstances do the criteria yield different designs? The two theorems presented below provide a partial answer.

Theorem 4. *Consider 2_{IV}^{n-m} designs $D \subset F_{qq}$ with $5N/16+1 \leq n \leq N/2$. If $n \leq N/2-4$ and a design has MA, then it is not possible to maximize ${}_2^{\#}C_2(D)$, so the MA design cannot have GMC.*

Proof. When $5N/16+1 \leq n \leq N/2$, Butler (2003) proved that if D is an MA design, then $D \subset F_{qq}$ and $F_{qq}\backslash D$ has MA among the designs in F_{qq}. So the number of independent factors in $F_{qq}\backslash D$ is $\min(N/2-n, q)$.

According to Part (c) of Lemma 1, if a design $D \subset F_{qq}$ and ${}_2^{\#}C_2(D)$ is maximized, then $\bar{g}(F_{qq}\backslash D)$ is minimized. Due to the structure of F_{qq}, the design $F_{qq}\backslash D$ has resolution at least IV. Utilizing Part (b) of Theorem 1, we find that the maximum number of independent factors in $F_{qq}\backslash D$ is at most $\lfloor \log_2(N/2-n-1)\rfloor+2$, where $\lfloor x\rfloor$ is the integer part of x. When $5N/16+1 \leq n \leq N/2$, we have

$$\lfloor \log_2(\frac{N}{2}-n-1)\rfloor+2 \leq \lfloor \log_2(\frac{N}{2}-\frac{5N}{16}-2)\rfloor+2 = \lfloor \log_2(\frac{3N}{16}-2)\rfloor+2 < q.$$

However, for $n \leq N/2-4$, $\lfloor \log_2(N/2-n-1)\rfloor+2 < N/2-n$. Therefore if the MA 2^{n-m} design D with $n \leq N/2-4$ maximizes ${}_2^{\#}C_2$, then the number of independent factors in the $F_{qq}\backslash D$ must be less than $\min(N/2-n, q)$. This contradiction establishes the conclusion in Theorem 4.

The next example is used to illustrate the result in Theorem 4.

Example 3. Consider the case when $N=32$ and $n=12 \leq 32/2-4$. According to Butler (2003), the MA 2^{12-7} design is isomorphic to

$$\begin{aligned} D_6 &= \{125, 135, 235, 1235, 45, 145, 245, 1245, 345, 1345, 2345, 12345\} \\ &= F_{55}\backslash\{5, 15, 25, 35\}. \end{aligned}$$

Note that the design D_6 is obtained by deleting the four columns $\{5, 15, 25, 35\}$ from F_{55} and the deleted four columns are independent. From Example 1, $D_4 = F_{55}\backslash\{5, 15, 25, 125\}$ is the GMC 2^{12-7} design up to isomorphism. However, the deleted four columns $\{5, 15, 25, 125\}$ are not independent. Thus the MA design D_6 is different from D_4 and does not have GMC.

The following is a second result identifying conditions where MA and GMC differ, but now with $n \geq N/2$.

Theorem 5. *Suppose $2^{r-1} \le N-1-n \le 2^r-1$ for some r and $n \ge N/2$. If $4 \le n+2^r-N \le 2^{r-1}-4$ with $4 \le r \le q-1$, then the families of MA and GMC designs are mutually exclusive.*

Proof. Using Lemma 4 of Chen and Hedayat (1996), Butler (2003) proved that, if $n \ge N/2$ and D is an MA design, then $F_{qq} \subset D$ and $D \backslash F_{qq}$ has MA among the designs in H_{q-1}. By repeatedly applying this result and Lemma 4 of Chen and Hedayat (1996), we can prove the following stronger result: if D has MA, then $S_{qr} \subset D$ and $D \backslash S_{qr}$ has MA among the designs in H_r, and hence the number of independent factors in $D \backslash S_{qr}$ is $\min(n+2^r-N, r)$.

According to Lemma 5 and the discussion before Theorem 3, if D has GMC, then $S_{qr} \subset D$, $D \backslash S_{qr}$ has resolution at least IV, and $\bar{g}(D \backslash S_{qr})$ is minimized. By Part (b) of Theorem 1, the number of independent factors in $D \backslash S_{qr}$ is at most $\lfloor \log_2(n+2^r-N-1) \rfloor + 2$. However, for $4 \le n+2^r-N \le 2^{r-2}$ with $r \ge 4$, we can easily check that

$$\lfloor \log_2(n+2^r-N-1) \rfloor + 2 < \min(n+2^r-N, r),$$

which means that, in this region every GMC design differs from an MA design. When $2^{r-2}+1 \le n+2^r-N \le 2^{r-1}-4$ with $r \ge 5$, from Part (b) of Theorem 1 there are r independent factors in $D \backslash S_{qr}$ and $D \backslash S_{qr} \subset T_r$. By Lemma 1 (a) with q replaced by r, and assuming $2^{r-2}+1 \le n+2^r-N \le 2^{r-1}-4$, for any $\gamma \in H_{r-1}$ we have

$$B_2(D \backslash S_{qr}, \gamma) = B_2(T_r \backslash (D \backslash S_{qr}), \gamma) + n + 2^r - N - 2^{r-2} \ge 1,$$

and therefore $\bar{g}(D \backslash S_{qr}) = 2^{r-1}-1$, which is a constant. So if D has GMC, then $D \backslash S_{qr} \subset T_r$ and ${}_2^{\#}C_2(D \backslash S_{qr})$ is maximized. If D is also an MA design, then $D \backslash S_{qr}$ also has MA. However, as in the proof of Theorem 4, we can show that $D \backslash S_{qr}$ does not have MA among all the designs taken from H_r. Consequently, if D has GMC then it cannot have MA. This completes the proof.

Zhang and Mukerjee (2009a) found that, for $N-1-n = 11$, the GMC and MA designs are different. This result is a special case of Theorem 5 with $r = 4$. The following example provides an application of Theorem 5.

Example 4. Suppose $N = 32$ and $n = 20$. According to Butler (2003), the MA 2^{20-15} design is isomorphic to

$$D_7 = \{124, 134, 234, 1234\} \cup F_{55}.$$

Note that the four columns $\{124, 134, 234, 1234\}$ joined into the design D_7 are independent. Now, for the GMC 2^{20-15} design D_5 in Example 2, the four columns $\{34, 134, 234, 1234\}$ joined into D_5 with the same F_{55} are not independent. As a result, among regular 2^{20-15} designs, the GMC and MA designs are different.

Acknowledgement

The authors would like to thank an associate editor and two referees for constructive suggestions and comments that led to a significant improvement of the article. Also, we express our special thanks to Professor Peter Hooper at the University of Alberta who helped us edit the paper. This work was supported by the Natural Sciences and Engineering Research Council of Canada, the start-up grant of University of Alberta, the NNSF of China grant Nos. 10871104, 10901092, 10771123, and 10826059, and the NSF of Shandong Province of China grant No. Q2007A05.

Appendix: Outline of the Proof for Theorem 1.

For Part (a).

When $r = q$, Part (a) can be validated directly. We only need to consider the case $r \le q-1$. The main idea of the proof is as follows. Suppose that $S_1 \subset H_q$ is a design with s factors, where $2^{r-1} \le s \le 2^r - 1$ for some $r \le q-1$. If S_1 has $h+1$ ($r \le h \le q-1$) independent factors, then we can find a design S_1^* with h independent factors such that $\bar{g}(S_1^*) < \bar{g}(S_1)$. The proof consists of three steps.

Let a denote the factor q. Under isomorphism, we assume $a \in S_1$ and S_1 can be represented as

$$S_1 = Q \cup \{a, ab_1, ab_2, \ldots, ab_l\},$$

where $Q \subset H_h$ and has h independent factors, and $\{b_1, \ldots, b_l\} \subset H_h$. Without loss of generality, we assume that $\{b_1, \ldots, b_t\} \subset Q$ and $\{b_{t+1}, \ldots, b_l\} \subset H_h \backslash Q$, and let

$$S_2 = Q \cup \{a, ab_1, \ldots, ab_t\} \cup \{b_{t+1}, \ldots, b_l\}.$$

In Step 1, we prove that $\bar{g}(S_2) \le \bar{g}(S_1)$. The details are in Lemma 6 of the supplement.

Next, we join Q and $\{b_{t+1}, \ldots, b_l\}$ together and still denote it by Q. Then S_2 has the form

$$S_2 = Q \cup \{a, ab_1, \ldots, ab_t\},$$

where $Q \subset H_h$ and has h independent factors, and $\{b_1, \ldots, b_t\} \subset Q$. When $2^{r-1} \le s \le 2^r - 1$, the number of factors in Q is smaller than $2^r - 1$. Therefore there are at least two factors c_1 and c_2 in Q such that $c = c_1c_2 \notin Q$. Under isomorphism, we assume that there is some t_0 such that

$$\{c, cb_1, cb_2, \ldots, cb_{t_0}\} \subset H_h \backslash Q \quad \text{and} \quad \{cb_{t_0+1}, \ldots, cb_t\} \subset Q.$$

Let

$$S_3 = Q \cup \{c, cb_1, cb_2, \ldots, cb_{t_0}\} \cup \{ab_{t_0+1}, \ldots, ab_t\}.$$

In Step 2, we show that $\bar{g}(S_3) \leq \bar{g}(S_2)$, especially, $\bar{g}(S_3) < \bar{g}(S_2)$ if $t_0 = t$. The details are in Lemma 7 of the supplement.

In Step 3, we repeat the same process from S_2 to S_3 until $t_0 = t$, i.e., $S_3 \subset H_h$. Then S_3 has h independent factors and $\bar{g}(S_3) < \bar{g}(S_2) \leq \bar{g}(S_1)$. Thus, Part (a) is proved.

For Part (b).

The idea of the proof for the first half of Part (b) is similar to that for Part (a). Suppose that S is a design with resolution at least IV that has s factors, and $2^{r-2} + 1 \leq s \leq 2^{r-1}$ for some $r \leq q$. If S has $h+1$ $(r \leq h \leq q-1)$ independent factors, then we can construct a resolution IV design S^* with s factors and h independent factors such that $\bar{g}(S^*) < \bar{g}(S)$. The details are in Lemma 8 of the supplement.

For the proof of the second half of Part(b), by Butler (2003), it suffices to show that $A_i(S) = 0$ for all odd numbered i's. Under the assumptions in Part (b), if $A_i(S) \neq 0$ for some odd number i, then $A_5(S) > 0$, for details see Lemma 9 of the supplement. With this result, we could find a design $S^* \subset F_{qr}$ such that $\bar{g}(S^*) < \bar{g}(S)$, which is a contradiction to the assumption that $\bar{g}(S)$ is minimized. Then Part (b) follows.

For Part (c).

First, we show that the four designs consisting of the first or last s columns of F_{qr} or T_r are isomorphic by observing the structure of F_{qr} or T_r. Next, it suffices to show that the design S consisting of the first s columns of F_{qr} is the unique one which maximizes (2.4). Mathematical induction is used to prove this result.

References

Ai, M. Y. and Zhang, R. C. (2004). s^{n-m} designs containing clear main effects or clear two-factor interactions. *Statist. Probab. Lett.* **69**, 151-160.

Bruen, A. A., Haddad, L. and Wehlau, D. L. (1998). Binary codes and caps. *J. Combinat. Des.* **6**, 275-284.

Butler, N. A. (2003). Some theory for constructing minimum aberration fractional factorial designs. *Biometrika* **90**, 233-238.

Butler, N. A. (2007). Results for two-level fractional factorial designs of resolution IV or more. *J. Statist. Plann. Inference* **137**, 317-323.

Chen, B. J., Li, P. F., Liu, M. Q. and Zhang, R. C. (2006). Some results on blocked regular 2-level fractional factorial designs with clear effects. *J. Statist. Plann. Inference* **136**, 4436-4449.

Chen, H. and Cheng, C. -S. (2006). Doubling and projection: A method of constructing two level designs of resolution IV. *Ann. Statist.* **34**, 546-558.

Chen, H. and Hedayat, A. S. (1996). 2^{n-l} designs with weak minimum aberration. *Ann. Statist.* **24**, 2536-2548.

Chen, H. and Hedayat, A. S. (1998). 2^{n-m} designs with resolution III and IV containing clear two-factor interactions. *J. Statist. Plann. Inference* **75**, 147-158.

Chen, J., Sun, D. X. and Wu, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *Internat. Statist. Rev.* **61**, 131-145.

Chen, J. and Wu, C. F. J. (1991). Some results on s^{n-k} fractional factorial designs with minimum aberration or optimal moments. *Ann. Statist.* **19**, 1028-1041.

Cheng, C. -S. and Mukerjee, R. (1998). Regular fractional factorial designs with minimum aberration and maximum estimation capacity. *Ann. Statist.* **26**, 2289-2300.

Cheng, C. -S., Steinberg, D. M. and Sun, D. X. (1999). Minimum aberration and model robustness for two-level factorial designs. *J. Roy. Statist. Soc. B* **61**, 85-93.

Cheng, C. -S. and Tang, B. (2005). A general theory of minimum aberration and its applications. *Ann. Statist.* **33**, 944-958.

Franklin, M. F. (1984). Constructing tables of minimum aberration p^{n-m} designs. *Technometrics* **26**, 225-232.

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

Li, P. F., Chen, B. J., Liu, M. Q. and Zhang, R. C. (2006). A note on minimum aberration and clear criteria. *Statist. Probab. Lett.* **76**, 1007-1011.

Mukerjee, R. and Wu, C. F. J. (2006). *A Modern Theory of Factorial Designs.* Springer Series in Statistics, Springer, New York.

Sun, D. X. (1993). Estimation capacity and related topics in experimental designs. PhD dissertation. University of Waterloo, Waterloo.

Tang, B., Ma, F., Ingram, D. and Wang, H. (2002). Bounds on the maximum number of clear two-factor interactions for 2^{m-p} designs of resolution III and IV. *Canad. J. Statist.* **30**, 127-136.

Tang, B. and Wu, C. F. J. (1996). Characterization of minimum aberration 2^{n-k} designs in terms of their complementary designs. *Ann. Statist.* **25**, 1176-1188.

Wu, C. F. J. and Chen, Y. (1992). A graph-aided method for planning two-level experiments when certain interactions are important. *Technometrics* **34**, 162-175.

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

Wu, H. Q. and Wu, C. F. J. (2002). Clear two-factor interaction and minimum aberration. *Ann. Statist.* **30**, 1496-1511.

Xu, H. and Cheng, C. -S. (2008). A complementary design theory for doubling. *Ann. Statist.* **36**, 445-457.

Yang, J. F., Li, P. F., Liu, M. Q. and Zhang, R. C. (2006). $2^{(n_1-n_2)-(k_1-k_2)}$ fractional factorial split-plot designs containing clear effects. *J. Statist. Plann. Inference* **136**, 4450-4458.

Zhang, R. C., Li, P., Zhao, S. L. and Ai, M. Y. (2008). A general minimum lower-order confounding criterion for two-level regular designs. *Statist. Sinica* **18**, 1689-1705.

Zhang, R. C. and Mukerjee, R. (2009a). Characterization of general minimum lower order confounding via complementary sets. *Statist. Sinica* **19**, 363-375.

Zhang, R. C. and Mukerjee, R. (2009b). General minimum lower order confounding in block designs using complementary sets. *Statist. Sinica* **19**, 1787-1802.

Zhang, R. C. and Shao, Q. (2001). Minimum aberration $(s^2)s^{n-k}$ designs. *Statist. Sinica* **11**, 213-223.

Zhao, S. L. and Zhang, R. C. (2008). 2^m4^n designs with resolution III or IV containing clear two-factor interaction components. *Statist. Papers* **49**, 441-454.

Department of Mathematical and Statistical Sciences, University of Alberta Edmonton, AB T6G 2G1, Canada.

E-mail: pengfei@stat.ualberta.ca

School of Mathematical Sciences, Qufu Normal University, Qufu 273165, China.

E-mail: zhaoshli758@126.com

KLAS and School of Mathematics and Statistics, Northeast Normal University, Changchun 130024, China.

LPMC and School of Mathematical Sciences, Nankai University, Tianjin 300071, China

E-mail: zhrch@nankai.edu.cn rczhang@nenu.edu.cn

(Received December 2008; accepted May 2010)