

A BAYESIAN APPROACH TO THE SELECTION OF TWO-LEVEL MULTI-STRATUM FACTORIAL DESIGNS

BY MING-CHUNG CHANG* AND CHING-SHUI CHENG*†

*Academia Sinica** and *University of California, Berkeley*†

In a multi-stratum factorial experiment, there are multiple error terms (strata) with different variances that arise from complicated structures of the experimental units. For unstructured experimental units, minimum aberration is a popular criterion for choosing regular fractional factorial designs. One difficulty in extending this criterion to multi-stratum factorial designs is that the formulation of a wordlength pattern based on which minimum aberration is defined requires an order of desirability among the relevant words, but a natural order is often lacking. Furthermore, a criterion based only on wordlength patterns does not account for the different stratum variances. Mitchell, Morris, and Ylvisaker (1995) proposed a framework for Bayesian factorial designs. A Gaussian process is used as the prior for the treatment effects, from which a prior distribution of the factorial effects is induced. This approach is applied to study optimal and efficient multi-stratum factorial designs. Good surrogates for the Bayesian criteria that can be related to wordlength and generalized wordlength patterns for regular and nonregular designs, respectively, are derived. A tool is developed for eliminating inferior designs and reducing the designs that need to be considered without requiring any knowledge of stratum variances. Numerical examples are used to illustrate the theory in several settings.

1. Introduction. In a multi-stratum factorial experiment, there are multiple error terms (strata) with different variances. For example, in an experiment conducted in several days, suppose the levels of some treatment factors are difficult to change and must be kept the same throughout the day, while the levels of the other factors can be changed from run to run on the same day. Then the precision of the estimates of the main effects of hard-to-change factors depends on the between-day variability, and that of the main effects of easy-to-change factors depends on the between-run variability on the same day. Typically the former is greater than the latter. Such an experiment is said to have two *strata*. The two strata arise from the structure of the experimental units that some larger units (those associated with the days, called *whole-plots*) are split into smaller units (those associated with the runs, called *subplots*). If the experiment is to be blocked, with the whole-plots grouped into more homogeneous blocks, then we will have a third stratum

MSC 2010 subject classifications: 62K05, 62K15

Keywords and phrases: A-criterion, block structure, D-criterion, Gaussian process, (M.S)-criterion, minimum aberration, wordlength pattern

and the associated between-block variability is expected to be greater than the between-whole-plot variability within the same block. In general, the strata are determined by the structure of experimental units, called a *block structure*.

When the experimental units are unstructured, the minimum aberration (MA) criterion proposed by Fries and Hunter [11] is a popular criterion for choosing *regular* fractional factorial designs under the hierarchical assumption that lower-order effects are more important than higher-order effects and effects of the same order are equally important. This criterion is based on the so-called *wordlength pattern* and has been extended to nonregular designs (Tang and Deng [20]). It has also been extended to several types of multi-stratum fractional factorial designs, such as block designs, split-plot designs, and blocked split-plot designs, by modifying the wordlength pattern. The modifications are often ad hoc and without strong justifications. The formulation of a wordlength pattern requires an order of desirability among the relevant words, but a natural order is often lacking. Furthermore, a criterion based only on wordlength patterns does not account for the different stratum variances. Cheng, Steinberg, and Sun [5] showed that minimum aberration is a good surrogate for maximum *estimation capacity*, a model-robust criterion that maximizes the number of estimable models among some potential models. For nonregular designs, since different estimable models may be estimated with different efficiencies, it is no longer enough to compare the number of estimable models. In this case, the *information capacity* criterion (Sun [18]) of maximizing the average efficiency over a set of potential models is more appropriate. One can apply the information capacity criterion to compare multi-stratum factorial designs, under which different estimable models may also be estimated with different efficiencies due to different stratum variances. Cheng and Tsai [6, 7] adopted this approach and derived a surrogate for the maximum information capacity criterion. This line of work, however, assumes that the three-factor and higher-order interactions are negligible, and can only be applied to orthogonal regular designs.

Mitchell, Morris, and Ylvisaker [16] proposed a framework for Bayesian fractional factorial designs. A Gaussian process commonly used in the literature of computer experiments (Sacks, Welch, Mitchell, and Wynn [17]) to model unknown deterministic response functions is used as the prior for the treatment effects, from which a prior distribution for the factorial effects is induced. This approach, which provides more flexibility in incorporating the prior knowledge, was further developed by Kerr [15] and Joseph [12] for studying optimal fractional factorial designs. Joseph, Ai, and Wu [13] used the same Bayesian approach to inspire a minimum aberration criterion for mixed two- and four-level designs with unstructured units. The main objective of the present article is to apply this approach to multi-stratum factorial designs. Another work relevant to ours is Ai, Kang, and Joseph [1] in which a Bayesian approach was applied to study blocked fractional factorial designs with fixed block effects.

Some preliminary materials, including block structures, treatment factorial effects, strata, orthogonal designs, Bayesian approach, and statistical models, are presented in Section 2. Explicit forms of Bayesian A- and D-optimality criteria are derived in Section 3. Generally

analytical results on Bayesian optimal fractional factorial designs are difficult to prove. In Section 4 we provide one such result in the setting of regular half-fractions with two blocks; earlier, Bayesian optimal half-fractions without blocking was obtained in [15]. In Section 5, we derive good surrogates for the Bayesian criteria that can also be applied to non-regular and nonorthogonal designs. For orthogonal blocking of complete factorial designs, the surrogate criterion reduces to the usual minimum aberration criterion. In the case of fractional factorial designs for unstructured experimental units, the surrogate criterion is shown to be a refinement of minimum and generalized minimum aberration for regular and nonregular designs, respectively. For nonregular multi-stratum designs, our approach provides a stronger justification than naive modifications of the usual wordlength patterns from regular to nonregular designs. A tool is developed for eliminating inferior designs without any knowledge of the stratum variances, thereby reducing the designs that need to be considered. Examples in several settings are provided in Section 6 to illustrate the theory.

2. Preliminaries.

2.1. *Unit factors and block structures.* Let Ω be a set of N experimental units. An $n_{\mathcal{F}}$ -level *unit factor* \mathcal{F} can be considered as a partition of Ω into $n_{\mathcal{F}}$ disjoint nonempty subsets. Each subset, called an \mathcal{F} -class, consists of units that have the same level of \mathcal{F} . Given two factors \mathcal{F}_1 and \mathcal{F}_2 , we say that \mathcal{F}_1 is *nested* in \mathcal{F}_2 , (\mathcal{F}_1 is finer than \mathcal{F}_2 , or \mathcal{F}_2 is coarser than \mathcal{F}_1), denoted by $\mathcal{F}_1 \prec \mathcal{F}_2$, if any two units in the same \mathcal{F}_1 -class are also in the same \mathcal{F}_2 -class and $\mathcal{F}_1 \neq \mathcal{F}_2$. We write $\mathcal{F}_1 \preceq \mathcal{F}_2$ if $\mathcal{F}_1 \prec \mathcal{F}_2$ or $\mathcal{F}_1 = \mathcal{F}_2$.

The coarsest factor, denoted by \mathcal{U} and called the universal factor, is the single-level factor with all the units in the same class. On the other hand, the finest factor, denoted by \mathcal{E} and called the equality factor, is the N -level factor with each class consisting of one single unit. A block structure \mathfrak{B} is a collection of unit factors on the same Ω and we include \mathcal{U} and \mathcal{E} in each block structure. For example, the block structure of a usual block design with $N = bk$ units in b blocks of size k can be regarded as a collection $\{\mathcal{U}, \mathcal{B}, \mathcal{E}\}$ of three unit factors, where \mathcal{E} is a bk -level factor corresponding to bk units, and \mathcal{B} is a b -level block factor that partitions the bk units into b blocks each of size k . We have $\mathcal{E} \preceq \mathcal{B} \preceq \mathcal{U}$. This is also the block structure of a split-plot experiment with b whole-plots each consisting of k subplots. An experiment with unstructured units is considered to have the block structure $\{\mathcal{U}, \mathcal{E}\}$.

A factor is said to be *uniform* if all its classes are of the same size. Given two unit factors \mathcal{F}_1 and \mathcal{F}_2 , the *supremum* of \mathcal{F}_1 and \mathcal{F}_2 , denoted by $\mathcal{F}_1 \vee \mathcal{F}_2$, is the factor such that (i) $\mathcal{F}_1, \mathcal{F}_2 \preceq \mathcal{F}_1 \vee \mathcal{F}_2$, and (ii) $\mathcal{F}_1 \vee \mathcal{F}_2 \preceq \mathcal{G}$ for all \mathcal{G} such that $\mathcal{F}_1, \mathcal{F}_2 \preceq \mathcal{G}$. We say that \mathcal{F}_1 and \mathcal{F}_2 are *orthogonal* if \mathcal{F}_1 and \mathcal{F}_2 have proportional frequencies in each $(\mathcal{F}_1 \vee \mathcal{F}_2)$ -class, i.e., for each $(\mathcal{F}_1 \vee \mathcal{F}_2)$ -class Γ , if both the i th \mathcal{F}_1 -class and the j th \mathcal{F}_2 -class are contained in Γ , then

$$n_{ij} = \frac{n_i n_j}{|\Gamma|},$$

where n_{i+} , n_{+j} , and n_{ij} are the numbers of units in, respectively, the i th \mathcal{F}_1 -class, the j th \mathcal{F}_2 -class, and the intersection of these two classes, and $|I|$ is the number of units in I . Throughout this paper, we only consider block structures \mathfrak{B} that satisfy the following conditions:

$$(2.1) \quad \text{All the factors in } \mathfrak{B} \text{ are uniform and pairwise orthogonal.}$$

$$(2.2) \quad \mathcal{E} \in \mathfrak{B}, \mathcal{U} \in \mathfrak{B}.$$

$$(2.3) \quad \mathcal{F}, \mathcal{G} \in \mathfrak{B} \Rightarrow \mathcal{F} \vee \mathcal{G} \in \mathfrak{B}.$$

We note that most of the block structures encountered in practice satisfy (2.1), (2.2), and (2.3).

The relation between the units and the levels of \mathcal{F} can be described by an $N \times n_{\mathcal{F}}$ incidence matrix $\mathbf{X}_{\mathcal{F}}$ with 0 and 1 entries such that the (i, j) th entry of $\mathbf{X}_{\mathcal{F}}$ is 1 if and only if the i th unit is in the j th \mathcal{F} -class. In particular,

$$\mathbf{X}_{\mathcal{U}} = \mathbf{1}_N,$$

where $\mathbf{1}_N$ is the $N \times 1$ vector of 1's.

2.2. Treatment factorial effects. Suppose there are n two-level treatment factors. Denote each treatment combination by $\mathbf{x} = (x_1, \dots, x_n)^T$, where $x_i = 0$ or 1 is the level of the i th treatment factor. Let $\boldsymbol{\alpha}$ be a $2^n \times 1$ vector with components $\alpha(\mathbf{x})$, where $\alpha(\mathbf{x})$ is the effect of treatment combination \mathbf{x} . We can express $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha} = \mathbf{P}\boldsymbol{\beta}$, where \mathbf{P} is a $2^n \times 2^n$ model matrix for a 2^n complete factorial experiment and $\boldsymbol{\beta} = (\beta_{S_1}, \dots, \beta_{S_{2^n}})^T$ is the vector of treatment factorial effects β_{S_l} , one for each $S_l \subseteq \{1, \dots, n\}$. Specifically, $\beta_{\emptyset} = \frac{1}{2^n} \sum_{\mathbf{x}} \alpha(\mathbf{x})$ is the mean and the corresponding column of \mathbf{P} consists of 1's. For $S = \{i\}$, $1 \leq i \leq n$, β_S is a main effect contrast of factor i , and in the associated column of \mathbf{P} , the entry corresponding to \mathbf{x} is 1 if $x_i = 1$, and is -1 if $x_i = 0$. Furthermore, for $S = \{i_1, \dots, i_k\}$, β_S is a k -factor interaction contrast and the corresponding column of \mathbf{P} is the Hadamard product of the columns corresponding to the main effects of factors i_1, \dots, i_k , where the Hadamard product of $\mathbf{u} = (u_1, \dots, u_h)^T$ and $\mathbf{v} = (v_1, \dots, v_h)^T$ is $\mathbf{u} \odot \mathbf{v} = (u_1 v_1, \dots, u_h v_h)^T$. For convenience, β_S is often written as $\mathbf{i}_1 \dots \mathbf{i}_k$ (or a combination of letters if each factor is labeled by a letter), and is called a *word* of length k . Often we do not distinguish among S , β_S , and the corresponding words.

Under the Bayesian framework, $\boldsymbol{\beta}$ is treated as a random vector. Specifying the distribution of $\boldsymbol{\beta}$ is a crucial step. We take the approach proposed in [16]; see also the discussions in Section 10.11 of Cheng [4]. Here we regard $\boldsymbol{\alpha}$, a function of the factor settings $\mathbf{x} = (x_1, \dots, x_n)$, as a realization of a Gaussian process. Specifically, suppose the distribution of $\boldsymbol{\alpha}$ is multivariate normal with zero mean and covariance matrix $\sigma^2 \mathbf{R}$. It can be seen that the columns of \mathbf{P} are mutually orthogonal; thus $\boldsymbol{\beta} = \frac{1}{2^n} \mathbf{P}^T \boldsymbol{\alpha}$ and hence

$$(2.4) \quad \boldsymbol{\beta} \sim \text{N}(\mathbf{0}, 2^{-2n} \sigma^2 \mathbf{P}^T \mathbf{R} \mathbf{P}).$$

It was shown in [16] that if $\boldsymbol{\alpha}$ is stationary, i.e., the correlation between $\alpha(\mathbf{x}_1)$ and $\alpha(\mathbf{x}_2)$ depends only on the factors where \mathbf{x}_1 and \mathbf{x}_2 differ, then the β_S 's, $S \subseteq \{1, \dots, n\}$, are independent. In this case $\text{cov}(\boldsymbol{\beta})$ is diagonal and

$$(2.5) \quad (2^{-n/2} \mathbf{P}^T) \text{cov}(\boldsymbol{\alpha}) (2^{-n/2} \mathbf{P}) = 2^n \text{cov}(\boldsymbol{\beta})$$

is the spectral decomposition of $\text{cov}(\boldsymbol{\alpha})$. It follows that the columns of \mathbf{P} , which define the β_S 's, $S \subseteq \{1, \dots, n\}$, are eigenvectors, and the variances of the β_S 's are the eigenvalues of $\text{cov}(\boldsymbol{\alpha})$ divided by 2^n .

As an example, assume that the Gaussian process has the covariance function

$$\text{cov}(\alpha(\mathbf{x}_1), \alpha(\mathbf{x}_2)) = \sigma^2 \prod_{1 \leq j \leq n: x_{1j} \neq x_{2j}} \rho_j,$$

where $0 < \rho_1, \dots, \rho_n < 1$. Then we have

$$\text{var}(\beta_S) = \frac{\sigma^2}{2^n} \left\{ \prod_{1 \leq i \leq n, i \in S} (1 - \rho_i) \right\} \left\{ \prod_{1 \leq i \leq n, i \notin S} (1 + \rho_i) \right\}, \quad \text{cov}(\beta_S, \beta_{S'}) = 0 \text{ if } S \neq S'.$$

This implies that if $S \subset S'$, then $\text{var}(\beta_S) > \text{var}(\beta_{S'})$. This property, consistent with the hierarchical assumption, was referred to as the property of nested decreasing interaction variances in [15]. In addition, if $\rho_1 = \dots = \rho_n = \rho$, then we have $\text{var}(\beta_S) = \frac{\sigma^2}{2^n} (1 - \rho)^{|S|} (1 + \rho)^{n-|S|}$, where $|S|$ is the number of elements in S . An alternative parametrization is to let $r = (1 - \rho)/(1 + \rho)$ and $\tau^2 = \sigma^2(1 + r)^{-n}$, which leads to

$$(2.6) \quad \text{var}(\beta_S) = \tau^2 r^{|S|}.$$

In this case of *isotropic* priors, factorial effects of the same order can be regarded as equally important; thus MA designs are expected to perform well. In the rest of the paper, we assume that $\boldsymbol{\alpha}$ is stationary.

2.3. Design construction and defining words. The construction of a multi-stratum fractional factorial design involves the selection of a subset of treatment combinations and assigning them to various classes of the unit factors in the block structure. For *regular* designs, this is done by solving linear equations. With the levels represented by the two elements 0 and 1 of the field \mathbb{Z}_2 under modulo 2 addition and multiplication, the treatment combinations can be identified with the 2^n vectors in the n -dimensional space \mathbb{Z}_2^n .

For each nonzero vector \mathbf{a} in \mathbb{Z}_2^n , the two equations $\mathbf{a}^T \mathbf{x} = 0$ and $\mathbf{a}^T \mathbf{x} = 1$ divide the treatment combinations \mathbf{x} into two disjoint sets, say H_0 and H_1 , respectively. Let $S = \{1 \leq i \leq n : a_i \neq 0\}$, then the factorial effect β_S defined in Section 2.2 is a difference of $\sum_{\mathbf{x}: \mathbf{x} \in H_0} \alpha(\mathbf{x})$ and $\sum_{\mathbf{x}: \mathbf{x} \in H_1} \alpha(\mathbf{x})$ divided by 2^n .

A system of p linear equations $\mathbf{a}_i^T \mathbf{x} = b_i, i = 1, \dots, p$, where $\mathbf{a}_1, \dots, \mathbf{a}_p$ are linearly independent, has 2^{n-p} solutions, which constitute a regular 2^{n-p} fractional factorial design.

The $2^p - 1$ nonzero linear combinations of $\mathbf{a}_1, \dots, \mathbf{a}_p$ (and the corresponding β_S) are called *treatment defining words (effects)*. Let A_k be the number of treatment defining words of length k . Then (A_1, \dots, A_n) is called the *wordlength pattern* of the design. The 2^p choices of b_1, \dots, b_p divide the 2^n treatment combinations into 2^p disjoint sets, each of which can be used as a 2^{n-p} fractional factorial design. Similarly, h linearly independent vectors $\mathbf{b}_1, \dots, \mathbf{b}_h$ such that $\mathbf{a}_1, \dots, \mathbf{a}_p, \mathbf{b}_1, \dots, \mathbf{b}_h$ are linearly independent can be used to divide the treatment combinations in the fractional factorial design into 2^h blocks of equal size. The vectors $\mathbf{a}_1, \dots, \mathbf{a}_p$ are called independent treatment defining words, and $\mathbf{b}_1, \dots, \mathbf{b}_h$ are called independent block defining words. The non-zero linear combinations of $\mathbf{a}_1, \dots, \mathbf{a}_p, \mathbf{b}_1, \dots, \mathbf{b}_h$ that are not treatment defining words are called *block defining words (effects)*.

2.4. *Statistical model.* Let $\mathbf{y} = (y_1, \dots, y_N)^T$ be the responses under a fractional factorial design, and the N experimental units have a block structure $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1, \dots, \mathcal{F}_m\}$. Throughout this paper, let $\mathcal{F}_0 = \mathcal{U}$ and $\mathcal{F}_m = \mathcal{E}$. Suppose

$$(2.7) \quad \mathbf{y} = \mathbf{X}_{\mathcal{T}}\boldsymbol{\alpha} + \sum_{i=0}^m \mathbf{X}_{\mathcal{F}_i}\boldsymbol{\gamma}^{\mathcal{F}_i},$$

where $\mathbf{X}_{\mathcal{T}}$ is an $N \times 2^n$ unit-treatment incidence matrix and $\boldsymbol{\gamma}^{\mathcal{F}_i} = (\gamma_1^{\mathcal{F}_i}, \dots, \gamma_{n_{\mathcal{F}_i}}^{\mathcal{F}_i})^T$, where $\gamma_j^{\mathcal{F}_i}$ is the effect of the j th level of unit factor \mathcal{F}_i (for example, block effects, whole-plot effects, and subplot effects). We assume that the γ 's are independent, with each $\gamma_j^{\mathcal{F}_i}$ following an $N(0, \sigma_{\mathcal{F}_i}^2)$ distribution, and that they are independent of $\boldsymbol{\alpha}$. Then

$$(2.8) \quad \mathbf{y}|\boldsymbol{\beta} \sim N(\mathbf{X}_{\mathcal{T}}\mathbf{P}\boldsymbol{\beta}, \sum_{i=0}^m \sigma_{\mathcal{F}_i}^2 \mathbf{X}_{\mathcal{F}_i}\mathbf{X}_{\mathcal{F}_i}^T).$$

Let $\mathbf{U} = \mathbf{X}_{\mathcal{T}}\mathbf{P}$ and $\mathbf{V} = \sum_{i=0}^m \sigma_{\mathcal{F}_i}^2 \mathbf{X}_{\mathcal{F}_i}\mathbf{X}_{\mathcal{F}_i}^T$. Then \mathbf{U} is the full model matrix under the design. Each column of \mathbf{U} , except for the column of 1's, corresponds to a factorial effect. Throughout this paper, we only consider single-replicate complete factorial designs or fractional factorial designs in which no treatment combination is observed more than once. Then \mathbf{U} consists of N rows of \mathbf{P} . Since $\mathbf{P}\mathbf{P}^T = 2^n \mathbf{I}_{2^n}$, we have

$$\mathbf{U}\mathbf{U}^T = 2^n \mathbf{I}_N.$$

If the design is a regular 2^{n-p} fractional factorial design, then the 2^n columns of \mathbf{U} can be partitioned into 2^{n-p} sets of size 2^p , with the columns in the same set corresponding to aliased effects. Any two columns of \mathbf{U} in the same alias set are either identical or can be obtained from each other by changing the signs of all the entries. Without loss of generality, we assume that all the aliased columns of \mathbf{U} are identical. For each nonempty $S \subseteq \{1, \dots, n\}$, let $\mathcal{A}(\beta_S)$ be the set of β_S and all its aliases, and let $\mathcal{A}(\beta_{\emptyset})$ be the defining

contrast subgroup, i.e., the set consisting of the mean and the treatment defining words. For convenience, we also write the $\mathcal{A}(\beta_S)$'s as $\mathcal{A}_0, \dots, \mathcal{A}_{2^{n-p}-1}$, with $\mathcal{A}_0 = \mathcal{A}(\beta_\emptyset)$.

Under a block structure satisfying (2.1), (2.2), and (2.3), the covariance matrix \mathbf{V} has $m + 1$ eigenspaces $W_{\mathcal{F}_0}, \dots, W_{\mathcal{F}_m}$, with one eigenspace associated with each of the $m + 1$ unit factors, where $W_{\mathcal{F}_0} = W_{\mathcal{U}}$ is the one-dimensional space consisting of all the vectors with constant entries, and each other eigenvector defines a unit contrast; furthermore, the eigenspaces are determined by the block structure and do not depend on the entries of \mathbf{V} . The readers are referred to Bailey [2] and [4] for these results as well as how to determine the $W_{\mathcal{F}_i}$'s from \mathfrak{B} . Let the corresponding eigenvalues be $\xi_{\mathcal{F}_0}, \dots, \xi_{\mathcal{F}_m}$. Then for each $\mathbf{c} \in W_{\mathcal{F}_i}$, we have $\text{var}(\mathbf{c}^T \mathbf{y} | \boldsymbol{\beta}) = \|\mathbf{c}\|^2 \xi_{\mathcal{F}_i}$. The eigenspaces $W_{\mathcal{F}_0}, \dots, W_{\mathcal{F}_m}$ are called *strata* and $\xi_{\mathcal{F}_0}, \dots, \xi_{\mathcal{F}_m}$ are referred to as stratum variances. Furthermore,

$$(2.9) \quad \xi_{\mathcal{F}} = \sum_{\mathcal{G} \in \mathfrak{B}: \mathcal{G} \preceq \mathcal{F}} \frac{N}{n_{\mathcal{G}}} \sigma_{\mathcal{G}}^2.$$

Thus if $\mathcal{F}_i \preceq \mathcal{F}_j$, then $\xi_{\mathcal{F}_i} \leq \xi_{\mathcal{F}_j}$. The case where $\gamma_1^{\mathcal{F}_i}, \dots, \gamma_{n_{\mathcal{F}_i}}^{\mathcal{F}_i}$ are unknown constants (fixed effects) can be treated by letting $\sigma_{\mathcal{F}_i}^2 = \infty$, which leads to $\xi_{\mathcal{F}_j} = \infty$ if $\mathcal{F}_i \preceq \mathcal{F}_j$.

EXAMPLE 2.1. *A block design has the block structure $\mathfrak{B} = \{\mathcal{U}, \mathcal{B}, \mathcal{E}\}$. Let y_{ij} be the observation on the j th unit in the i th block, $i = 1, \dots, b$, $j = 1, \dots, k$. Then under (2.7),*

$$y_{ij} = \alpha_{t(i,j)} + \gamma^{\mathcal{U}} + \gamma_i^{\mathcal{B}} + \gamma_{i,j}^{\mathcal{E}},$$

where $t(i, j)$ is the label of the treatment assigned to the j th unit in the i th block. Write $\gamma^{\mathcal{U}}$ as μ , $\gamma_i^{\mathcal{B}}$ as β_i , and $\gamma_{i,j}^{\mathcal{E}}$ as $\epsilon_{i,j}$. Then μ , β_i 's, and $\epsilon_{i,j}$'s are independent random effects with zero means and $\text{var}(\mu) = \sigma_{\mathcal{U}}^2$, $\text{var}(\beta_i) = \sigma_{\mathcal{B}}^2$, $\text{var}(\epsilon_{i,j}) = \sigma_{\mathcal{E}}^2$. Each eigenvector in $W_{\mathcal{B}}$ is orthogonal to the vectors of 1's, with all the entries corresponding to the units in the same block being equal; thus it defines a between-block contrast. The eigenspace $W_{\mathcal{E}}$ is orthogonal to both $W_{\mathcal{U}}$ and $W_{\mathcal{B}}$. It follows that each of its vectors defines a between-unit contrast in the same block. We call $W_{\mathcal{B}}$ and $W_{\mathcal{E}}$ inter- and intra-block strata, respectively. For each $\mathbf{c} \in W_{\mathcal{B}}$ (respectively, $W_{\mathcal{E}}$), we have $\text{var}(\mathbf{c}^T \mathbf{y} | \boldsymbol{\alpha}) = \|\mathbf{c}\|^2 \xi_{\mathcal{B}}$ (respectively, $\|\mathbf{c}\|^2 \xi_{\mathcal{E}}$). The two eigenvalues $\xi_{\mathcal{B}}$ and $\xi_{\mathcal{E}}$ are called inter- and intra-block variances, respectively. By (2.9), we have $\xi_{\mathcal{B}} = k\sigma_{\mathcal{B}}^2 + \sigma_{\mathcal{E}}^2$ and $\xi_{\mathcal{E}} = \sigma_{\mathcal{E}}^2$. Thus $\xi_{\mathcal{B}} \geq \xi_{\mathcal{E}}$. The readers are referred to Chapter 5 of [4] for detailed derivations of strata for some simple block structures.

For each nonempty $S \subseteq \{1, \dots, n\}$, let \mathbf{u}_S be the column of \mathbf{U} corresponding to β_S . Then under a regular fractional factorial design, β_S is a treatment defining word if and only if $\mathbf{u}_S = \mathbf{1}_N$; otherwise, $\mathbf{1}_N^T \mathbf{u}_S = 0$. Thus the word count A_k can be expressed as

$$(2.10) \quad A_k = \frac{1}{N} \sum_{S: |S|=k} \|\mathbf{P}_{W_{\mathcal{U}}} \mathbf{u}_S\|^2,$$

where $\mathbf{P}_{W_{\mathcal{U}}}$ is the orthogonal projection matrix onto $W_{\mathcal{U}}$.

A regular multi-stratum fractional factorial design is said to be *orthogonal* if, for each column vector \mathbf{u} of \mathbf{U} , we have $\mathbf{u} \in W_{\mathcal{F}_i}$ for some i . In this case, we say that the factorial effect corresponding to \mathbf{u} (together with its aliases) is estimated in stratum $W_{\mathcal{F}_i}$. For example, the blocked regular designs constructed by the method described in Section 2.3 are orthogonal, with all the block defining effects estimated in the interblock stratum (we say that they are *confounded* with blocks) and the effects that are neither block nor treatment defining effects estimated in the intrablock stratum. In this case, for each nonempty $S \subseteq \{1, \dots, n\}$, we have that β_S is a block defining word if and only if $\mathbf{P}_{W_{\mathcal{B}}}\mathbf{u}_S = \mathbf{u}_S$, where $\mathbf{P}_{W_{\mathcal{B}}}$ is the orthogonal projection matrix onto $W_{\mathcal{B}}$. Let B_k be the number of block defining words of length k ; then similar to (2.10), we have

$$(2.11) \quad B_k = \frac{1}{N} \sum_{S:|S|=k} \|\mathbf{P}_{W_{\mathcal{B}}}\mathbf{u}_S\|^2.$$

We refer the readers to Chapters 13 and 14 of [4] for a detailed treatment of the construction of orthogonal regular multi-stratum designs and how to determine the factorial effects estimated in each stratum.

In general, for orthogonal regular multi-stratum designs, we define the word counts in various strata as follows: for $k = 0, \dots, n$ and $i = 0, \dots, m$,

$$B_{k,i} = \text{the number of } S \subseteq \{1, \dots, n\} \text{ such that } |S| = k \text{ and } \mathbf{u}_S \in W_{\mathcal{F}_i}.$$

Then

$$(2.12) \quad B_{k,i} = \frac{1}{N} \sum_{S:|S|=k} \|\mathbf{P}_{W_{\mathcal{F}_i}}\mathbf{u}_S\|^2.$$

Note that for regular fractional factorial designs with unstructured units, $(B_{1,0}, \dots, B_{n,0})$ is the usual wordlength pattern (A_1, \dots, A_n) . For nonregular and nonorthogonal designs, the word counts $B_{k,i}$ can be defined directly via (2.12). In the case of nonregular designs with unstructured units, this yields the generalized wordlength pattern in the minimum G_2 -aberration criterion introduced by Tang and Deng [20].

3. Optimality criteria. Bayesian optimality criteria, such as the D- and A-criteria, are based on the posterior covariance matrix $\text{cov}(\boldsymbol{\beta}|\mathbf{y})$. From (2.4) and (2.8), the posterior distribution of $\boldsymbol{\beta}$ given \mathbf{y} can easily be obtained by standard results on multivariate normal distributions. We have that $\boldsymbol{\beta}|\mathbf{y}$ is normal with

$$(3.1) \quad \text{cov}(\boldsymbol{\beta}|\mathbf{y}) = \boldsymbol{\Sigma}_{\beta} - \boldsymbol{\Sigma}_{\beta}\mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}_{\beta}\mathbf{U}^T + \mathbf{V})^{-1}\mathbf{U}\boldsymbol{\Sigma}_{\beta},$$

where $\boldsymbol{\Sigma}_{\beta} = 2^{-2n}\sigma^2\mathbf{P}^T\mathbf{R}\mathbf{P}$. We say that a design is Bayesian A-optimal if it minimizes $\text{tr}(\text{cov}(\boldsymbol{\beta}|\mathbf{y})) = \sum_{S \subseteq \{1, \dots, n\}} \text{var}(\beta_S|\mathbf{y})$, and Bayesian D-optimal if it minimizes $\det(\text{cov}(\boldsymbol{\beta}|\mathbf{y}))$.

For any treatment combination \mathbf{x} , $\alpha(\mathbf{x})$ can be expressed as $\mathbf{p}_\mathbf{x}^T \boldsymbol{\beta}$, where $\mathbf{p}_\mathbf{x}^T$ is the row of \mathbf{P} corresponding to \mathbf{x} . Thus the best linear unbiased predictor of $\alpha(\mathbf{x})$ given \mathbf{y} is $\mathbf{p}_\mathbf{x}^T E(\boldsymbol{\beta}|\mathbf{y})$, with conditional prediction error $E\{[\mathbf{p}_\mathbf{x}^T \boldsymbol{\beta} - \mathbf{p}_\mathbf{x}^T E(\boldsymbol{\beta}|\mathbf{y})]^2 | \mathbf{y}\} = E\{[\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})]^T \mathbf{p}_\mathbf{x} \mathbf{p}_\mathbf{x}^T [\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})] | \mathbf{y}\}$. It follows that the total conditional prediction error over all the treatment combinations is equal to

$$\begin{aligned} \sum_{\mathbf{x} \in \text{full factorial}} \text{var}(\mathbf{p}_\mathbf{x}^T \boldsymbol{\beta} | \mathbf{y}) &= \sum_{\mathbf{x} \in \text{full factorial}} E\{[\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})]^T \mathbf{p}_\mathbf{x} \mathbf{p}_\mathbf{x}^T [\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})] | \mathbf{y}\} \\ &= E\{[\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})]^T (\mathbf{P}^T \mathbf{P}) [\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{y})] | \mathbf{y}\} \\ &= 2^n \sum_{S \subseteq \{1, \dots, n\}} \text{var}(\beta_S | \mathbf{y}), \end{aligned}$$

where the last equality holds since $\mathbf{P}^T \mathbf{P} = 2^n \mathbf{I}_{2^n}$. Thus the Bayesian A-optimality is equivalent to minimizing the overall conditional prediction error of the treatment effects.

The following result gives an explicit form of $\text{cov}(\boldsymbol{\beta}|\mathbf{y})$ for orthogonal regular fractional factorial designs.

THEOREM 3.1. *Under an orthogonal regular 2^{n-p} design, where the experimental units have a block structure $\mathfrak{B} = \{\mathcal{F}_i : i = 0, \dots, m\}$ satisfying (2.1), (2.2), and (2.3), if β_S is estimated in $W_{\mathcal{F}_i}$, then*

$$(3.2) \quad \text{var}(\beta_S | \mathbf{y}) = \text{var}(\beta_S) - \frac{[\text{var}(\beta_S)]^2}{\sum_{\beta: \beta \in \mathcal{A}(\beta_S)} \text{var}(\beta) + 2^{-(n-p)} \xi_{\mathcal{F}_i}}.$$

If β_S and $\beta_{S'}$ are in different alias sets, then $\text{cov}(\beta_S, \beta_{S'} | \mathbf{y}) = 0$; if they are in the same alias set and are estimated in $W_{\mathcal{F}_i}$, then

$$\text{cov}(\beta_S, \beta_{S'} | \mathbf{y}) = - \frac{\text{var}(\beta_S) \text{var}(\beta_{S'})}{\sum_{\beta: \beta \in \mathcal{A}(\beta_S)} \text{var}(\beta) + 2^{-(n-p)} \xi_{\mathcal{F}_i}}.$$

PROOF. We first deal with the term $(\mathbf{U} \boldsymbol{\Sigma}_\beta \mathbf{U}^T + \mathbf{V})^{-1}$ in (3.1).

Note that $\mathbb{R}^N = \oplus_{\mathcal{F}: \mathcal{F} \in \mathfrak{B}} W_{\mathcal{F}}$, where $N = 2^{n-p}$. Also $\mathbb{R}^N = \oplus_{j=0}^{N-1} T_j$, where T_0 is spanned by the vector with all entries equal to 1, and each T_j , $1 \leq j \leq N-1$, is spanned by a factorial effect contrast of $n-p$ basic factors. Hence we have

$$\oplus_{\mathcal{F}: \mathcal{F} \in \mathfrak{B}} W_{\mathcal{F}} = \oplus_{j=0}^{N-1} T_j.$$

Since the design is orthogonal, each column of \mathbf{U} is a basis vector of $W_{\mathcal{F}_i}$ for some i . We can extract N distinct columns of \mathbf{U} and standardize them to unit length to form a $2^{n-p} \times 2^{n-p}$ orthogonal matrix $\tilde{\mathbf{P}}$. Then we have the following eigen-decomposition:

$$(3.3) \quad \mathbf{V} = \tilde{\mathbf{P}} \boldsymbol{\Lambda}_{\mathfrak{B}} \tilde{\mathbf{P}}^T,$$

where $\mathbf{\Lambda}_{\mathfrak{B}}$ is a diagonal matrix with a diagonal entry equal to $\xi_{\mathcal{F}_i}$ if the corresponding column in $\tilde{\mathbf{P}}$ is a basis vector of $W_{\mathcal{F}_i}$.

On the other hand, restricting the Gaussian process to the treatment combinations in the regular fractional factorial design, analogous to (2.5), we have

$$(3.4) \quad \text{cov}(\mathbf{X}_{\mathcal{T}}\boldsymbol{\alpha}) = 2^{n-p}\tilde{\mathbf{P}}\tilde{\mathbf{\Lambda}}_{\beta}\tilde{\mathbf{P}}^T,$$

where $\tilde{\mathbf{\Lambda}}_{\beta}$ is a $2^{n-p} \times 2^{n-p}$ diagonal matrix with each diagonal entry equal to $\sum_{S \in \mathcal{A}} \text{var}(\beta_S)$ for some alias set \mathcal{A} . Then since $\mathbf{U}\boldsymbol{\Sigma}_{\beta}\mathbf{U}^T = \text{cov}(\mathbf{U}\boldsymbol{\beta}) = \text{cov}(\mathbf{X}_{\mathcal{T}}\boldsymbol{\alpha})$, by (3.4), we have

$$(3.5) \quad \mathbf{U}\boldsymbol{\Sigma}_{\beta}\mathbf{U}^T = 2^{n-p}\tilde{\mathbf{P}}\tilde{\mathbf{\Lambda}}_{\beta}\tilde{\mathbf{P}}^T.$$

It follows from (3.1), (3.3), and (3.5) that

$$\text{cov}(\boldsymbol{\beta}|\mathbf{y}) = \boldsymbol{\Sigma}_{\beta} - \boldsymbol{\Sigma}_{\beta}\mathbf{U}^T\tilde{\mathbf{P}}(2^{n-p}\tilde{\mathbf{\Lambda}}_{\beta} + \mathbf{\Lambda}_{\mathfrak{B}})^{-1}\tilde{\mathbf{P}}^T\mathbf{U}\boldsymbol{\Sigma}_{\beta}.$$

The theorem is proved by noting that the inner product of a column of \mathbf{U} and a column of $\tilde{\mathbf{P}}$ is $2^{(n-p)/2}$ if they are in the same alias set and is zero otherwise. \square

Let $\boldsymbol{\xi}$ and \mathbf{v} be the vectors of the $\xi_{\mathcal{F}_i}$'s and $\text{var}(\beta_S)$'s, respectively. Since $\sum_{S: S \subseteq \{1, \dots, n\}} \text{var}(\beta_S)$ is a constant, by (3.2), among the orthogonal regular designs, a design d^* is Bayesian A-optimal for $(\boldsymbol{\xi}, \mathbf{v})$ if and only if it maximizes

$$\Phi_A(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{S: S \subseteq \{1, \dots, n\}} \frac{[\text{var}(\beta_S)]^2}{\sum_{\beta: \beta \in \mathcal{A}(\beta_S)} \text{var}(\beta) + e_S},$$

where $e_S = 2^{-(n-p)}\xi_{\mathcal{F}_i}$ if β_S is estimated in $W_{\mathcal{F}_i}$.

Under the prior distribution in (2.6), we have the following explicit form of $\Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$.

THEOREM 3.2. *Under (2.6), for an orthogonal regular fractional factorial design d ,*

$$\Phi_A(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{j=0}^{2^{n-p}-1} \frac{\tau^4 \sum_{i=1}^n r^{2i} N_i^{(j)}}{\tau^2 \sum_{i=1}^n r^i N_i^{(j)} + e_j},$$

where $e_j = 2^{-(n-p)}\xi_{\mathcal{F}_s}$ if the effects in \mathcal{A}_j are estimated in $W_{\mathcal{F}_s}$ and $N_i^{(j)}$ is the number of words of length i in \mathcal{A}_j . In particular, $(N_1^{(0)}, \dots, N_n^{(0)}) = (A_1, \dots, A_n)$ is the usual wordlength pattern.

THEOREM 3.3. *Among the orthogonal regular 2^{n-p} fractional factorial designs, a design d^* is Bayesian D-optimal if and only if it minimizes*

$$(3.6) \quad \Phi_D(d; \boldsymbol{\xi}, \mathbf{v}) = \prod_{j=0}^{2^{n-p}-1} \frac{e_j}{\sum_{\beta: \beta \in \mathcal{A}_j} \text{var}(\beta) + e_j}.$$

Under (2.6), $\Phi_D(d; \boldsymbol{\xi}, \mathbf{v}) = \prod_{j=0}^{2^{n-p}-1} \frac{e_j}{\tau^2 \sum_{i=1}^n r^i N_i^{(j)} + e_j}$.

PROOF. By (3.1), $\text{cov}(\boldsymbol{\beta}|\mathbf{y}) = \boldsymbol{\Sigma}_\beta \mathbf{W}$, where $\mathbf{W} = \mathbf{I}_{2^n} - \mathbf{U}^T(\mathbf{U}\boldsymbol{\Sigma}_\beta\mathbf{U}^T + \mathbf{V})^{-1}\mathbf{U}\boldsymbol{\Sigma}_\beta$. Since $\det(\boldsymbol{\Sigma}_\beta \mathbf{W}) = \det(\boldsymbol{\Sigma}_\beta) \det(\mathbf{W})$ and $\det(\boldsymbol{\Sigma}_\beta) = \prod_{S: S \subseteq \{1, \dots, n\}} \text{var}(\beta_S)$, which is a constant, it is sufficient to show that $\det(\mathbf{W})$ is equal to the quantity in (3.6).

By Theorem 3.1, $\text{cov}(\boldsymbol{\beta}|\mathbf{y})$ is a block-diagonal matrix, and so is \mathbf{W} . Let $\mathbf{W} = \text{diag}(\mathbf{W}_0, \dots, \mathbf{W}_{N-1})$, where each \mathbf{W}_j is a $2^p \times 2^p$ matrix corresponding to \mathcal{A}_j . Then $\det(\mathbf{W}) = \prod_{j=0}^{N-1} \det(\mathbf{W}_j)$. If the effects in \mathcal{A}_j are estimated in $W_{\mathcal{F}_i}$, then

$$\mathbf{W}_j = \frac{1}{T_j} \begin{pmatrix} T_j - x_1 & -x_2 & \cdots & -x_{2^p} \\ -x_1 & T_j - x_2 & \cdots & -x_{2^p} \\ \vdots & \vdots & \ddots & \vdots \\ -x_1 & -x_2 & \cdots & T_j - x_{2^p} \end{pmatrix},$$

where $T_j = \sum_{\beta: \beta \in \mathcal{A}_j} \text{var}(\beta) + 2^{-(n-p)} \xi_{\mathcal{F}_i}$ and the x_l 's are the $\text{var}(\beta)$'s for $\beta \in \mathcal{A}_j$. Thus

$$\begin{aligned} \det(\mathbf{W}_j) &= \frac{1}{(T_j)^{2^p}} \det \begin{pmatrix} T_j - x_1 & -x_2 & \cdots & T_j - \sum_{i=1}^{2^p} x_i \\ -x_1 & T_j - x_2 & \cdots & T_j - \sum_{i=1}^{2^p} x_i \\ \vdots & \vdots & \ddots & \vdots \\ -x_1 & -x_2 & \cdots & T_j - \sum_{i=1}^{2^p} x_i \end{pmatrix} \\ &= \frac{T_j - \sum_{i=1}^{2^p} x_i}{(T_j)^{2^p}} \det \begin{pmatrix} T_j - x_1 & -x_2 & \cdots & 1 \\ -x_1 & T_j - x_2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ -x_1 & -x_2 & \cdots & 1 \end{pmatrix} \\ &= \frac{T_j - \sum_{i=1}^{2^p} x_i}{(T_j)^{2^p}} \det \begin{pmatrix} T_j & 0 & \cdots & 0 \\ 0 & T_j & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ &= \frac{e_j}{\sum_{\beta: \beta \in \mathcal{A}_j} \text{var}(\beta) + e_j}. \end{aligned}$$

□

We note that in implementing the criteria in Theorem 3.2 and Theorem 3.3 under (2.6), one needs to specify the parameters τ, r and the $\xi_{\mathcal{F}_i}$'s.

When some unit factors have fixed effects, we can define the A- and D-criteria by letting the corresponding stratum variances go to infinity and taking the limits of $\Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$ and $\Phi_D(d; \boldsymbol{\xi}, \mathbf{v})$. This amounts to imposing improper priors on the corresponding fixed unit effects.

A useful result used in [7] is that for a Schur concave function $f(x_1, \dots, x_h)$, if (x_1, \dots, x_h) is majorized by (y_1, \dots, y_h) , then $f(x_1, \dots, x_h) \geq f(y_1, \dots, y_h)$. In addition, if f is increasing in each x_i , then a good surrogate for maximizing $f(x_1, \dots, x_h)$ is the (M.S)-criterion

due to Eccleston and Hedayat [10]: maximize $\sum_{i=1}^h x_i$, and then minimize $\sum_{i=1}^h x_i^2$ among those that maximize $\sum_{i=1}^h x_i$.

4. An analytical result. General analytical results in the present setting are difficult to prove. For unstructured units, minimum aberration half fractions were shown to be Bayesian A- and D-optimal in [15]. In the case of blocked 2^{n-1} fractional factorial designs with two blocks of size 2^{n-2} , we show below that, under some mild conditions, the minimum aberration designs according to a criterion of Cheng and Wu [9] are Bayesian D- and A-optimal. To construct a design, we need a treatment defining word to define the half fraction and another treatment interaction (block defining word) to divide the 2^{n-1} treatment combinations in the half fraction into two blocks. The block defining effect and its alias are then confounded with blocks. The other factorial effects are estimated in $W_{\mathcal{E}}$.

The following result can be used to help choose optimal block defining word when the treatment defining word is fixed.

LEMMA 4.1. *Among the 2^{n-1} designs with two blocks of size 2^{n-2} and a given treatment defining word of length L , those that confound two aliased treatment factorial effects involving $n - \lfloor \frac{L}{2} \rfloor$ and $n - \lceil \frac{L}{2} \rceil$ factors with blocks are Bayesian A- and D-optimal under the prior specified in (2.6).*

PROOF. Let d be a 2^{n-1} design with a treatment defining word of length L and two blocks of size 2^{n-2} . Under d , one pair of aliased words of lengths $u + x$ and $u + L - x$ are confounded with blocks, where $x \in \{0, 1, \dots, \lfloor L/2 \rfloor\}$ and $u \in \{0, 1, \dots, n - L\}$. Let $f(x, u) = \bar{\Phi}_A(d; \boldsymbol{\xi}, \mathbf{v}) - \Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$, where $\bar{\Phi}_A(d; \boldsymbol{\xi}, \mathbf{v})$ is obtained from $\Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$ by replacing $\xi_{\mathcal{B}}$ with $\xi_{\mathcal{E}}$. Then since $\Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$ is a constant not depending on the block defining word, maximizing $\Phi_A(d; \boldsymbol{\xi}, \mathbf{v})$ is the same as minimizing $f(x, u)$. We have

$$\begin{aligned}
(4.1) \quad f(x, u) &= \frac{\tau^4 r^{2u} (r^{2x} + r^{2L-2x})}{\tau^2 r^u (r^x + r^{L-x}) + \xi_{\mathcal{E}}/2^{n-1}} - \frac{\tau^4 r^{2u} (r^{2x} + r^{2L-2x})}{\tau^2 r^u (r^x + r^{L-x}) + \xi_{\mathcal{B}}/2^{n-1}} \\
&= \frac{\xi_{\mathcal{B}} - \xi_{\mathcal{E}}}{2^{n-1}} \cdot \frac{(r^x + r^{L-x})^2 - 2r^L}{\{r^x + r^{L-x} + \frac{\xi_{\mathcal{B}}/\tau^2}{2^{n-1}r^u}\} \cdot \{r^x + r^{L-x} + \frac{\xi_{\mathcal{E}}/\tau^2}{2^{n-1}r^u}\}} \\
&= \frac{\xi_{\mathcal{B}} - \xi_{\mathcal{E}}}{2^{n-1}} \cdot \frac{1}{\{1 + \frac{1}{r^x + r^{L-x}} \frac{\xi_{\mathcal{B}}/\tau^2}{2^{n-1}r^u}\} \{1 + \frac{1}{r^x + r^{L-x}} \frac{\xi_{\mathcal{E}}/\tau^2}{2^{n-1}r^u}\}} \\
&\quad - \frac{\xi_{\mathcal{B}} - \xi_{\mathcal{E}}}{2^{n-1}} \cdot \frac{2r^L}{\{r^x + r^{L-x} + \frac{\xi_{\mathcal{B}}/\tau^2}{2^{n-1}r^u}\} \{r^x + r^{L-x} + \frac{\xi_{\mathcal{E}}/\tau^2}{2^{n-1}r^u}\}}.
\end{aligned}$$

It is easy to see from (4.1) that when x is fixed, since $\xi_{\mathcal{B}} > \xi_{\mathcal{E}}$ and $0 < r < 1$, $f(x, u)$ attains the minimum at $u^* = n - L$, the largest possible value of u . Then from the last expression of $f(x, u)$ above, $f(x, n - L)$ is minimized at $x^* = \lfloor \frac{L}{2} \rfloor$ since $r^x + r^{L-x}$ attains its minimum at $x^* = \lfloor \frac{L}{2} \rfloor$. Thus $f(x, u)$ is minimized at $x^* = \lfloor \frac{L}{2} \rfloor$ and $u^* = n - L$. Then we have $u^* + x^* = n - \lceil \frac{L}{2} \rceil$ and $u^* + L - x^* = n - \lfloor \frac{L}{2} \rfloor$.

For the D-optimality, by Theorem 3.3, minimizing $\Phi_D(d; \boldsymbol{\xi}, \mathbf{v})$ is equivalent to maximizing $\prod_{i=0}^{2^{n-1}-1} (v_i + \xi_i/2^{n-1})$, where each v_i is of the form $\tau^2 r^u (r^x + r^{L-x})$, exactly one of the ξ_i 's is equal to ξ_B , and all the others are equal to ξ_E . Since $\prod_{i=0}^{2^{n-1}-1} (v_i + \xi_i/2^{n-1})$ is a Schur concave function of $(v_0 + \xi_0/2^{n-1}, \dots, v_{2^{n-1}-1} + \xi_{2^{n-1}-1}/2^{n-1})$, a blocking scheme is D-optimal if its $(v_0 + \xi_0/2^{n-1}, \dots, v_{2^{n-1}-1} + \xi_{2^{n-1}-1}/2^{n-1})$ is majorized by that of any other blocking scheme. Without loss of generality, suppose $v_0 \geq \dots \geq v_{2^{n-1}-1}$. Then since $\xi_B > \xi_E$, it is easy to see that the blocking scheme with $\xi_{2^{n-1}-1} = \xi_B$ produces a $(v_0 + \xi_0/2^{n-1}, \dots, v_{2^{n-1}-1} + \xi_{2^{n-1}-1}/2^{n-1})$ that is majorized by that of any other blocking scheme. Thus it suffices to show that $r^u (r^x + r^{L-x})$ is minimized at $u^* = n - L$ and $x^* = \lfloor \frac{L}{2} \rfloor$. This holds since $r^u (r^x + r^{L-x})$ is a decreasing function of both u and x ($\leq L/2$), and u^* and x^* are the largest possible values of u and x , respectively. \square

EXAMPLE 4.1. *Among the orthogonal regular 2^{4-1} designs with two blocks of size 4 and treatment defining word $\mathbf{1}$ (respectively, $\mathbf{12}$, $\mathbf{123}$, or $\mathbf{1234}$), the one with the block defining word $\mathbf{234}$ (= $\mathbf{1234}$), (respectively, $\mathbf{134}$ (= $\mathbf{234}$), $\mathbf{124}$ (= $\mathbf{34}$), or $\mathbf{24}$ (= $\mathbf{13}$)), is the best under the Bayesian A- and D-criteria.*

We now show that under some mild conditions the best design in Lemma 4.1 with $L = n$ (that is, a maximum resolution design) is a Bayesian D-optimal 2^{n-1} design with two blocks of size 2^{n-2} .

THEOREM 4.1. *Let d^* be the 2^{n-1} resolution n design with two blocks of size 2^{n-2} that confounds two aliased treatment factorial effects involving $n - \lceil \frac{n}{2} \rceil$ and $n - \lfloor \frac{n}{2} \rfloor$ factors with blocks. For a given $h > 0$, suppose*

1. $r \leq (\sqrt{h^2 + 4} - h)^2/4$,
2. $\xi_B \leq (2^{n-1} \tau^2 r^{n-1.5})h$.

Then d^ is Bayesian D-optimal among all the orthogonal regular 2^{n-1} designs with two blocks of size 2^{n-2} under the prior specified in (2.6).*

REMARK 4.1. *It can be seen that the optimal design in Theorem 4.1 has minimum aberration with respect to the wordlength pattern W_1 proposed in [9]. We note that the assumption $\xi_B \leq (2^{n-1} \tau^2 r^{n-1.5})h$ is equivalent to $\xi_B \leq (2^n \tau^2 r^n) (\frac{r^{-1.5}}{2})h$, where $2^n \tau^2 r^n$ is the prior variance of a normalized n -factor interaction. It can be seen that $f(h) = (\sqrt{h^2 + 4} - h)^2/4$, the upper bound on r in Theorem 4.1, is decreasing in h . Therefore when h is large (implying ξ_B can be large), d^* is expected to be optimal under smaller r 's. Kang and Joseph [14] argued that $r = 1/3$ is a good choice when there is no other prior knowledge. We have $f(1.15) \approx 1/3$, indicating that for $r = 1/3$, the result in Theorem 4.1 holds for $\xi_B \leq (2^n \tau^2 r^n) (3^{1.5} \cdot \frac{1.15}{2}) \approx 2.988(2^n \tau^2 r^n)$.*

PROOF OF THEOREM 4.1. It is sufficient to show that $\Phi_D(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_D(d_2; \boldsymbol{\xi}, \mathbf{v})$ for any two blocked 2^{n-1} designs d_1 and d_2 with treatment defining words $\{1, \dots, L\}$ and

$\{1, \dots, L-1\}$, respectively ($L \geq 3$), and the block defining words as prescribed in Lemma 4.1.

Let $H_1 = \{1, \dots, L\}$, $H_2 = \{1, \dots, L-1\}$, and $U_0 = \{I, H_1, H_2, H_1 \Delta H_2\}$, where $H_1 \Delta H_2 = (H_1 \cup H_2) \setminus (H_1 \cap H_2)$. Then all the 2^n words can be partitioned into 2^{n-2} classes, each of which other than U_0 is a coset of U_0 with respect to the group operation Δ . Denote these cosets by $U_j, j = 1, \dots, 2^{n-2} - 1$. Then each U_j consists of two pairs of aliased words under both d_1 and d_2 . It follows from Theorem 3.3 that for $d = d_1$ or d_2 , $\Phi_D(d; \boldsymbol{\xi}, \mathbf{v}) = \prod_{j=0}^{2^{n-2}-1} \Phi_{D,j}(d; \boldsymbol{\xi}, \mathbf{v})$, where $\Phi_{D,j}(d; \boldsymbol{\xi}, \mathbf{v})$ is the contribution of U_j to the product in (3.6).

Under d_1 , H_1 is aliased with I (which corresponds to the mean). Thus H_1 and I (of lengths L and 0 , respectively) are the only effects estimated in $W_{\mathcal{U}}$. The other two words in U_0 , H_2 and $H_1 \Delta H_2$, have lengths $L-1$ and 1 , respectively, and cannot be estimated in $W_{\mathcal{B}}$, since, by Lemma 4.1, the two effects confounded with blocks have lengths $n - \lceil \frac{L}{2} \rceil$ and $n - \lfloor \frac{L}{2} \rfloor$, respectively. Thus H_2 and $H_1 \Delta H_2$ are estimated in $W_{\mathcal{E}}$. Likewise, under d_2 , I and H_2 are estimated in $W_{\mathcal{U}}$, and the other two effects in U_0 , H_1 and $H_1 \Delta H_2$, are estimated in $W_{\mathcal{E}}$.

Without loss of generality, suppose that under d_1 , the two effects estimated in $W_{\mathcal{B}}$, say H and $H \Delta H_1$, are in U_1 . By Lemma 4.1, we may assume that $|H| = n - \lceil \frac{L}{2} \rceil$ and $|H \Delta H_1| = n - \lfloor \frac{L}{2} \rfloor$. Then $|H \Delta H_1| = n - \lceil \frac{L-1}{2} \rceil$ and $|H \Delta (H_1 \Delta H_2)| = n - \lfloor \frac{L-1}{2} \rfloor$. It follows that $H \Delta H_1$ and $H \Delta (H_1 \Delta H_2)$ are the two words that are aliased and estimated in $W_{\mathcal{B}}$ under d_2 . Then $H \Delta (H_1 \Delta H_2)$ and $H \Delta H_2$ are aliased and estimated in $W_{\mathcal{E}}$ under d_1 , while H and $H \Delta H_2$ are aliased and estimated in $W_{\mathcal{E}}$ under d_2 .

For all $j \geq 2$, all the effects in U_j are estimated in $W_{\mathcal{E}}$. Based on this and the information from the previous two paragraphs, one can write down $\Phi_{D,0}(d_1; \boldsymbol{\xi}, \mathbf{v})$, $\Phi_{D,0}(d_2; \boldsymbol{\xi}, \mathbf{v})$, $\Phi_{D,1}(d_1; \boldsymbol{\xi}, \mathbf{v})$, and $\Phi_{D,1}(d_2; \boldsymbol{\xi}, \mathbf{v})$. When the experimental units are unstructured, Kerr [15] proved $\Phi_D(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_D(d_2; \boldsymbol{\xi}, \mathbf{v})$ by showing that $\Phi_{D,j}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,j}(d_2; \boldsymbol{\xi}, \mathbf{v})$ for all $j = 0, \dots, 2^{n-2} - 1$. Effectively she has shown that $\Phi_{D,j}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,j}(d_2; \boldsymbol{\xi}, \mathbf{v})$ for all $j \geq 2$ in our setting. Thus it is sufficient to show $\Phi_{D,j}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,j}(d_2; \boldsymbol{\xi}, \mathbf{v})$ for $j = 0, 1$.

For $\Phi_{D,0}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,0}(d_2; \boldsymbol{\xi}, \mathbf{v})$, we need to show

$$(1 + r^L + \frac{\xi_{\mathcal{U}}}{2^{n-1}\tau^2})(r + r^{L-1} + \frac{\xi_{\mathcal{E}}}{2^{n-1}\tau^2}) \geq (1 + r^{L-1} + \frac{\xi_{\mathcal{U}}}{2^{n-1}\tau^2})(r + r^L + \frac{\xi_{\mathcal{E}}}{2^{n-1}\tau^2}).$$

This inequality is equivalent to $\frac{\xi_{\mathcal{U}}}{2^{n-1}\tau^2} \geq \frac{\xi_{\mathcal{E}}}{2^{n-1}\tau^2} - 1 + r$, which always holds since $\xi_{\mathcal{U}} \geq \xi_{\mathcal{E}}$ and $r < 1$.

For $\Phi_{D,1}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,1}(d_2; \boldsymbol{\xi}, \mathbf{v})$ when L is odd, we need to show

$$(r^{-\frac{1}{2}} + r^{\frac{1}{2}} + \delta_{\mathcal{B}})(r^{-\frac{1}{2}} + r^{\frac{1}{2}} + \delta_{\mathcal{E}}) \geq (2r^{\frac{1}{2}} + \delta_{\mathcal{B}})(2r^{-\frac{1}{2}} + \delta_{\mathcal{E}}),$$

where $\delta_{\mathcal{B}} = \frac{2^{-(n-1)}\xi_{\mathcal{B}}}{\tau^2 r^{n-L/2}}$ and $\delta_{\mathcal{E}} = \frac{2^{-(n-1)}\xi_{\mathcal{E}}}{\tau^2 r^{n-L/2}}$. This is equivalent to

$$(4.2) \quad 1 - r - (\delta_{\mathcal{B}} - \delta_{\mathcal{E}})r^{\frac{1}{2}} \geq 0.$$

It follows from the assumption $\xi_{\mathcal{B}} \leq \{2^{n-1}\tau^2r^{n-1.5}\}h$ that $0 < \delta_{\mathcal{E}} \leq \delta_{\mathcal{B}} \leq h$. Then a sufficient condition for (4.2) is $1 - r - hr^{\frac{1}{2}} \geq 0$. This holds if $r \leq (\sqrt{h^2 + 4} - h)^2/4$.

For $\Phi_{D,1}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,1}(d_2; \boldsymbol{\xi}, \mathbf{v})$ when L is even, we need to show

$$(2 + \delta_{\mathcal{B}})(r^{-1} + r + \delta_{\mathcal{E}}) \geq (1 + r + \delta_{\mathcal{B}})(1 + r^{-1} + \delta_{\mathcal{E}}),$$

which can be simplified as

$$(4.3) \quad 1 - r - r(\delta_{\mathcal{B}} - \delta_{\mathcal{E}}) \geq 0.$$

Since $r^{1/2} > r$, (4.3) is implied by (4.2). \square

We can also establish a version of Theorem 4.1 for the Bayesian A-criterion by replacing the conditions $\Phi_{D,i}(d_1; \boldsymbol{\xi}, \mathbf{v}) \leq \Phi_{D,i}(d_2; \boldsymbol{\xi}, \mathbf{v})$, $i = 0, 1$, in the proof with $\Phi_{A,i}(d_1; \boldsymbol{\xi}, \mathbf{v}) \geq \Phi_{A,i}(d_2; \boldsymbol{\xi}, \mathbf{v})$, $i = 0, 1$.

5. More complicated block structures: a surrogate criterion. For a setting as simple as regular half fractions with two blocks, the proof of Theorem 4.1 is already quite tedious. The determination of optimal designs with more complicated block structures is very challenging. In this section, we derive a more manageable two-stage surrogate criterion that can be applied to nonregular designs as well. Notice that the popular minimum aberration criterion is itself a surrogate for another statistically more meaningful criterion; see [5]. Interestingly the first stage of our surrogate criterion can be expressed in terms of wordlength and generalized wordlength patterns for regular and nonregular designs, respectively. The surrogate criterion depends on the stratum variances, but as in [7], a useful tool is developed for eliminating many inferior designs without having to know the stratum variances; then one only needs to select from a much smaller ‘‘essentially complete class’’.

We first note that another expression of $\text{cov}(\boldsymbol{\beta}|\mathbf{y})$ instead of (3.1) is

$$\text{cov}(\boldsymbol{\beta}|\mathbf{y}) = \left(\mathbf{U}^T \mathbf{V}^{-1} \mathbf{U} + \boldsymbol{\Sigma}_{\beta}^{-1} \right)^{-1} = \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \boldsymbol{\Sigma}_{\beta}^{-1} \right)^{-1}.$$

Thus we have

$$\det[\text{cov}(\boldsymbol{\beta}|\mathbf{y})^{-1}] = \det \left(\boldsymbol{\Sigma}_{\beta}^{-1} \right) \det \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_{\beta} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right).$$

Since $\det(\boldsymbol{\Sigma}_{\beta}^{-1})$ is a constant, maximizing $\det[\text{cov}(\boldsymbol{\beta}|\mathbf{y})^{-1}]$ is equivalent to maximizing $\det \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_{\beta} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right)$. A design is said to be Bayesian (M.S)-optimal if it maximizes $\text{tr} \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_{\beta} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right)$ and minimizes $\text{tr} \left[\left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_{\beta} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right)^2 \right]$ among the designs that maximize $\text{tr} \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_{\beta} \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right)$. Since $\text{tr}(\mathbf{A}\mathbf{B}) =$

$\text{tr}(\mathbf{BA})$ for any two compatible matrices \mathbf{A} and \mathbf{B} , and $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^T$ for any orthogonal projection matrix \mathbf{P} , we have

$$\begin{aligned} \text{tr} \left(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n} \right) &= \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \text{tr} \left[\boldsymbol{\Sigma}_\beta^{1/2} (\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} \boldsymbol{\Sigma}_\beta^{1/2})^T (\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U}) \right] + 2^n \\ &= \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \text{tr} \left[(\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} \boldsymbol{\Sigma}_\beta^{1/2})^T (\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} \boldsymbol{\Sigma}_\beta^{1/2}) \right] + 2^n. \end{aligned}$$

If all the factorial effects involving k factors have identical prior variance v_k , $k = 0, 1, \dots, n$ (say, under (2.6)), then

$$\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \text{tr} \left[(\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} \boldsymbol{\Sigma}_\beta^{1/2})^T (\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} \boldsymbol{\Sigma}_\beta^{1/2}) \right] = \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \sum_{k=0}^n v_k \left(\sum_{S:|S|=k} \|\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{u}_S\|^2 \right).$$

Since $\sum_{i=0}^m \mathbf{P}_{W_{\mathcal{F}_i}} = \mathbf{I}_N$, one can replace $\mathbf{P}_{W_{\mathcal{F}_m}}$ with $\mathbf{I}_N - \sum_{i=0}^{m-1} \mathbf{P}_{W_{\mathcal{F}_i}}$. Then

$$\begin{aligned} \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \sum_{k=0}^n v_k \left(\sum_{S:|S|=k} \|\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{u}_S\|^2 \right) &= c - \sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m} - \xi_{\mathcal{F}_i}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \sum_{k=0}^n v_k \left(\sum_{S:|S|=k} \|\mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{u}_S\|^2 \right) \\ &= c - N \sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m} - \xi_{\mathcal{F}_i}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \sum_{k=0}^n v_k B_{k,i}(d), \end{aligned}$$

where $c = \frac{1}{\xi_{\mathcal{F}_m}} \sum_{k=0}^n v_k \sum_{S:|S|=k} \|\mathbf{u}_S\|^2$, which is a constant, and $B_{k,i}(d)$ is as defined in (2.12). Furthermore, since $B_{0,0}(d) = 1$ and $B_{0,i}(d) = 0$ for all $i > 0$, maximizing $\text{tr}(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n})$ is equivalent to minimizing

$$(5.1) \quad \Phi_M(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m} - \xi_{\mathcal{F}_i}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) \sum_{k=1}^n v_k B_{k,i}(d).$$

For the designs that minimize $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$, minimizing $\text{tr}[(\sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}} \boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U} + \mathbf{I}_{2^n})^2]$ reduces to minimizing

$$\begin{aligned} \Phi_S(d; \boldsymbol{\xi}, \mathbf{v}) &= \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}^2} \text{tr} \left[(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U})^2 \right] \\ &\quad + 2 \sum_{0 \leq l < s \leq m} \frac{1}{\xi_{\mathcal{F}_l} \xi_{\mathcal{F}_s}} \text{tr} \left[(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_l}} \mathbf{U}) (\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_s}} \mathbf{U}) \right]. \end{aligned}$$

Thus a Bayesian (M.S)-optimal design d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ and minimizes $\Phi_S(d; \boldsymbol{\xi}, \mathbf{v})$ among those that minimize $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$. Note that $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ can be expressed in terms of the generalized wordlength patterns.

Under orthogonal regular designs, $\Phi_S(d; \boldsymbol{\xi}, \mathbf{v})$ can further be simplified. It can be seen that in this case $(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_l}} \mathbf{U})(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_s}} \mathbf{U})$ is a zero matrix when $l \neq s$. Thus we have

$$\Phi_S(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^m \frac{1}{\xi_{\mathcal{F}_i}^2} \text{tr} \left[(\boldsymbol{\Sigma}_\beta \mathbf{U}^T \mathbf{P}_{W_{\mathcal{F}_i}} \mathbf{U})^2 \right].$$

For $i = 1, \dots, m$, let h_i be the number of alias sets estimated in $W_{\mathcal{F}_i}$, and $N_k^{(i,j)}(d)$ be the number of words of length k in the j th alias set estimated in $W_{\mathcal{F}_i}$, $j = 1, \dots, h_i$. For $i = 0$, let $h_0 = 1$ and $N_k^{(0,1)}(d) = B_{k,0}(d)$. Then $B_{k,i}(d) = \sum_{j=1}^{h_i} N_k^{(i,j)}(d)$. Some routine calculation yields

$$(5.2) \quad \Phi_S(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^m \frac{N^2}{\xi_{\mathcal{F}_i}^2} \sum_{j=1}^{h_i} \left[\sum_{k=0}^n v_k N_k^{(i,j)}(d) \right]^2.$$

Furthermore, if d is a complete factorial design, then $B_{k,0}(d) = 0$ for $k \geq 1$ and $\sum_{k=0}^n N_k^{(i,j)} = 1$ for all i, j . It follows that $\sum_{j=1}^{h_i} \left[\sum_{k=0}^n v_k N_k^{(i,j)}(d) \right]^2 = \sum_{k=0}^n v_k^2 B_{k,i}(d)$. Then since $\sum_{i=0}^m B_{k,i}(d) = C_k^n$, by (5.2) we have

$$(5.3) \quad \Phi_S(d; \boldsymbol{\xi}, \mathbf{v}) = c' - N^2 \sum_{i=1}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}^2} - \frac{1}{\xi_{\mathcal{F}_i}^2} \right) \sum_{k=1}^n v_k^2 B_{k,i}(d),$$

where c' is a constant. Thus an orthogonal multi-stratum complete factorial design is Bayesian (M.S)-optimal if it maximizes $\sum_{i=1}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}^2} - \frac{1}{\xi_{\mathcal{F}_i}^2} \right) \sum_{k=1}^n v_k^2 B_{k,i}(d)$ among those that minimize $\sum_{i=1}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}^2} - \frac{1}{\xi_{\mathcal{F}_i}^2} \right) \sum_{k=1}^n v_k B_{k,i}(d)$.

PROPOSITION 5.1. *In the case of unstructured experimental units, a regular design minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ if and only if it minimizes $\sum_{k=1}^n v_k B_{k,0}(d)$. If $v_k \gg v_{k+1}$ for all k , then a good surrogate is to sequentially minimize $B_{1,0}(d), \dots, B_{n,0}(d)$; that is, the usual minimum aberration criterion.*

PROOF. For unstructured units, $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1\}$. By (5.1), minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ is equivalent to minimizing $\sum_{k=1}^n v_k B_{k,0}(d)$, where $(B_{1,0}(d), \dots, B_{n,0}(d))$ is the wordlength pattern of d . \square

The condition $v_k \gg v_{k+1}$ for all k amounts to that the lower-order effects are much more important than high-order ones. For unstructured units, Proposition 5.1 formally connects our criterion with minimum aberration, a popular criterion under the commonly made effect hierarchy assumption when the prior knowledge about the underlying system is vague. Similarly, for nonregular designs, generalized minimum aberration is a good surrogate for

minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$. Thus the Bayesian (M.S)-criterion developed here can be viewed as a refinement of minimum and generalized minimum aberration. One may also need to carry out the S-step.

Under the same assumption as in Proposition 5.1, it was shown in [12] that minimum aberration designs minimize the posterior variance of the mean (intercept), but this result cannot be extended to nonregular designs. As mentioned earlier, minimum aberration regular half-fractions were shown to be Bayesian D- and A-optimal in [15].

Blocking of complete 2^n factorial designs is related to the construction of 2^{n-p} regular fractional factorial designs in that a defining contrast subgroup that defines a fractional factorial design divides the 2^n treatment combinations into 2^p sets of size 2^{n-p} . Each of the 2^p sets can be used as a 2^{n-p} fractional factorial design. On the other hand, these sets together form 2^p blocks of a complete factorial. The treatment defining words of a fractional factorial design are those that are confounded with blocks in the corresponding blocking of the complete factorial. Under the hierarchical assumption, it is desirable not to confound too many lower order effects with blocks. Sun, Wu, and Chen [19] extended the minimum aberration criterion from fractional factorial designs to blocked complete factorial designs. A blocked complete factorial design is said to have minimum aberration if it sequentially minimizes B_1, B_2, \dots , where B_i is the number of block defining words of length i . A result similar to Proposition 5.1 holds for optimal blocking of complete factorial designs.

PROPOSITION 5.2. *Suppose the 2^n treatment combinations in a complete factorial design are to be divided into 2^p blocks of size 2^{n-p} . If $v_k \gg v_{k+1}$ for all k , then minimum aberration is a good surrogate for the Bayesian (M.S)-optimality over orthogonal designs.*

PROOF. In this case, $m = 2$. By the discussions preceding Proposition 5.1, since $\xi_{\mathcal{F}_1} > \xi_{\mathcal{F}_2}$, the M-step is to minimize $\sum_{k=1}^n v_k B_{k,1}(d)$ and the S-step is to maximize $\sum_{k=1}^n v_k^2 B_{k,1}(d)$. When $v_k \gg v_{k+1}$ for all k , minimizing $\sum_{k=1}^n v_k B_{k,1}(d)$ is achieved by sequentially minimizing $B_{1,1}(d), B_{2,1}(d), \dots$, i.e., minimum aberration. By the same assumption, the designs that minimize $\sum_{k=1}^n v_k B_{k,1}(d)$ have the same $B_{k,1}$ values. It follows that they also have the same value of $\sum_{k=1}^n v_k^2 B_{k,1}(d)$. Thus the S-step is not needed. \square

In the following theorem we provide a necessary and sufficient condition for a design to minimize $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all *feasible* stratum variances. Here $\boldsymbol{\xi}$ is said to be feasible if $\mathcal{F}_i \prec \mathcal{F}_j \Rightarrow \xi_{\mathcal{F}_i} \leq \xi_{\mathcal{F}_j}$.

THEOREM 5.1. *Suppose \mathfrak{B} is a block structure satisfying (2.1), (2.2), and (2.3). Then a necessary and sufficient condition for a design d^* to minimize $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\boldsymbol{\xi}$ is that it minimizes $\sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d)$ for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ such that*

$$(5.4) \quad \mathcal{F} \in \mathfrak{G}, \mathcal{F}' \in \mathfrak{B}, \text{ and } \mathcal{F} \prec \mathcal{F}' \Rightarrow \mathcal{F}' \in \mathfrak{G}.$$

We first show the necessity part since it is straightforward and is useful for understanding the implications and utilities of the theorem. The proof of the sufficiency part will be

deferred to the end of this section. Suppose d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\boldsymbol{\xi}$. For any $\mathfrak{G} \in \mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4), let $\boldsymbol{\xi}$ be such that $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_0}$ for all $\mathcal{F} \in \mathfrak{G}$ and $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_m}$ for all $\mathcal{F} \notin \mathfrak{G}$. Then since $\xi_{\mathcal{F}_m} < \xi_{\mathcal{F}_0}$, it is easy to see that such a $\boldsymbol{\xi}$ is feasible. In this case,

$$(5.5) \quad \Phi_M(d; \boldsymbol{\xi}, \mathbf{v}) = \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_0}} \right) \sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d).$$

It follows that d^* minimizes $\sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d)$. This proves the necessity part.

By (5.5), we can interpret the conclusion of Theorem 5.1 as that d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\boldsymbol{\xi}$ if and only if, for all \mathfrak{B} satisfying (5.4), it minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for the cases with $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_0}$ for all $\mathcal{F} \in \mathfrak{G}$ and $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_m}$ for all $\mathcal{F} \notin \mathfrak{G}$. Since $\sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d)$ does not depend on $\boldsymbol{\xi}$ and only involves the unit factors in \mathfrak{G} , it is enough to consider only the case $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_0} = \infty$ for all $\mathcal{F} \in \mathfrak{G}$ and $\xi_{\mathcal{F}} = \xi_{\mathcal{F}_m}$ for all $\mathcal{F} \notin \mathfrak{G}$. In other words, in verifying the necessary and sufficient condition, one only needs to consider the block structures $\mathfrak{G} \cup \{\mathcal{E}\}$ (note that $\mathcal{F}_m = \mathcal{E}$), where all the unit factors in \mathfrak{G} have fixed effects. Thus by checking the minimization of $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for a small number of extreme cases with fixed unit effects (and no knowledge about $\boldsymbol{\xi}$ is needed to do this), one is able to conclude the strong property of minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all feasible $\boldsymbol{\xi}$. Even if a design with such a strong optimality property does not exist, suppose $\sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d_1) \leq \sum_{i: \mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d_2)$ for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4), with strict inequality for at least one such \mathfrak{G} ; then d_2 is worse than d_1 and we say that d_2 is *inadmissible*. This provides a simple way of eliminating inferior designs and substantially reducing the designs that need to be considered. Since many designs are ruled out from consideration, one can also use the actual Bayesian A- and D-criterion values to compare the remaining designs.

By (5.3) and the same argument as in the proof of Theorem 5.1, the following result for complete factorial designs can be established.

THEOREM 5.2. *Let d^* be an orthogonal 2^n complete factorial design with a block structure \mathfrak{B} satisfying (2.1), (2.2), and (2.3). Then d^* is Bayesian (M.S)-optimal for all feasible $\boldsymbol{\xi}$ if and only if, for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4), d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ and minimizes $\Phi_S(d; \boldsymbol{\xi}, \mathbf{v})$ among the designs that minimize $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$, where $\xi_{\mathcal{F}_i} = \infty$ for all $\mathcal{F}_i \in \mathfrak{G}$ and $\xi_{\mathcal{F}_i} = \xi_{\mathcal{F}_m}$ for all $\mathcal{F}_i \notin \mathfrak{G}$.*

EXAMPLE 5.1 (A chain of nested unit factors). *Suppose $\mathfrak{B} = \{\mathcal{F}_0, \dots, \mathcal{F}_m\}$, where $\mathcal{F}_i \prec \mathcal{F}_j$ for all $i > j$. The \mathfrak{G} 's that satisfy (5.4) are the m sets $\{\mathcal{F}_0\}, \{\mathcal{F}_0, \mathcal{F}_1\}, \dots, \{\mathcal{F}_0, \dots, \mathcal{F}_{m-1}\}$. It follows from Theorem 5.1 that a design d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\boldsymbol{\xi}$ such that $i > j \Rightarrow \xi_{\mathcal{F}_i} \leq \xi_{\mathcal{F}_j}$ provided that it minimizes $\sum_{i=0}^l \sum_{k=1}^n v_k B_{k,i}(d)$ for each $l = 0, 1, \dots, m-1$. By the discussions preceding Theorem 5.2, minimizing $\sum_{i=0}^l \sum_{k=1}^n v_k B_{k,i}(d)$ is the same as minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ under the block structure \mathfrak{B} with $\xi_{\mathcal{F}_0} = \dots = \xi_{\mathcal{F}_l} = \infty$ and $\xi_{\mathcal{F}_{l+1}} = \dots = \xi_{\mathcal{F}_m}$. It can be seen that (i) for $l > 0$, this is equivalent to minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ under the block structure $\mathfrak{B}_l = \{\mathcal{F}_0, \mathcal{F}_l, \mathcal{F}_m\}$ with $\xi_{\mathcal{F}_l} = \xi_{\mathcal{F}_0} = \infty$, i.e., an*

experiment with fixed block effects, where each block consists of the units with the same level of \mathcal{F}_l , and (ii) for $l = 0$, it is equivalent to minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ under the block structure $\mathfrak{B}_0 = \{\mathcal{F}_0, \mathcal{F}_m\}$, i.e., an experiment with unstructured units. In particular, for $\mathfrak{B} = \{\mathcal{E}, \mathcal{B}, \mathcal{U}\}$, a blocked factorial experiment, if a design d^* minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for an experiment with fixed block effects ($\xi_{\mathcal{B}} = \infty$) as well as an experiment with unstructured units ($\xi_{\mathcal{B}} = \xi_{\mathcal{E}}$), then it minimizes $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ for all $\xi_{\mathcal{B}} \geq \xi_{\mathcal{E}}$.

Under $v_1 \gg \cdots \gg v_n$, a good surrogate for minimizing $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ is to sequentially minimize $\sum_{i=0}^{m-1} (\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}) B_{1,i}(d), \dots, \sum_{i=0}^{m-1} (\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}) B_{n,i}(d)$. Thus the minimization of $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v})$ leads to a minimum aberration criterion based on the wordlength pattern

$$(5.6) \quad \left(\sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) B_{1,i}(d), \dots, \sum_{i=0}^{m-1} \left(\frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}} \right) B_{n,i}(d) \right).$$

Such a criterion depends on the stratum variances. On the other hand, a good surrogate for minimizing $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} \sum_{k=1}^n v_k B_{k,i}(d)$ is to sequentially minimize $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_{1,i}(d), \dots, \sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_{n,i}(d)$. For each subset \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4), this induces a minimum aberration criterion based on the wordlength pattern

$$(5.7) \quad \left(\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_{1,i}(d), \dots, \sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_{n,i}(d) \right).$$

Similar to Theorem 5.1, the following holds for such surrogate minimum aberration criteria.

COROLLARY 5.1. *A necessary and sufficient condition for a design d^* to have minimum aberration with respect to the wordlength pattern (5.6) for all feasible $\boldsymbol{\xi}$ is that it has minimum aberration with respect to (5.7) for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4).*

It follows that if d_1 is at least as good as d_2 with respect to the MA criterion based on (5.7) for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4), and is better than d_2 for at least one such \mathfrak{G} , then d_2 is dominated by d_1 for feasible $\boldsymbol{\xi}$'s with respect to the MA criterion based on (5.6) and can be ruled out from consideration. The conclusion drawn for experiments with a chain of nested unit factors (in particular, blocked factorial experiments) studied in Example 5.1 also applies to the surrogate MA criteria formulated above.

REMARK 5.1. *Theorem 5.1 has a similar flavor to Theorem 4.1 of [7], which provided a necessary and sufficient condition for a design to be optimal with respect to a surrogate for the maximum information capacity criterion considered there for all feasible $\boldsymbol{\xi}$. In applying Theorem 4.1 of [7], one has to check certain conditions for all subsets \mathfrak{G}' of $\mathfrak{B} \setminus \{\mathcal{F}_0\}$ such that*

$$(5.8) \quad \mathcal{F} \in \mathfrak{G}', \mathcal{F}' \in \mathfrak{B} \text{ and } \mathcal{F}' \prec \mathcal{F} \Rightarrow \mathcal{F}' \in \mathfrak{G}'.$$

We note that (5.4) is equivalent to (5.8) if we take \mathfrak{G}' to be $\mathfrak{B} \setminus \mathfrak{G}$. This, coupled with some tedious technical arguments which are omitted here, indicates that optimal orthogonal regular designs based on the approach presented in this paper are expected to perform well, if not optimally, under the criterion studied in [7]. In addition to the assumption that the three-factor and higher-order interactions are negligible, the most serious drawback of the approach in [7] is that it cannot be applied to nonregular designs and nonorthogonal regular designs. The approach proposed in this paper overcomes such difficulties.

REMARK 5.2. Theorem 5.1, Theorem 5.2, and Corollary 5.1 are our main tools for eliminating inferior designs. While Corollary 5.1 is specifically for minimum aberration versions of the criteria under the effect hierarchy assumption that $v_k \gg v_{k+1}$ for all k , Theorem 5.1 and Theorem 5.2 can be used for all v_k values.

The sufficiency part of Theorem 5.1 can be proved in a similar fashion to that of Theorem 4.1 of [7].

PROOF OF THE SUFFICIENCY PART OF THEOREM 5.1. Let $\eta_i = \frac{1}{\xi_{\mathcal{F}_m}} - \frac{1}{\xi_{\mathcal{F}_i}}$ and $B_i(d) = \sum_{k=1}^n v_k B_{k,i}(d)$, $i = 0, \dots, m-1$. Then $\Phi_M(d; \boldsymbol{\xi}, \mathbf{v}) = \sum_{i=0}^{m-1} \eta_i B_i(d)$. We have to show that a design that minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i(d)$ for all subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ satisfying (5.4) also minimizes $\sum_{i=0}^{m-1} \eta_i B_i(d)$ for all feasible $\boldsymbol{\xi}$.

In the following a more general result will be proved: for any subset \mathfrak{B}' of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$, if d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i(d)$ for all subsets \mathfrak{G} of \mathfrak{B}' such that

$$(5.9) \quad \mathcal{F} \in \mathfrak{G}, \mathcal{F}' \in \mathfrak{B}' \text{ and } \mathcal{F} \prec \mathcal{F}' \Rightarrow \mathcal{F}' \in \mathfrak{G},$$

then d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{B}'} \eta_i B_i(d)$ for all feasible $\boldsymbol{\eta}$. Here $\boldsymbol{\eta} = (\eta_0, \dots, \eta_{m-1})^T$, and $\boldsymbol{\eta}$ is feasible if all the η_i 's are nonnegative and $\mathcal{F}_i \preceq \mathcal{F}_j \Rightarrow \eta_i \leq \eta_j$. If this is true, then the result in Theorem 5.1 follows by taking $\mathfrak{B}' = \mathfrak{B} \setminus \{\mathcal{F}_m\}$.

We prove the statement in the previous paragraph by mathematical induction on the number of unit factors in \mathfrak{B}' . It is clearly true when \mathfrak{B}' consists of one single unit factor. Now suppose that it holds for all subsets of $\mathfrak{B} \setminus \{\mathcal{F}_m\}$ with fewer than s unit factors, $s \geq 2$; we show that if $|\mathfrak{B}'| = s$ and d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i(d)$ for all subsets \mathfrak{G} of \mathfrak{B}' satisfying (5.9), then d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{B}'} \eta_i B_i(d)$ for all feasible $\boldsymbol{\eta}$. Under the given assumption, by taking $\mathfrak{G} = \mathfrak{B}'$, we have that

$$(5.10) \quad d^* \text{ minimizes } \sum_{i:\mathcal{F}_i \in \mathfrak{B}'} B_i(d).$$

If all the η_i 's for which $\mathcal{F}_i \in \mathfrak{B}'$ are equal, say they are all equal to η , then, by (5.10), d^* minimizes $\eta \sum_{i:\mathcal{F}_i \in \mathfrak{B}'} B_i(d) = \sum_{i:\mathcal{F}_i \in \mathfrak{B}'} \eta_i B_i(d)$.

On the other hand, suppose not all the η_i 's for which $\mathcal{F}_i \in \mathfrak{B}'$ are equal. Let η be the smallest value of such η_i 's and let $\mathfrak{B}^* = \{\mathcal{F}_i \in \mathfrak{B}' : \eta_i > \eta\}$. Then \mathfrak{B}^* is nonempty and

$|\mathfrak{B}^*| < s$. For each i such that $\mathcal{F}_i \in \mathfrak{B}^*$, let $\eta_i^* = \eta_i - \eta$; then $\eta_i^* > 0$. Also, $\eta_i^* \leq \eta_{i'}$ if $\mathcal{F}_i \prec \mathcal{F}_{i'}$ and $\mathcal{F}_i, \mathcal{F}_{i'} \in \mathfrak{B}^*$. Furthermore,

$$\sum_{i:\mathcal{F}_i \in \mathfrak{B}'} \eta_i B_i(d) = \sum_{i:\mathcal{F}_i \in \mathfrak{B}^*} \eta_i^* B_i(d) + \eta \sum_{i:\mathcal{F}_i \in \mathfrak{B}'} B_i(d).$$

By (5.10), it suffices to show that d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{B}^*} \eta_i^* B_i(d)$. Since $|\mathfrak{B}^*| < s$, by the induction hypothesis, it remains to show that d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i(d)$ for all subsets \mathfrak{G} of \mathfrak{B}^* satisfying the following condition:

$$(5.11) \quad \mathcal{F} \in \mathfrak{G}, \mathcal{F}' \in \mathfrak{B}^* \text{ and } \mathcal{F} \prec \mathcal{F}' \Rightarrow \mathcal{F}' \in \mathfrak{G}.$$

Suppose a subset \mathfrak{G} of \mathfrak{B}^* satisfies (5.11). By the assumption on \mathfrak{B}' , d^* minimizes $\sum_{i:\mathcal{F}_i \in \mathfrak{G}} B_i(d)$ provided that \mathfrak{G} also satisfies (5.9). That is, given $\mathcal{F} \in \mathfrak{G}$, $\mathcal{F}' \in \mathfrak{B}'$, and $\mathcal{F} \prec \mathcal{F}'$, we want to show $\mathcal{F}' \in \mathfrak{G}$. Because $\mathcal{F} \in \mathfrak{G} \subseteq \mathfrak{B}^*$, by the definition of \mathfrak{B}^* , we have $\eta_{\mathcal{F}} > \eta$. Also, $\eta_{\mathcal{F}'} \geq \eta_{\mathcal{F}}$ since $\mathcal{F} \prec \mathcal{F}'$. Thus we have $\eta_{\mathcal{F}'} \geq \eta_{\mathcal{F}} > \eta$, which leads to $\mathcal{F}' \in \mathfrak{B}^*$. Then by (5.11), $\mathcal{F}' \in \mathfrak{G}$. □

6. Examples. In this section, we consider the selection of Bayesian D-optimal designs under the prior specification (2.6) for some specific block structures. We first use Theorem 5.1 (or Corollary 5.1 under the assumption $v_1 \gg \dots \gg v_n$) to eliminate inadmissible designs, and then compute $\det(\text{cov}(\boldsymbol{\beta}|\mathbf{y}))$ for selected values of $\boldsymbol{\xi}, \sigma^2$, and r to compare the remaining designs. The value of σ^2 is arbitrarily chosen, while the values of r are chosen to be space-filling in the interval (0,1). Values of $\boldsymbol{\xi}$ are chosen for the cases of fixed unit effects, random unit effects, and unstructured units. In the following examples, the i th treatment factor is denoted by \mathbf{i} if $i \leq 9$ and \mathbf{t}_i if $i \geq 10$.

6.1. Blocking of regular fractional factorials. The block structure is $\mathfrak{B} = \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}$ with $\mathcal{F}_2 \prec \mathcal{F}_1 \prec \mathcal{F}_0$, where $\mathcal{F}_1 = \mathcal{B}$ is the block factor. We require that no treatment main effect be aliased with other treatment main effects nor confounded with blocks. Then $B_{1,0}(d) = B_{2,0}(d) = B_{1,1}(d) = 0$. As mentioned in Example 5.1, the subsets \mathfrak{G} of $\mathfrak{B} \setminus \{\mathcal{F}_2\}$ satisfying (5.4) are $\mathfrak{G}_1 = \{\mathcal{F}_0\}$ and $\mathfrak{G}_2 = \{\mathcal{F}_0, \mathcal{F}_1\}$, and that checking the sufficient condition in Theorem 5.1 amounts to examining the performances of the designs in experiments with unstructured units as well as a blocked experiment with fixed block effects. Suppose a 2^{6-2} design is to be run in 4 blocks of size 4. We apply Theorem 5.1 with $\sigma^2 = 15, 20$ and $r = 0.1, 0.3, 0.5, 0.7, 0.9$. In all these cases, complete searches show that the design d^* defined by $\mathbf{5} = \mathbf{134}, \mathbf{6} = \mathbf{123}, \mathcal{B}_1 = \mathbf{13}, \mathcal{B}_2 = \mathbf{124}$, where \mathcal{B}_1 and \mathcal{B}_2 are independent block defining words, is the only design (up to isomorphism) that minimizes both $\sum_{k=1}^n v_k B_{k,0}(d) + \sum_{k=1}^n v_k B_{k,1}(d)$ and $\sum_{k=1}^n v_k B_{k,0}(d)$. By Theorem 5.1, d^* is Bayesian (M.S)-optimal for all feasible $\boldsymbol{\xi}$ and the selected values of σ^2 and r .

The wordlength patterns (5.7) induced by \mathfrak{G}_1 and \mathfrak{G}_2 are $(B_{3,0}(d), \dots, B_{6,0}(d))$ and $(B_{2,0}(d) + B_{2,1}(d), \dots, B_{6,0}(d) + B_{6,1}(d))$. We found d^* as the only design (up to isomorphism) that sequentially minimizes both $B_{3,0}(d), \dots, B_{6,0}(d)$ and $B_{2,0}(d) + B_{2,1}(d), \dots, B_{6,0}(d) + B_{6,1}(d)$. Thus d^* also has minimum aberration with respect to the wordlength pattern (5.6) for all feasible ξ . Example 14.12 of [4, p. 316] identified the same design to be optimal with respect to the surrogate for the maximum information capacity criterion considered there.

Suppose a 2^{13-8} design is to be run in 8 blocks of size 4. Denote the independent block defining words by \mathcal{B}_1 , \mathcal{B}_2 , and \mathcal{B}_3 . Let d_1 be defined by $\mathbf{6} = \mathbf{12}, \mathbf{7} = \mathbf{13}, \mathbf{8} = \mathbf{14}, \mathbf{9} = \mathbf{234}, \mathbf{t}_{10} = \mathbf{1234}, \mathbf{t}_{11} = \mathbf{235}, \mathbf{t}_{12} = \mathbf{245}, \mathbf{t}_{13} = \mathbf{345}, \mathcal{B}_1 = \mathbf{23}, \mathcal{B}_2 = \mathbf{24}, \mathcal{B}_3 = \mathbf{15}$, d_2 be defined by $\mathbf{6} = \mathbf{123}, \mathbf{7} = \mathbf{124}, \mathbf{8} = \mathbf{134}, \mathbf{9} = \mathbf{234}, \mathbf{t}_{10} = \mathbf{125}, \mathbf{t}_{11} = \mathbf{135}, \mathbf{t}_{12} = \mathbf{235}, \mathbf{t}_{13} = \mathbf{145}, \mathcal{B}_1 = \mathbf{13}, \mathcal{B}_2 = \mathbf{14}, \mathcal{B}_3 = \mathbf{15}$, and d_3 be defined by $\mathbf{6} = \mathbf{12345}, \mathbf{7} = \mathbf{123}, \mathbf{8} = \mathbf{124}, \mathbf{9} = \mathbf{135}, \mathbf{t}_{10} = \mathbf{145}, \mathbf{t}_{11} = \mathbf{134}, \mathbf{t}_{12} = \mathbf{234}, \mathbf{t}_{13} = \mathbf{15}, \mathcal{B}_1 = \mathbf{12}, \mathcal{B}_2 = \mathbf{13}, \mathcal{B}_3 = \mathbf{45}$. At the first stage of screening out inferior designs, we apply Corollary 5.1 as well as Theorem 5.1 with $\sigma^2 = 15, 20$ and $r = 0.1, 0.3, 0.5, 0.7, 0.9$. In all these cases, each design other than d_1, d_2, d_3 has the same performance as or is inferior to at least one of them. However, unlike the example in the previous paragraph, since there are designs that have the same performance as d_1, d_2 , or d_3 in the first-stage screening but are not isomorphic to them, such designs should be retained in the second-stage screening, where $\det(\text{cov}(\boldsymbol{\beta}|\mathbf{y}))$, with $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0, 0.1, 0.2, 0.5, 1$ and $r = 0.1, 0.3, 0.5, 0.7, 0.9$, are compared. In all these cases, again d_1, d_2 , and d_3 are at least as good as any other design. For the MA surrogates,

$$\begin{aligned} (B_{3,0}(d_1), \dots, B_{13,0}(d_1)) &= (4, 39, 32, 48, 56, 39, 32, 0, 4, 1, 0), \\ (B_{3,0}(d_2), \dots, B_{13,0}(d_2)) &= (0, 55, 0, 96, 0, 87, 0, 16, 0, 1, 0), \\ (B_{3,0}(d_3), \dots, B_{13,0}(d_3)) &= (4, 38, 32, 52, 56, 33, 32, 4, 4, 0, 0), \\ \sum_{0 \leq i \leq 1} (B_{2,i}(d_1), \dots, B_{13,i}(d_1)) &= (22, 80, 163, 320, 452, 416, 311, 192, 70, 16, 5, 0), \\ \sum_{0 \leq i \leq 1} (B_{2,i}(d_2), \dots, B_{13,i}(d_2)) &= (36, 0, 365, 0, 848, 0, 651, 0, 140, 0, 7, 0), \\ \sum_{0 \leq i \leq 1} (B_{2,i}(d_3), \dots, B_{13,i}(d_3)) &= (30, 36, 255, 240, 452, 472, 255, 240, 30, 36, 1, 0). \end{aligned}$$

Since d_1 has the smallest $B_{2,0}(d) + B_{2,1}(d)$ among the three designs, it performs the best in the case $\xi_{\mathcal{F}_0} = \xi_{\mathcal{F}_1} = \infty$, and is expected to be optimal when $\xi_{\mathcal{F}_1}$ (interblock variance) is sufficiently greater than $\xi_{\mathcal{F}_2}$ (intra-block variance). On the other hand, since d_2 has the smallest $B_{3,0}(d)$, it performs the best when $\xi_{\mathcal{F}_1} = \xi_{\mathcal{F}_2}$, and is expected to be optimal when $\xi_{\mathcal{F}_1}$ is not too larger than $\xi_{\mathcal{F}_2}$. We have $B_{3,0}(d_3) = B_{3,0}(d_1), B_{4,0}(d_3) < B_{4,0}(d_1)$ and $B_{3,0}(d_3) > B_{3,0}(d_2)$; also, $B_{2,0}(d_2) + B_{2,1}(d_2) > B_{2,0}(d_3) + B_{2,1}(d_3) > B_{2,0}(d_1) + B_{2,1}(d_1)$. Thus d_3 is better than d_1 but worse than d_2 under $\xi_{\mathcal{F}_1} = \xi_{\mathcal{F}_2}$; it is better than d_2 but worse than d_1 under $\xi_{\mathcal{F}_0} = \xi_{\mathcal{F}_1} = \infty$. We expect d_3 to be optimal in cases where neither d_1

nor d_2 is optimal. These observations are confirmed by comparisons of the three designs we have made based on $\det(\text{cov}(\boldsymbol{\beta}|\mathbf{y}))$ with various values of $\boldsymbol{\xi}$, r , and σ^2 . Note that d_1 and d_2 were also identified in Example 14.13 of [4, p. 316] as “admissible” designs under a surrogate criterion for maximum information capacity. We point out the error that d_3 is also admissible but was left out in that example. The conclusions we draw on the comparison of d_1 and d_2 are the same as in Example 14.13 of [4].

6.2. Blocking of strip-plots. Example 14.17 of [4, p. 323] considered blocked strip-plot designs with six row treatment factors, four column treatment factors, and 32 experimental units in 2 blocks, where the 16 units in each block are arranged in 4 rows and 4 columns. The block structure is $\{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3, \mathcal{F}_4\}$, where $\mathcal{F}_0 = \mathcal{U}$, $\mathcal{F}_4 = \mathcal{E}$, \mathcal{F}_1 divides the 32 units into two blocks of size 16, \mathcal{F}_2 divides the 16 units in each block into 4 rows of size 4, and \mathcal{F}_3 divides the 16 units in each block into 4 columns of size 4, with $\mathcal{F}_4 \preceq \mathcal{F}_3, \mathcal{F}_2 \preceq \mathcal{F}_1 \preceq \mathcal{F}_0$. The main effects of row-treatment factors must be estimated in the between-row stratum within blocks and the main effects of column treatment factors must be estimated in the between-column stratum within blocks. The \mathfrak{G} 's satisfying (5.4) are $\{\mathcal{F}_0\}, \{\mathcal{F}_0, \mathcal{F}_1\}, \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2\}, \{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_3\}$, and $\{\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3\}$. Checking the conditions in Theorem 5.1 or Corollary 5.1 for these \mathfrak{G} 's amounts to examining performances of the designs in the following extreme scenarios, respectively: 32 unstructured experimental units, 2 blocks of size 16, 4 rows of size 4 nested in each of 2 blocks, 4 columns of size 4 nested in each of 2 blocks, and 4 rows and 4 columns nested in each of two blocks, where all the blocks, rows, and columns have fixed effects. The two admissible designs identified in Example 14.17 of [4, p. 323] are also the best according to the approach presented in this paper.

6.3. Blocking of nonregular fractional factorials. We present an application to nonregular and nonorthogonal blocked fractional factorial designs. The treatment defining word count A_k and block defining word count B_k for orthogonal regular designs can also be expressed, respectively, as in (2.10) and (2.11). By interpreting A_k and B_k as the corresponding expressions in (2.10) and (2.11), one can extend the wordlength patterns of orthogonal regular designs to nonorthogonal/nonregular designs. This was used by Cheng, Li, and Ye [8] to extend the two wordlength patterns $W_1(d)$ ([9]) and $W_2(d)$ (Chen and Cheng [3] and [9]) for blocked regular designs to $W_1^{G_2}(d)$ and $W_2^{G_2}(d)$, respectively, for blocked nonregular designs. Under orthogonal regular designs, the consequences of the treatment and block defining words on effect aliasing and confounding are quite clear; however, due to complex aliasing and nonorthogonality, it is difficult to interpret and justify $W_1^{G_2}(d)$ and $W_2^{G_2}(d)$. The following example shows how the approach presented here, which is based on statistically meaningful criteria, can be applied to nonorthogonal/nonregular designs.

Consider 16-run blocked factorial designs with four blocks of size four that are constructed by the method of replacement in [8] for five two-level treatment factors. Two optimal designs $7.21/2+3$ and $7.21/6+7$ based on $W_1^{G_2}$ and $W_2^{G_2}$, respectively, were ob-

tained in [8]. Design 7.21/2+3, denoted by d_1 hereafter, is obtained by projecting the 16-run Hall's design of type II listed in [8] onto factors $\mathbf{1}, \mathbf{4}, \mathbf{6}, \mathbf{8}, \mathbf{t}_{10}, \mathbf{t}_{12}, \mathbf{t}_{15}$, with $\mathbf{4}$ and $\mathbf{6}$ as the block generators. Design 7.21/6+7, denoted by d_2 hereafter, is obtained by projecting the same 16-run Hall's design onto factors $\mathbf{1}, \mathbf{4}, \mathbf{6}, \mathbf{8}, \mathbf{t}_{10}, \mathbf{t}_{12}, \mathbf{t}_{15}$, with \mathbf{t}_{12} and \mathbf{t}_{15} as the block generators. A complete search based on the results presented in this paper screens out all the designs except d_1 and d_2 . Design d_1 has $(B_{3,0}(d_1), B_{4,0}(d_1), B_{5,0}(d_1)) = (0, 0, 1)$ and $(B_{2,0}(d_1) + B_{2,1}(d_1), \dots, B_{5,0}(d_1) + B_{5,1}(d_1)) = (3, 3, 0, 1)$, and is optimal for the case $\xi_{\mathcal{F}_1} = \xi_{\mathcal{F}_2}$ ($\mathfrak{G} = \{\mathcal{F}_0\}$), i.e., the case of unstructured units (therefore as well as when the interblock variance is not too larger than the intrablock variance). In contrast, d_2 has $(B_{3,0}(d_2), B_{4,0}(d_2), B_{5,0}(d_2)) = (0, 1, 0)$ and $(B_{2,0}(d_2) + B_{2,1}(d_2), \dots, B_{5,0}(d_2) + B_{5,1}(d_2)) = (2, 4, 1, 0)$, and is optimal for the case $\xi_{\mathcal{F}_0} = \xi_{\mathcal{F}_1} = \infty$ ($\mathfrak{G} = \{\mathcal{F}_0, \mathcal{F}_1\}$), i.e., the case of fixed block effects (therefore as well as when the interblock variance is sufficiently greater than the intrablock variance). Define the D-efficiency of d_2 relative to d_1 as the ratio of $[\det(\text{cov}(\boldsymbol{\beta}|\mathbf{y}))]^{1/2^5}$ under d_1 over that under d_2 . Figure 6.1 displays such relative efficiencies for $\xi_{\mathcal{F}_0} = \infty, \xi_{\mathcal{F}_2} = 8, v_k = \tau^2 r^k, \tau^2 = \sigma^2(1+r)^{-5}$ with $\sigma^2 = 20, 0 < r < 1$, and (a) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0$, (b) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0.01$, (c) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0.1$, (d) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 1$. This figure and the results we have obtained for other values of σ^2, r and $\boldsymbol{\xi}$ confirmed the conclusion drawn above. In panel (d) we see that d_1 dominates d_2 completely (for all possible values of r) when the interblock and intrablock variances are equal. The figure shows clearly that as the ratio of interblock variance over intrablock variance becomes greater than 1, d_2 is better for some small values of r (the lower-order treatment factorial effects are much more important than the higher-order ones), and as this ratio increases, the range where d_2 is better gets larger. Eventually for infinite interblock variance (panel (a)), d_2 dominates d_1 completely. Hence we suggest that d_2 be used under fixed block effects.

7. Concluding remarks. In this paper, we use the Bayesian approach proposed in [16] to study the selection of multi-stratum factorial designs. Our criterion is to minimize the determinant (or trace) of the posterior covariance matrix. We derive a good surrogate criterion which can be related to the generalized wordlength patterns. This allows us to deal with nonorthogonal/nonregular designs, and provides a stronger justification for the use of generalized wordlength patterns than naive modifications of the usual wordlength patterns from regular to nonregular designs. We also provide a useful tool for screening out inferior designs to facilitate the search for optimal designs. Applications of such a tool are illustrated in several examples. Since the optimal designs in these examples are obtained through exhaustive computational searches, an efficient algorithm to handle highly fractionated designs with large run sizes would be desirable.

There are some possible directions to extend this work. One is to consider a more general prior distribution than (2.6). We point out that many results in Sections 3 and 5, including Theorems 3.1, 3.3, 5.1, 5.2, and Corollary 5.1, can be extended as long as $\boldsymbol{\Sigma}_\beta$ is a diagonal matrix, or equivalently, the underlying Gaussian process is stationary. Another extension is to consider multi-level treatment factors. However, this is not a trivial task since the

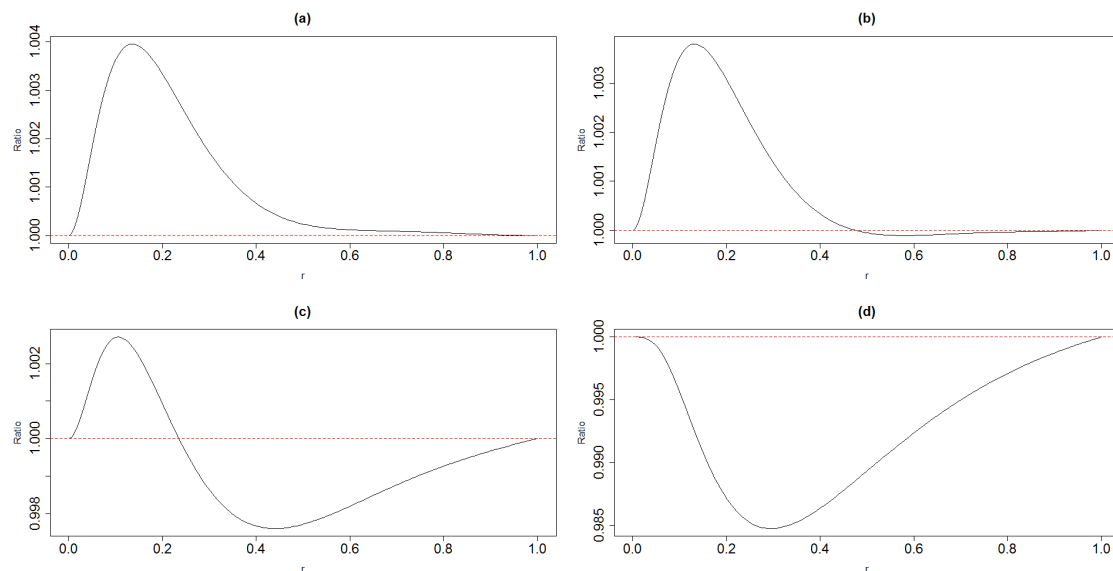


FIG 6.1. D -efficiencies of d_2 relative to d_1 for $0 < r < 1, \xi_{\mathcal{F}_0} = \infty, \xi_{\mathcal{F}_2} = 8$; (a) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0$; (b) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0.01$; (c) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 0.1$; (d) $\xi_{\mathcal{F}_2}/\xi_{\mathcal{F}_1} = 1$

prior distribution of treatment factorial effects induced by a Gaussian process is much more complicated than the two-level case, especially when the treatment factors are quantitative. This extension needs further investigations.

Acknowledgment. We would like to thank the referees for their helpful and constructive comments and suggestions, which have led to an improved paper.

REFERENCES

- [1] AI, M. AND KANG, L. AND JOSEPH, V. R. (2009). Bayesian optimal blocking of factorial designs. *J. Statist. Plann. Inference* **139** 3319-3328.
- [2] BAILEY, R. A. (2008). *Design of Comparative Experiments*. Cambridge Univ. Press, Cambridge.
- [3] CHEN, H. AND CHENG, C. S. (1999). Theory of optimal blocking of 2^{n-m} designs. *Ann. Statist.* **27** 1948-1973.
- [4] CHENG, C. S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments*. CRC Press, Boca Raton.
- [5] CHENG, C. S., STEINBERG, D. M. AND SUN, D. X. (1999). Minimum aberration and model robustness for two-level fractional factorial designs. *J. R. Statist. Soc. Ser. B* **61** 85-93.
- [6] CHENG, C. S. AND TSAI, P. W. (2009). Optimal two-level regular fractional factorial block and split-plot designs. *Biometrika* **96** 83-93.
- [7] CHENG, C. S. AND TSAI, P. W. (2011). Multistratum fractional factorial designs. *Statist. Sinica* **21** 1001-1021.
- [8] CHENG, S. W., LI, W. AND YE, K. Q. (2004). Blocked nonregular two-level factorial designs. *Technometrics* **46** 269-279.

- [9] CHENG, S. W. AND WU, C. F. J. (2002). Choice of optimal blocking schemes in two-level and three-level designs. *Technometrics* **44** 269-277.
- [10] ECCLESTON, J. A. AND HEDAYAT, A. (1974). On the theory of connected designs: characterization and optimality. *Ann. Statist.* **2** 1238-1255.
- [11] FRIES, A. AND HUNTER, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22** 601-608.
- [12] JOSEPH, V. R. (2006). A Bayesian approach to the design and analysis of fractionated experiments. *Technometrics* **48** 219-229.
- [13] JOSEPH, V. R., AI, M. AND WU, C.F.J. (2009). Bayesian-inspired minimum aberration two- and four-level designs. *Biometrika* **96** 95-106.
- [14] KANG, L. AND JOSEPH, V. R. (2009). Bayesian optimal single arrays for robust parameter design. *Technometrics* **51** 250-261.
- [15] KERR, M. K. (2001). Bayesian optimal fractional factorials. *Statist. Sinica* **11** 605-630.
- [16] MITCHELL, T. J., MORRIS, M. D. AND YLVIKAKER, D. (1995). Two-level fractional factorials and Bayesian prediction. *Statist. Sinica* **15** 559-573.
- [17] SACKS, J., WELCH, W. J., MITCHELL, T. J. AND WYNN, H. P. (1989). Design and analysis of computer experiments (with comments). *Statist. Sci.* **4** 409-435.
- [18] SUN, D. X. (1993). *Estimation Capacity and Related Topics in Experimental Designs*. Ph.D. Dissertation, University of Waterloo.
- [19] SUN, D. X., WU, C. F. J. AND CHEN, Y. (1997). Optimal blocking schemes for 2^n and 2^{n-p} designs. *Technometrics* **39** 298-307.
- [20] TANG, B. AND DENG, L. Y. (1999). Minimum G_2 -aberration for nonregular fractional factorial designs. *Ann. Statist.* **27** 1914-1926

128 ACADEMIA ROAD, SECTION 2, NANKANG,
TAIPEI 11529, TAIWAN
E-MAIL: mcchang@stat.sinica.edu.tw
cheng@stat.sinica.edu.tw