The Annals of Statistics
1996, Vol. 24, No. 6, 2536–2548

2^{n-l} DESIGNS WITH WEAK MINIMUM ABERRATION[1]

BY HEGANG CHEN AND A. S. HEDAYAT

Case Western Reserve University and University of Illinois

Since not all 2^{n-l} fractional factorial designs with maximum resolution are equally good, Fries and Hunter introduced the minimum aberration criterion for selecting good 2^{n-l} fractional factorial designs with the same resolution. We modify the concept of minimum aberration and define weak minimum aberration and show the usefulness of this new design concept. Using some techniques from finite geometry, we construct 2^{n-l} fractional factorial designs of resolution III with weak minimum aberration. Further, several families of 2^{n-l} fractional factorial designs of resolution III and IV with minimum aberration are obtained.

1. Introduction and definitions. Fractional factorial designs with factors at two levels are very important in factor screening experiments and many scientific investigations and the goal of this paper is to contribute to this area of experiment design. A 2^{-l}th fraction of a 2^n factorial design consisting of 2^{n-l} distinct combinations will be referred to as a 2^{n-l} fractional factorial design. An important characteristic of a fractional factorial design is its resolution. A design is of resolution r if no c-factor effect is confounded with any other effect containing less than $r - c$ factors. Often experimenters prefer to use a design with the highest possible resolution, but not all 2^{n-l} fractional factorial designs with maximum resolution are equally good. Fries and Hunter (1980) introduced the minimum aberration criterion for selecting good 2^{n-l} fractional factorial designs with the same resolution. In general, the minimum aberration criterion gives a good measure of the estimation capacity of a fractional factorial design. Chen and Wu (1991) and Chen (1992) constructed 2^{n-l} fractional factorial designs with minimum aberration for $l = 3$, 4 and 5. However, for large n and fixed $l(= 3, 4, 5)$, the number of runs N $(= 2^{n-l})$ in their designs is very large. Hedayat and Pesotan (1992, 1995) and Wu and Chen (1992) discussed and presented interesting results which are directly related to the concept of aberration in designs.

In the following, we introduce some notation and definitions concerning resolution and aberration in a design [see Fries and Hunter (1980) and Franklin (1984) for detailed discussions on these concepts]. We also define and discuss a new criterion called weak minimum aberration.

A 2^{n-l} fractional factorial design is a design that uses a 2^{-l} fraction of the whole 2^n runs from an experiment based on n factors each at two levels. Fur-

Received October 1994; revised March 1996.

[1]Research sponsored by NSF Grant DMS-93-04014 and the National Cancer Institute Grant PO1-CA48112.

AMS 1991 *subject classifications*. Primary 62K15; secondary 62K05.

Key words and phrases. Fractional factorial design, regular fraction, minimum aberration design, resolution, wordlength pattern.

ther, a fractional factorial design is said to be regular if the set of its treatment combinations forms a subgroup or a coset of a subgroup. Hereafter, all fractions will be regular fractions. For related additional information concerning fractional factorial designs, see Raktoe, Hedayat and Federer (1981).

The numbers $1, 2, \ldots, n$ attached to the factors are called letters and a product (juxtaposition) of any subset of these letters is called a word. The number of letters in a word is called the length of the word. Associated with every 2^{n-l} fractional factorial design is a set of l words $W_1, \ldots, W_l$ called generators. The set of distinct words formed by all possible products involving the l generators gives the defining relation of the fraction. Let $D(2^{n-l})$ be a 2^{n-l} fractional factorial design, and let $A_i(D)$ be the number of words of length i in the defining relation of $D(2^{n-l})$. Let $W(D)$ be the vector whose entries are $A_1(D), A_2(D), \ldots$:

$$W(D) = (A_1(D), A_2(D), \ldots, A_n(D)),$$

where $W(D)$ is referred to as the wordlength pattern of $D(2^{n-l})$. With this notation, the resolution of $D(2^{n-l})$ is the smallest i with positive $A_i(D)$ in $W(D)$.

DEFINITION 1. A 2^{n-l} fractional factorial design is said to have a maximum resolution R if no other 2^{n-l} fractional factorial design has larger resolution than R. Let D_1 and D_2 be two 2^{n-l} fractional factorial designs with wordlength patterns $W(D_1)$ and $W(D_2)$, and let s be the smallest integer such that $A_s(D_1) \neq A_s(D_2)$ in these two wordlength patterns. Then D_1 is said to have less aberration than D_2 if $A_s(D_1) < A_s(D_2)$. A 2^{n-l} fractional factorial design is said to have a minimum aberration if no other 2^{n-l} fractional factorial design has less aberration.

EXAMPLE 1. There are precisely five different $D(2^{9-5})$ fractional factorial designs [see Pu (1989)], namely,

$$\begin{aligned}
D_1&: I = 12345 = 126 = 237 = 348 = 1239,\\
D_2&: I = 12345 = 126 = 137 = 238 = 1239,\\
D_3&: I = 12345 = 126 = 137 = 148 = 1239,\\
D_4&: I = 12345 = 126 = 147 = 238 = 349,\\
D_5&: I = 12345 = 126 = 137 = 148 = 2349.
\end{aligned}$$

All these designs have maximum resolution III, but with different wordlength patterns

$$\begin{aligned}
W(D_1) &= (0, 0, 7, 9, 6, 6, 3, 0, 0),\\
W(D_2) &= (0, 0, 8, 10, 4, 4, 4, 1, 0),\\
W(D_3) &= (0, 0, 6, 10, 8, 4, 2, 1, 0),\\
W(D_4) &= (0, 0, 6, 9, 9, 6, 0, 0, 1),\\
W(D_5) &= (0, 0, 4, 14, 8, 0, 4, 1, 0).
\end{aligned}$$

Looking through these wordlength patterns, clearly D_5 has minimum aberration.

Both resolution and minimum aberration are defined under the assumptions: (a) lower-order interactions are more important than higher-order interactions, and (b) interactions of the same order are equally important. For 2^{n-l} designs of maximum resolution $R_{\max}$, the most important problem is to minimize the number of words of length $R_{\max}$. The numbers of words of length $R_{\max}+1$, $R_{\max}+2$ or higher are less important. Although combinatorial complexity of the defining relation makes the relation between lengths and estimability less certain, minimizing the number of shortest-length words generally leads to the estimability of more lower-order interactions, or less stringent assumptions. For example, if we assume that three-factor and higher-order interactions are negligible, designs of maximum resolution IV with the minimum number of words of length 4 should be good designs.

The concept of weak minimum aberration is a natural and useful modification of minimum aberration and is defined below.

DEFINITION 2. A 2^{n-l} fractional factorial design with maximum resolution $R_{\max}$ is said to have a weak minimum aberration if it has the minimum number of words of length $R_{\max}$.

We will use tools and terminology from finite geometry to define fractional factorial designs. In Section 2, we study the relationship of wordlength patterns between fractional factorial designs and their complementary designs in the whole factorial. In Section 3, we construct 2^{n-l} fractional factorial designs of resolution III with weak minimum aberration and several families of 2^{n-l} fractional factorial designs of resolutions III and IV with minimum aberration. Finally, in Section 4, we classify several families of $2^{n-(n-k)}$ designs with minimum aberration.

2. Some properties of wordlength patterns. Let $D(2^{n-l})$ be a 2^{n-l} fractional factorial design. The points in $D(2^{n-l})$ can be represented as column vectors as follows:

$$D(2^{n-l}) = \{\mathbf{x}\colon \mathbf{x} = B_n\mathbf{u},\ \mathbf{u} \in EG(n-l,2)\}, \tag{1}$$

where B_n is an $n \times (n-l)$ matrix of rank $n-l$ over the finite field $GF(2)$ and $EG(n-l,2)$ is the Euclidean geometry of dimension $n-l$ over $GF(2)$. The matrix B_n is called the factor representation of the fractional factorial design $D(2^{n-l})$. One such matrix B_n can be obtained by writing down the coordinates of n points of $PG(n-l-1,2)$ as rows, where $PG(n-l-1,2)$ is the projective geometry of dimension $n-l-1$ over $GF(2)$. Then a fractional factorial design as in (1) is determined by a set of n points of $PG(n-l-1,2)$. We note that $D(2^{n-l})$ is of resolution at least r if and only if no $(r-1)$ points of B_n are dependent.

LEMMA 1. *Let B_n be a factor representation of a fractional factorial design $D(2^{n-l})$. The resolution of $D(2^{n-l})$ is larger than or equal to 3 if and only if B_n is a set of n distinct points of $PG(n-l-1,2)$.*

For $2^{k-1} < n \le 2^k - 1$, the maximum resolution of a $2^{n-(n-k)}$ fractional factorial design is equal to III [Bose (1947)]. A $2^{n-(n-k)}$ fractional factorial design of resolution III is determined by a subset of n distinct points of $PG(k-1,2)$. Let $B_n = (\mathbf{a}_1, \dots, \mathbf{a}_n)^T$ be a subset of n distinct points of $PG(k-1,2)$. Such a subset can be obtained by deleting $2^k - 1 - n$ points from $PG(k-1,2)$. Without loss of generality, we can represent all points of $PG(k-1,2)$ as

$$\underbrace{\mathbf{a}_1, \dots, \mathbf{a}_n}_{B_n}, \underbrace{\mathbf{a}_{n+1}, \dots, \mathbf{a}_{2^k-1}}_{\overline{B}_n}, \tag{2}$$

where the first n points are all points of B_n and $\overline{B}_n$ denotes all points of $PG(k-1,2)\backslash B_n = \{\mathbf{a}_{n+1}, \dots, \mathbf{a}_{2^k-1}\}$. Let D_n and $\overline{D}_n$ be respectively the two fractional factorial designs corresponding to B_n and $\overline{B}_n$ as their factor representations. We call $\bar{D}_n$ the complementary design of D_n. Further, let $W(D_n)$ and $W(\overline{D}_n)$ be their corresponding wordlength patterns. Constructing a $2^{n-(n-k)}$ fractional factorial design of resolution III with weak minimum aberration is equivalent to deleting $2^k - 1 - n$ points from $PG(k-1,2)$ so that D_n has the minimum number of words of length 3. Here $A_3(D_n)$ in $W(D_n)$ and $A_3(\overline{D}_n)$ in $W(\overline{D}_n)$ are the numbers of one-dimensional subspaces of $PG(k-1,2)$ among points of B_n and $\overline{B}_n$ (one-dimensional subspaces of projective geometry are also called lines).

We shall now study the relationship between $A_3(D_n)$ and $A_3(\overline{D}_n)$. Let D be a fractional factorial design with (2) as its factor representation and $W(D)$ as its wordlength pattern. A word of length 3 in the defining relation of D corresponds to three points which form a line of $PG(k-1,2)$. Therefore, $A_3(D)$ in $W(D)$ is the number of lines of $PG(k-1,2)$, namely,

$$A_3(D) = \frac{(2^k-1)(2^k-2)}{(2^2-1)(2^2-2)}.$$

All lines of $PG(k-1,2)$ can be classified as one of the following three types:

1. Those lines containing one point from $\overline{B}_n$ and two points from B_n.
2. Those lines containing one point from B_n and two points from $\overline{B}_n$.
3. Those lines containing three points from $\overline{B}_n$ or from B_n.

Each pair of n points in B_n determines a line, but these $\binom{n}{2}$ lines are not all distinct. Indeed, $\binom{3}{2}A_3(D_n)$ pairs out of $A_3(D)$ lines of $PG(k-1,2)$ in B_n are duplicated. Therefore, the number of lines of type 1 is

$$\binom{n}{2} - \binom{3}{2}A_3(D_n). \tag{3}$$

Similarly, the number of lines of type 2 is

$$\binom{2^k-1-n}{2} - \binom{3}{2} A_3(\overline{D}_n), \tag{4}$$

and the number of lines of type 3 is

$$A_3(D_n) + A_3(\overline{D}_n). \tag{5}$$

Since the sum of (3), (4) and (5) is equal to the total number of lines in $PG(k-1, 2)$, that is,

$$\binom{n}{2} - \binom{3}{2} A_3(D_n) + \binom{2^k-1-n}{2} - \binom{3}{2} A_3(\overline{D}_n) + A_3(D_n) + A_3(\overline{D}_n)$$
$$= \frac{(2^k-1)(2^k-2)}{(2^2-1)(2^2-2)},$$

we can conclude the following relation between $A_3(D_n)$ and $A_3(\overline{D}_n)$:

$$A_3(D_n) = \frac{1}{2}\left[\binom{n}{2} + \binom{2^k-1-n}{2} - \frac{(2^k-1)(2^k-2)}{(2^2-1)(2^2-2)}\right] - A_3(\overline{D}_n). \tag{6}$$

From (6), we have the following lemma.

LEMMA 2. *Let the n rows of B_n be an n-subset of $PG(k-1, 2)$ and $\overline{B}_n = PG(k-1, 2)\backslash B_n$. Then B_n contains the minimum number of lines of $PG(k-1, 2)$ among all n-subsets of $PG(k-1, 2)$ if and only if $\overline{B}_n$ contains the maximum number of lines among all (2^k-1-n)-subsets of $PG(k-1, 2)$.*

For $n = 2^{k-1}$, Bose (1947) chose $\overline{B}_n = PG(k-2, 2)$ [embedded in $PG(k-1, 2)$]. The design D_n corresponding to B_n as its factor representation has resolution IV.

Let M be an m-subset of $PG(k-1, 2)$. The rank of M, denoted by rank(M), is the maximal number of independent points of M. Let $A_3(M)$ denote the number of lines in M. To search for an m-subset containing the maximum number of lines, the following lemmas will show that we only need to consider m-subsets with the minimum rank.

If M is a subset with rank $p+1$ ($\le k$), then M can be represented as

$$M = H \cup \{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_f\}, \tag{7}$$

where H is a subset of $PG(p-1, 2)$ [embedded in $PG(k-1, 2)$] with rank p, $\mathbf{a} \in PG(k-1, 2)\backslash PG(p-1, 2)$ and $\mathbf{b}_1, \dots, \mathbf{b}_f \in PG(p-1, 2)$.

LEMMA 3. *Let M be an m-subset of $PG(k-1, 2)$ with rank $p+1$ and having a form as in* (7). *Then there exists an m-subset M' of $PG(k-1, 2)$:*

$$M' = H' \cup \{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_t\}, \tag{8}$$

where H' is a subset of $PG(p-1, 2)$ with rank p, $\mathbf{a} \in PG(k-1, 2)\backslash PG(p-1, 2)$ and $\mathbf{b}_1, \dots, \mathbf{b}_t \in H'$, such that M' has at least as many lines as M.

PROOF. Let M have the representation as in (7), where $\mathbf{b}_1, \dots, \mathbf{b}_t \in H$ and $\mathbf{b}_{t+1}, \dots, \mathbf{b}_f \in PG(p-1,2)\backslash H$. Let $H_1^0 = \{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_t\}$, $H_2^0 = \{\mathbf{a}+\mathbf{b}_{t+1}, \dots, \mathbf{a}+\mathbf{b}_f\}$ and $H^0 = H_1^0 \cup H_2^0$. Consider

$$M' = H' \cup \{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_t\},$$

where $H' = H \cup \{\mathbf{b}_{t+1}, \dots, \mathbf{b}_f\}$ is a subset of $PG(p-1,2)$ with rank p. All lines in M can be classified into two classes. One class consists of all lines with points in H. Clearly, M' preserves all these lines. The other class of lines consists of lines with two points from H^0 and one point from H. Here M' preserves all lines in M containing two points from H_1^0 and one point from H. Lines in M containing two points from H_2^0 and one point from H correspond to lines in M' containing two points from $\{\mathbf{b}_{t+1}, \dots, \mathbf{b}_f\}$ and one point from H. Lines in M containing one point from $H_1^0\backslash\{\mathbf{a}\}$, one point from H_2^0 and one point from H correspond to lines in M' with one point from $\{\mathbf{b}_1, \dots, \mathbf{b}_t\}$, one point from $\{\mathbf{b}_{t+1}, \dots, \mathbf{b}_f\}$ and one point from H. Thus there are at least as many lines in M' as there are in M. Consequently, $A_3(M) \le A_3(M')$. □

From Lemma 3, it is sufficient to consider the representation given in (8) for counting the number of lines. It is easy to argue that, for $m = 2^r + q$, $0 \le q < 2^r$, the rank of any m-subset is at least $r+1$. However, if the rank exceeds $r+1$, then the following lemma gives additional information about M.

LEMMA 4. *Let M be an m-subset of $PG(k-1,2)$, $m = 2^r + q$ and $0 \le q < 2^r$, $r < k$. If the rank of M is larger than $r+1$, then there is an m-subset of $PG(k-1,2)$ with smaller rank whose number of lines is greater than the number of lines in M.*

PROOF. Let $\operatorname{rank}(M) = p+1 > r+1$. By Lemma 3, we may assume that

$$M = H \cup \{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_t\},$$

where H is a subset of $PG(p-1,2)$ with rank p, $\mathbf{a} \in PG(k-1,2)\backslash PG(p-1,2)$ and $\mathbf{b}_1, \dots, \mathbf{b}_t \in H$. Lines in M can be classified into two classes. One class contains all lines in H and another class contains all lines containing two points from $\{\mathbf{a}, \mathbf{a}+\mathbf{b}_1, \dots, \mathbf{a}+\mathbf{b}_t\}$ and one point from H. Since $\operatorname{rank}(M) > r+1$, $PG(p-1,2)\backslash H \ne \varnothing$, there are at least two points $\mathbf{s}_0^1$, $\mathbf{s}_0^2 \in H$ such that $\mathbf{c}_0 = \mathbf{s}_0^1 + \mathbf{s}_0^2 \in PG(p-1,2)\backslash H$. Consider the following process. Form the set

$$M_0 = H \cup \{\mathbf{c}_0, \mathbf{c}_0+\mathbf{b}_1, \dots, \mathbf{c}_0+\mathbf{b}_t\},$$

with $\mathbf{c}_0, \mathbf{c}_0+\mathbf{b}_1, \dots, \mathbf{c}_0+\mathbf{b}_{t_0} \notin H$ and $\mathbf{c}_0+\mathbf{b}_{t_0+1}, \dots, \mathbf{c}_0+\mathbf{b}_t \in H$ for some t_0 $(\le t)$. Therefore, M_0 can be represented as

$$M_0 = H \cup \{\mathbf{c}_0, \mathbf{c}_0+\mathbf{b}_1, \dots, \mathbf{c}_0+\mathbf{b}_{t_0}\}.$$

We observe that M_0 preserves all lines of the first class in M and its rank is p.

If $t = t_0$, we shall argue that M_0 has more lines than M. All lines of the second class in M correspond to lines in M_0 which contain two points from $\{\mathbf{c}_0, \mathbf{c}_0 + \mathbf{b}_1, \dots, \mathbf{c}_0 + \mathbf{b}_{t_0}\}$ and one point from H. Further, M_0 has at least one more line, namely, $\{\mathbf{s}_0^1, \mathbf{s}_0^2, \mathbf{c}_0\}$. Consequently, $A_3(M_0) > A_3(M)$.

If $t > t_0$, the number of points in M_0 is less than m. We shall now consider the lines in the second class in M and relate these lines to those in M_0. The lines in the second class in M can be conveniently classified into the following two types:

(a) Lines formed by $\{\mathbf{a}, \mathbf{a} + \mathbf{b}_i, \mathbf{b}_i\}$, $i = 1, \dots, t$.
(b) Lines formed by $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$, where $\mathbf{b}_i + \mathbf{b}_j \in H$.

Let us now go back to M_0. There are two possibilities for $\mathbf{c}_0$:

(i) $\mathbf{c}_0$ is not a sum of two points from $\{\mathbf{b}_1, \dots, \mathbf{b}_t\}$. All lines of type (a) correspond to lines in M_0 formed by $\{\mathbf{c}_0, \mathbf{c}_0 + \mathbf{b}_i, \mathbf{b}_i\}$. The lines of type (b) can be further classified into the following three cases:

(b1) Those lines $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$ with $1 \le i, j \le t_0$.
(b2) Those lines $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$ with $1 \le i \le t_0$, $t_0 + 1 \le j \le t$.
(b3) Those lines $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$ with $t_0 + 1 \le i, j \le t$.

The lines of type (b1) correspond to distinct new lines $\{\mathbf{c}_0 + \mathbf{b}_i, \mathbf{c}_0 + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$ in M_0. The lines of type (b3) have not been covered in M_0.

The lines of type (b2) correspond to the new lines $\{\mathbf{c}_0 + \mathbf{b}_i, \mathbf{c}_0 + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$ in M_0 which may not be distinct. There are no more than two distinct lines of type (b2) in M corresponding to the same line in M_0. Suppose that there are two lines of type (b2), $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$, $\{\mathbf{a} + \mathbf{b}_{i'}, \mathbf{a} + \mathbf{b}_{j'}, \mathbf{b}_{i'} + \mathbf{b}_{j'}\}$, $1 \le i, i' \le t_0$ and $t_0 + 1 \le j, j' \le t$, corresponding to the same line in M_0, that is, $\{\mathbf{c}_0 + \mathbf{b}_i, \mathbf{c}_0 + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\} = \{\mathbf{c}_0 + \mathbf{b}_{i'}, \mathbf{c}_0 + \mathbf{b}_{j'}, \mathbf{b}_{i'} + \mathbf{b}_{j'}\}$. Since $\mathbf{c}_0 + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j, \mathbf{c}_0 + \mathbf{b}_{j'}, \mathbf{b}_{i'} + \mathbf{b}_{j'} \in H$ and $\mathbf{c}_0 + \mathbf{b}_i, \mathbf{c}_0 + \mathbf{b}_{i'} \in M_0 \backslash H$, we have $\mathbf{b}_i = \mathbf{b}_{i'}$, $\mathbf{c}_0 + \mathbf{b}_j = \mathbf{b}_{i'} + \mathbf{b}_{j'}$ and $\mathbf{c}_0 + \mathbf{b}_i = \mathbf{b}_j + \mathbf{b}_{j'} \in M_0 \backslash H$. Therefore, we can assume that there is a new line of type (b3) $\{\mathbf{a} + \mathbf{b}_j, \mathbf{a} + \mathbf{b}_{j'}, \mathbf{b}_j + \mathbf{b}_{j'}\}$ which did not exist in M. The lost line $\{\mathbf{a} + \mathbf{b}_{i'}, \mathbf{a} + \mathbf{b}_{j'}, \mathbf{b}_{i'} + \mathbf{b}_{j'}\}$ will be compensated in the following process by corresponding to the new line of type (b3) $\{\mathbf{a} + \mathbf{b}_j, \mathbf{a} + \mathbf{b}_{j'}, \mathbf{b}_j + \mathbf{b}_{j'}\}$.

To cover the remaining lines in (b3) and the new lines of type (b3) resulting from the indistinctness of the corresponding lines in M_0 of the lines of type (b2), we add $(t - t_0)$ more point(s) to M_0 such that the new m-subset with rank p not only covers those lines which are not covered yet, but also has at least one more additional line than those in M. Since $\operatorname{rank}(M_0) = p \ge r + 1$, $PG(p - 1, 2) \backslash M_0 \ne \varnothing$, there are at least two points $\mathbf{s}_1^1$, $\mathbf{s}_1^2 \in M_0$ such that $\mathbf{s}_1^1 + \mathbf{s}_1^2 \in PG(p - 1, 2) \backslash M_0$. Let $\mathbf{c}_1 = \mathbf{s}_1^1 + \mathbf{s}_1^2 + \mathbf{b}_{t_0+1}$, and now build M_1 from M_0 by

$$M_1 = M_0 \cup \{\mathbf{c}_1 + \mathbf{b}_{t_0+1}, \dots, \mathbf{c}_1 + \mathbf{b}_t\}.$$

If all $\mathbf{c}_1 + \mathbf{b}_{t_0+1}, \dots, \mathbf{c}_1 + \mathbf{b}_t$ are in $PG(p - 1, 2) \backslash M_0$, those lines which are not covered correspond to those lines in M_1 formed by $\{\mathbf{c}_1 + \mathbf{b}_i, \mathbf{c}_1 + \mathbf{b}_j, \mathbf{b}_i + \mathbf{b}_j\}$, where $\mathbf{b}_i + \mathbf{b}_j \in H$ or $M_0 \backslash H$ and $t_0 < i, j \le t$. In this case, M_1 has at least one more additional line, namely, $\{\mathbf{s}_1^1, \mathbf{s}_1^2, \mathbf{c}_1 + \mathbf{b}_{t_0+1}\}$, than M. Otherwise, say, $\mathbf{c}_1 + \mathbf{b}_{t_0+1}, \dots, \mathbf{c}_1 + \mathbf{b}_{t_1} \in PG(p-1,2)\backslash M_0$ and $\mathbf{c}_1 + \mathbf{b}_{t_1+1}, \dots, \mathbf{c}_1 + \mathbf{b}_t$ are in M_0 or 0. In this case, we build M_1 as follows:

$$M_1 = M_0 \cup \{\mathbf{c}_1 + \mathbf{b}_{t_0+1}, \dots, \mathbf{c}_1 + \mathbf{b}_{t_1}\}.$$

If there is a $\mathbf{b}_i$ such that $\mathbf{c}_1 + \mathbf{b}_i = 0$, w.l.o.g. say $i = t_1 + 1$, then $\mathbf{b}_{t_1+1} = \mathbf{c}_1 = \mathbf{s}_1^1 + \mathbf{s}_1^2 + \mathbf{b}_{t_0+1}$. Since $\mathbf{b}_{t_1+1} + \mathbf{b}_i = \mathbf{c}_1 + \mathbf{b}_i \in M_1 \backslash M_0$ for $t_0 + 1 \le i \le t_1$, there are no remaining lines of the form $\{\mathbf{a} + \mathbf{b}_i, \mathbf{a} + \mathbf{b}_{t_1+1}, \mathbf{b}_{t_1+1} + \mathbf{b}_i\}$. Thus it does not affect the process. Clearly, $|M_1| > |M_0|$ (M_1 contains at least one more point, namely, $\mathbf{s}_1^1 + \mathbf{s}_1^2 = \mathbf{c}_1 + \mathbf{b}_{t_0+1}$). The lines uncovered in M_1 are similar to those at the first step, that is, the lines containing two points from $\{\mathbf{a}+\mathbf{b}_{t_1+1}, \dots, \mathbf{a}+\mathbf{b}_t\}$ and one point from M_0 or $M_1 \backslash M_0$ (which resulted from the indistinctness of the corresponding lines in M_1).

To cover all these lines, we repeat the same process on M_1. After v steps,

$$M_v = M_{v-1} \cup \{\mathbf{c}_v + \mathbf{b}_{t_{v-1}+1}, \dots, \mathbf{c}_v + \mathbf{b}_t\},$$

where $|M_v| = m$, $\operatorname{rank}(M_v) = p$ and, consequently, $A_3(M_v) > A_3(M)$.

(ii) $\mathbf{c}_0$ is a sum of two points from $\{\mathbf{b}_1, \dots, \mathbf{b}_t\}$, $\mathbf{c}_0$ has to be a sum of two points from $\{\mathbf{b}_{t_0+1}, \dots, \mathbf{b}_t\}$ and $t > t_0 + 1$. It is possible that there are several pairs $\{\mathbf{b}_{1_i}, \mathbf{b}_{2_i}\}$, $i = 1, \dots, h$, such that the sum of each pair is equal to $\mathbf{c}_0$ (all these pairs are distinct). Since $\mathbf{c}_0 = \mathbf{b}_{1_i} + \mathbf{b}_{2_i}$, two lines of type (a) $\{\mathbf{a}, \mathbf{a} + \mathbf{b}_{1_i}, \mathbf{b}_{1_i}\}$ and $\{\mathbf{a}, \mathbf{a} + \mathbf{b}_{2_i}, \mathbf{b}_{2_i}\}$ in M correspond to one line $\{\mathbf{c}_0, \mathbf{c}_0 + \mathbf{b}_{1_i}, \mathbf{b}_{1_i}\} = \{\mathbf{c}_0, \mathbf{c}_0 + \mathbf{b}_{2_i}, \mathbf{b}_{2_i}\}$ in M_0. Similarly, we can assume there is a new line of type (b3) $\{\mathbf{a} + \mathbf{b}_{1_i}, \mathbf{a} + \mathbf{b}_{2_i}, \mathbf{b}_{1_i} + \mathbf{b}_{2_i}\}$ which did not exist in M. The lost lines in the first step can be compensated in the following process by corresponding to the new lines of the form $\{\mathbf{a} + \mathbf{b}_{1_i}, \mathbf{a} + \mathbf{b}_{2_i}, \mathbf{b}_{1_i} + \mathbf{b}_{2_i}\}$, $i = 1, \dots, h$. By the same argument as in (i), the lemma is established. □

3. Main results. From Lemma 4, we can see that m-subsets containing the maximum number of lines must have the minimum rank. To search for the m-subsets containing the maximum number of lines, we only need to consider all m-subsets of $PG(r,2)$, $2^r \le m < 2^{r+1}$. By applying Lemma 2, the following theorem provides one structure of an m-subset containing the maximum number of lines.

THEOREM 1. *Let $m = 2^r + q$ and $0 \le q < 2^r$, $r < k$. Then the maximum number of lines in an m-subset of $PG(k-1,2)$ is*

$$\frac{(2^r-1)(2^r-2)}{(2^2-1)(2^2-2)} + \binom{q+1}{2}. \tag{9}$$

One structure of an m-subset of $PG(k-1,2)$ containing the maximum number of lines is

$$M = PG(r-1,2) \cup \{\mathbf{a}, \mathbf{a} + \mathbf{a}_{2^r-q}, \dots, \mathbf{a} + \mathbf{a}_{2^r-1}\}, \tag{10}$$

where $PG(r-1,2) = \{\mathbf{a}_1, \dots, \mathbf{a}_{2^r-1}\}$, and $\mathbf{a} \notin PG(r-1,2)$.

PROOF. By Lemma 4, we consider all m-subsets of $PG(r,2)$. The points of $PG(r,2)$ can be partitioned as follows:

$$PG(r-1,2), \mathbf{a}, \mathbf{a}+PG(r-1,2), \tag{11}$$

where $PG(r-1,2)=\{\mathbf{a}_1,\ldots,\mathbf{a}_{2^r-1}\}$, $\mathbf{a}+PG(r-1,2)=\{\mathbf{a}+\mathbf{a}_i|\mathbf{a}_i\in PG(r-1,2)\}$ and $\mathbf{a}\notin PG(r-1,2)$. An m-subset M of $PG(r,2)$ can be obtained by deleting 2^r-1-q points in (11). Let $\overline{M}=PG(r,2)\backslash M$, $\overline{M}=\{\mathbf{a}+\mathbf{a}_1,\ldots,\mathbf{a}+\mathbf{a}_{2^r-1-q}\}$. Note that $\overline{M}$ contains no lines, that is, $A_3(\overline{M})=0$. By Lemma 2, M in (10) contains the maximum number of lines among all m-subsets of $PG(k-1,2)$. There are $(2^r-1)(2^r-2)/(2^2-1)(2^2-2)$ lines in $PG(r-1,2)$; other lines in M are those containing two points from $\{\mathbf{a},\mathbf{a}+\mathbf{a}_{2^r-q},\ldots,\mathbf{a}+\mathbf{a}_{2^r-1}\}$ and one point from $PG(r-1,2)$. The number of lines of the latter type in M is $\binom{q+1}{2}$. The result (9) is the sum of these two numbers. $\square$

REMARK. The set (10) is a structure of an m-subset of $PG(k-1,2)$ containing the maximum number of lines. However, this structure is not unique. For convenience, a point of $PG(k-1,2)$ is denoted by $i_1i_2\ldots i_l$ if the i_1th, i_2th,$\ldots$,i_lth coordinates of this point are 1 and all others are 0. For example, when $k=4$, the following two 10-subsets both contain the maximum number of lines:

$$M_{10}=\{1,2,3,123,12,23,4,34,234,1234\}$$

and

$$M_{10}^*=\{1,2,3,13,12,23,123,4,14,1234\}.$$

Here M_{10}^* has a structure as in (10), and M_{10} does not have a structure as in (10). The following two $D(2^{21-16})$ designs, D_6 and D_7, are obtained by deleting M_{10} and M_{10}^* from $PG(3,2)$, respectively:

$$\begin{aligned} D_6\colon I &= 1256 = 1357 = 1458 = 2359 = 245t_{10} = 345t_{11} = 123t_{12} \\ &= 124t_{13} = 134t_{14} = 234t_{15} = 12345t_{16} = 45t_{17} \\ &= 35t_{18} = 25t_{19} = 15t_{20} = 1234t_{21}, \\ D_7\colon I &= 1256 = 1357 = 1458 = 2359 = 245t_{10} = 345t_{11} = 123t_{12} \\ &= 124t_{13} = 134t_{14} = 234t_{15} = 12345t_{16} = 24t_{17} \\ &= 34t_{18} = 1245t_{19} = 1345t_{20} = 2345t_{21}, \end{aligned}$$

where $t_{10},\ldots,t_{21}$ are factors $10,\ldots,21$. The wordlength patterns of D_6 and D_7 are

$$\begin{aligned} W(D_6) &= (0,0,40,220,641,1608,3640,6470,\ldots), \\ W(D_7) &= (0,0,40,221,640,1600,3648,6498,\ldots,). \end{aligned}$$

While both D_6 and D_7 have weak minimum aberration, D_6 has minimum aberration.

In general, the fractional factorial design with factor representation obtained by deleting $\overline{B}_n$ with a structure as in (10) from $PG(k-1,2)$ has weak minimum aberration. To search for minimum aberration designs, it is important to have a complete characterization of subsets with the maximum number of lines. This problem is currently under investigation.

THEOREM 2. *For $n = 2^k - 2^r - 1 - q$, $r < k$ and $0 \le q < 2^r$, and $m = 2^r + q$, the $2^{n-(n-k)}$ fractional factorial design D_n whose factor representation is obtained by deleting all points of an m-subset with structure as in* (10) *from $PG(k-1,2)$ has weak minimum aberration. The minimum number of words of length* 3, *$A_3(D_n)$, is equal to*

$$\tag{12} \tfrac{1}{3}[(2^{k-1}-1)(2^k - 3\cdot 2^r - 1) + 3(2^r - 2^{k-1})q + (4^r - 1)].$$

PROOF. Let us reconsider B_n and $\overline{B}_n$ in Section 2, $m = 2^k - 1 - n = 2^r + q$ and $n = 2^k - 2^r - 1 - q$. Assume that $\overline{B}_n$ is an m-subset with structure (10) containing the maximum number of lines, that is, $A_3(\overline{D}_n)$ is equal to (9). From (6),

$$A_3(D_n) = \frac{1}{2}\left[\binom{2^k - 2^r - q - 1}{2} + \binom{2^r + q}{2} - \frac{(2^k-1)(2^k-2)}{(2^2-1)(2^2-2)}\right] - \frac{(2^r-1)(2^r-2)}{(2^2-1)(2^2-2)} - \binom{q+1}{2}.$$

Upon simplification, we obtain (12). □

THEOREM 3. *For $n = 2^k - 2^r$, $r < k$, the $2^{n-(n-k)}$ fractional factorial design whose factor representation is obtained by deleting all points of $PG(r-1,2)$ from $PG(k-1,2)$ has minimum aberration. Specifically, for $n = 2^{k-1}$, the design is of resolution IV.*

PROOF. Since $\overline{B}_n = PG(r-1,2)$ is the only (2^r-1)-subset of rank r, the result follows by Theorem 1. □

4. Classification of $2^{n-(n-k)}$ fractional factorial designs with minimum aberration.

Let B_n be an n-subset of $PG(k-1,2)$ and $\overline{B}_n = PG(k-1,2)\backslash B_n$, and let D_n and $\overline{D}_n$ be their corresponding fractional factorial designs. If $\overline{B}_n$ with the maximum number of lines is unique, the design D_n has minimum aberration.

As indicated in the remark in Section 3, a subset with a maximum number of lines is not unique. To further identify designs with less aberration, Chen

(1993) studied the relationship between $A_4(D_n)$ in $W(D_n)$ and $A_4(\overline{D}_n)$ in $W(\overline{D}_n)$ and obtained the following result:

$$\begin{aligned} A_4(D_n) = \frac{1}{3}\Bigg[& 2A_3(\overline{D}_n) - A_3(D_n) + 2^{k-2}(2^k - n - 1)(2^k - n - 2) \\ & + \binom{n}{3} - 2\binom{2^k - n}{3} - \frac{(2^{k-2} - 1)(2^k - 1)(2^k - 2)}{(2^2 - 1)(2^2 - 2)}\Bigg] + A_4(\overline{D}_n). \end{aligned} \tag{13}$$

From (13), we see that $A_4(D_n)$ is minimized if and only if $A_4(\overline{D}_n)$ is minimized. If $\overline{B}_n$ with the maximum number of lines is not unique, we should choose $\overline{B}_n$ such that $A_4(\overline{D}_n)$ is minimized among $(2^k - 1 - n)$-subsets of $PG(k-1, 2)$ with the maximum number of lines. If the $(2^k - 1 - n)$-subset $\overline{B}_n$ with maximum number of $A_3(\overline{D}_n)$ and minimum number of $A_4(\overline{D}_n)$ is unique, then the design D_n with B_n as its factor representation has minimum aberration.

It is easy to see that the factor representation obtained by deleting one or two point(s) from $PG(k - 1, 2)$ is unique up to equivalence; hence the corresponding design has minimum aberration. Pu (1989) was first to classify all m-subsets of minimum rank for $m \leq 15$. The following m-subsets M_m, $m = 3, \dots, 9$, with maximum $A_3(M_m)$ are unique up to equivalence:

$m = 3$

$$M_3 = \{1, 2, 12\},$$

$m = 4$

$$M_4 = \{1, 2, 3, 23\},$$

$m = 5$

$$M_5 = \{1, 2, 3, 12, 13\},$$

$m = 6$

$$M_6 = \{1, 2, 3, 12, 13, 23\},$$

$m = 7$

$$M_7 = PG(2, 2),$$

$m = 8$

$$M_8 = \{1, 2, 3, 4, 13, 23, 12, 123\},$$

$m = 9$

$$M_9 = \{1, 2, 3, 4, 1234, 13, 23, 12, 123\}.$$

Deleting these subsets from any $PG(k - 1, 2)$ yields subsets corresponding to designs with minimum aberration.

The following M_m (without "*") with maximum $A_3(M_m)$ and minimum $A_4(M_m)$ are unique up to equivalence:

$m = 10$

$$M_{10} = \{1, 2, 3, 4, 1234, 12, 23, 34, 123, 234\},$$
$$M_{10}^* = \{1, 2, 3, 4, 1234, 12, 23, 13, 14, 123\},$$

$m = 11$

$$M_{11} = \{1, 2, 3, 4, 1234, 12, 13, 14, 23, 24, 34\},$$
$$M_{11}^* = \{1, 2, 3, 4, 1234, 14, 23, 24, 34, 234, 123\},$$

$m = 12$

$$M_{12} = \{1, 2, 3, 4, 1234, 12, 13, 14, 23, 124, 234, 134\},$$

$m = 13$

$$M_{13} = \{1, 2, 3, 4, 1234, 12, 13, 14, 23, 24, 34, 123, 124\},$$

$m = 14$

$$M_{14} = \{1, 2, 3, 4, 1234, 12, 13, 14, 23, 24, 34, 123, 124, 234\},$$

$m = 15$

$$M_{15} = PG(3, 2).$$

Each of the M_m subsets with "*" has the maximum number of $A_3(M_m)$, but its number of $A_4(M_m)$ is not minimized. Deleting these subsets from any $PG(k-1, 2)$ yields subsets corresponding to designs with weak minimum aberration. The designs derived by M_m without "*" have minimum aberration.

Acknowledgments. We would like to thank the referee and an Associate Editor for their helpful comments. Specifically, the revision benefited greatly from the insights and constructive suggestions of an Associate Editor.

REFERENCES

BOSE, R. C. (1947). Mathematical theory of the symmetrical factorial design. *Sankhyā Ser. A* **8** 107–166.

CHEN, H. (1993). Contributions to experimental designs. Ph.D. dissertation, Univ. Illinois, Chicago.

CHEN, J. (1992). Some results on 2^{n-k} fractional factorial designs and search for minimum aberration designs. *Ann. Statist.* **20** 2124–2141.

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.

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.

HEDAYAT, A. S. and PESOTAN, H. (1992). Two-level factorial designs for main effects. *Statist. Sinica* **2** 453–464.

HEDAYAT, A. S. and PESOTAN, H. (1997). Designs for two-level factorial experiments with linear models containing main effects and selected two-factor interactions. *J. Statist. Plann. Inference.* To appear.

PU, K. (1989). Contributions to fractional factorial designs. Ph.D. dissertation, Univ. Illinois, Chicago.

RAKTOE, B. L., HEDAYAT, A. S. and FEDERER, W. T. (1981). *Factorial Designs*. Wiley, New York.

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.

DEPARTMENT OF EPIDEMIOLOGY
AND BIOSTATISTICS
SCHOOL OF MEDICINE
CASE WESTERN RESERVE UNIVERSITY
CLEVELAND, OHIO 44106
E-MAIL: chen@hal.epbi.cwru.edu

DEPARTMENT OF MATHEMATICS, STATISTICS
AND COMPUTER SCIENCE (M/C 249)
322 SCIENCE AND ENGINEERING OFFICES
851 SOUTH MORGAN STREET
UNIVERSITY OF ILLINOIS
CHICAGO, ILLINOIS 60607-7045
E-MAIL: hedayat@uic.edu