Optimal Two-Level Regular Fractional Factorial Block and Split-Plot Designs

BY CHING-SHUI CHENG

Department of Statistics, University of California, Berkeley, California 94720, U.S.A.

cheng@stat.berkeley.edu

AND PI-WEN TSAI

Department of Mathematics, National Taiwan Normal University, Taipei 11677, Taiwan

pwtsai@math.ntnu.edu.tw

SUMMARY

We propose a general and unified approach to the selection of regular fractional factorial designs which can be applied to experiments that are unblocked, blocked, with random or fixed block effects, or have a split-plot structure. Our criterion is derived as a good surrogate for the model-robustness criterion of information capacity. In the case of random block effects, it takes the ratio of intra- and interblock variances into account. In most of the cases, up to 32 runs, we have examined, there exist designs that are optimal for all values of that ratio. Examples of optimal designs that depend on the ratio are provided. We also demonstrate that our criterion can further discriminate designs that cannot be distinguished by the existing minimum-aberration criteria.

Some key words: Alias set; Estimation capacity; Information capacity; Minimum aberration; Model robustness; Wordlength pattern.

1. INTRODUCTION

There has been considerable recent research in the selection and construction of good fractional factorial designs under the minimum-aberration criterion proposed by Fries & Hunter (1980). Cheng et al. (1999) provided a statistical justification of this popular criterion and demonstrated that it is a good surrogate for robustness against model uncertainty.

When running an experiment, the experimenters are advised to randomize the experimental runs. Complete randomization is performed when the experimental units have no particular structure. On the other hand, a constrained randomization that respects the structure of the experimental units is desirable if such a structure exists. A good example is when the method of blocking is used to control variations among the experimental units. Proper randomization in this case leads to two error terms, or strata, in Nelder's (1965) language. Two different variances, called inter- and intrablock variances, respectively, are associated with the two strata. When blocking is successfully done, one expects the intrablock variance to be smaller than the interblock variance. Then a factorial effect, main effect or interaction contrast, that is confounded with a block contrast is estimated with a larger variance than one that is orthogonal to block contrasts. This complicates the issue of selecting a design. One would expect the decision to depend in some way on prior knowledge about the ratio of the two variances if it is available.

Extensions of the minimum-aberration criterion to blocked fractional factorial designs have been proposed by several authors (Sitter et al., 1997; Chen & Cheng, 1999; Cheng & Wu, 2002) under the assumption that the underlying model has fixed block effects. One objective of this article is to propose a general approach that covers the cases of random as well as fixed block effects. Following Cheng et al. (1999), we use a measure of model robustness to motivate a criterion for selecting blocked fractional factorial designs. It suffices to say at this point that for models with fixed block effects, one version of our criterion reduces to what was proposed by Chen & Cheng (1999), and it is the same as the usual minimum-aberration criterion for unblocked designs when the inter- and intrablock variances are equal.

When designing a blocked fractional factorial experiment, typically we try to avoid confounding main effects with block contrasts. However, for practical considerations, sometimes one must confound certain main effect contrasts with blocks. Such designs, called split-plot designs, arise, for example, when it is difficult or expensive to change the levels of certain factors. Then each of these factors, called whole-plot factors, keeps the same level in each block, called a whole-plot, while the other factors, called subplot factors, are allowed to take different levels in the units, called subplots, within the same whole-plot.

Perhaps because the research on minimum-aberration criteria for blocked fractional factorial designs has been focused on the case of fixed block effects, while for split-plot designs one must assume that the whole-plot effects are random, these two lines of research have been rather disconnected, even though split-plot designs are a special kind of blocked factorial designs. Huang et al. (1998) and Bingham & Sitter (1999a, b) applied the usual minimum-aberration criterion, for unblocked factorials, to rank fractional factorial split-plot designs. This approach ignores the effect of the block structure on the precision of the estimators of factorial effects. In many situations the usual minimum-aberration criterion cannot distinguish several different split-plot designs even though some are clearly better than the others. Bingham & Sitter (2001) and Mukerjee & Fang (2002) used additional secondary criteria further to discriminate such designs. The second objective of this article is to apply the same approach we propose for blocked fractional factorial designs to split-plot designs. Our unified approach can also be applied to more complicated experiments with more than two strata, including row-column designs, split-split-plot designs, blocked split-plot designs (McLeod & Brewster, 2004), strip-plot designs (Miller, 1997) and more generally those with the block structures studied by Nelder (1965). These will be treated elsewhere.

Only two-level regular fractional factorial designs with orthogonal blocking are considered in this article; the case in which the number of levels is a prime power can be treated similarly. We make the hierarchical assumption that lower-order effects are more important than higher-order effects and effects of the same order are equally important, a situation in which minimum aberration is a suitable criterion. For simplicity, we further assume that three-factor and higher-order interactions are negligible.

2. RANDOMIZED BLOCK DESIGNS AND EXISTING CRITERIA

A consequence of randomization for a block design is that the vector of observations y can be assumed to have a simple covariance matrix V in which all the observations have the same variance and there are two possibly different covariances depending on whether or not two observations are in the same block (Bailey, 1981). Let b be the number of blocks and let s be the block size. Then V has three eigenvalues, ξ_0, ξ_1 and ξ_2, where ξ_0 is the common row sum of V, and under the assumption of no treatment effects, i.e., $E(y)$ is constant, $\xi_1 = E\{(b-1)^{-1}\sum_{i=1}^{b} s(y_{i.} - y_{..})^2\}$, and $\xi_2 = E[\{b(s-1)\}^{-1}\sum_{i=1}^{b}\sum_{j=1}^{s}(y_{ij} - y_{i.})^2]$, where y_{ij} is the observation on the jth unit in the ith block, $y_{i.} = \sum_{j=1}^{s} y_{ij}/s$, and $y_{..} = \sum_{i=1}^{b}\sum_{j=1}^{s} y_{ij}/bs$. It is clear from these equations that ξ_1 and ξ_2 are measures of between-block and within-block variability, and are called inter- and intrablock variance, respectively. Consider contrasts $l'y$ of the observations such that all the observations with the same factor-level combination have the same coefficient. We note that $\text{var}(l'y) = \xi_1 l'l$ if l falls in the eigenspace associated with ξ_1, in which case we say that the treatment contrast $E(l'y)$ is confounded with blocks; on the other hand, $\text{var}(l'y) = \xi_2 l'l$ if l falls in the eigenspace associated with ξ_2, in which case $E(l'y)$ is orthogonal to the blocks. We denote the ratio ξ_2/ξ_1 by r.

We shall assume that $r \leq 1$, i.e., $\xi_1 \geq \xi_2$. The inequality $\xi_1 > \xi_2$ holds whenever two different observations from the same block have a larger covariance than those from different blocks. One example is when the error term ε_{ij} associated with the observation on the jth unit in the ith block can be decomposed as $\varepsilon_{ij} = \beta_i + e_{ij}$, where $\{\beta_i\}$ and $\{e_{ij}\}$ are mutually uncorrelated random variables with $E(\beta_i) = E(e_{ij}) = 0$, and $\text{var}(\beta_i) = \sigma_B^2$, $\text{var}(e_{ij}) = \sigma_e^2$; that is, we have a model with random block effects. In this case, the covariances are σ_B^2 or 0 for two observations in the same or different blocks, respectively. Thus the condition $\sigma_B^2 > \sigma_e^2$, as is often assumed in the literature, is not necessary.

Note that r can be thought of as the efficiency of the estimator of a contrast confounded with blocks relative to that of one orthogonal to blocks. Under models with fixed-block effects, such efficiency is zero since contrasts confounded with blocks are not estimable. Therefore, models with fixed block effects can be thought to have $r = 0$. On the other hand, $r = 1$ corresponds to when the inter- and intrablock variances are equal, i.e., blocking does not affect precision of the estimators.

A regular 2^{n-p} fractional factorial design with n two-level factors in 2^{n-p} runs is determined by p independent treatment defining effects. Under such a design, the defining contrast subgroup generated by the p independent effects contains $2^p - 1$ treatment factorial effects, called defining effects, defining contrasts or defining words. Let $A_i(d)$ be the number of defining effects of length i of a design d. The sequence $\{A_1(d), A_2(d), \ldots\}$ is called the wordlength pattern of d. A design is said to have minimum aberration if it sequentially minimizes $A_1(d), A_2(d), A_3(d), \ldots$, and so on.

To divide the 2^{n-p} treatment combinations into 2^q blocks each of size 2^{n-p-q}, where $q < n-p$, one needs another set of q independent defining effects, called block defining effects, which are independent of the treatment defining effects. It is convenient to think of such a blocked fractional factorial design as a $2^{n+q-(p+q)}$ design, where the $n+q$ factors are the n treatment factors and q block factors defined by the q additional independent block defining contrasts. However, the wordlength pattern of this $2^{n+q-(p+q)}$ design as defined in the previous paragraph is not enough to summarize the properties of a block design. We define $A_{i,j}(d)$ as the number of defining words in the $2^{n+q-(p+q)}$ design that contain i treatment factors and j block factors, and let $B_i(d) = \sum_{j:j\geq 1} A_{i,j}(d)$. All the existing minimum-aberration criteria for blocked fractional factorial designs are based on the two sequences $\{A_{i,0}(d)\}$ and $\{B_i(d)\}$; see Sitter et al. (1997), Chen & Cheng (1999) and Cheng & Wu (2002). Two criteria relevant to our study, the W_{CC}-criterion proposed by Chen & Cheng (1999) and the W_1-criterion proposed by Cheng & Wu (2002), are based on the sequential minimization of components of the following wordlength patterns:

$$W_{\text{CC}} = (3A_{3,0} + B_2, A_{4,0}, 10A_{5,0} + B_3, A_{6,0}, \ldots),$$

$$W_1 = (A_{3,0}, A_{4,0}, B_2, A_{5,0}, A_{6,0}, B_3, \ldots).$$

Note that there are two assumptions underlying these criteria: the block effects are assumed to be fixed, and $A_{1,0} = A_{2,0} = B_1 = 0$. When three-factor and higher order interactions are negligible, only the first three terms in W_1 and the first two terms in W_{CC} are relevant. In particular, we call a design W_1-optimal if it sequentially minimizes $A_{3,0}$, $A_{4,0}$ and B_2, and W_{CC}-optimal if it sequentially minimizes $3A_{3,0} + B_2$ and $A_{4,0}$.

3. New criteria for blocked fractional factorial designs

The minimum-aberration criterion based on W_{CC} was proposed as a surrogate for the model robustness criterion of maximum estimation capacity (Cheng et al., 1999). For any $1 \leq k \leq \frac{1}{2}n(n-1)$, define $E_k(d)$ as the number of models containing all the main effects and k two-factor interactions that can be estimated by a design d. Here k can be thought of as the number of active two-factor interactions. Following Cheng et al. (1999), Chen & Cheng (1999) showed that the E_k's are large if the number of two-factor interactions that are neither aliased with main effects nor confounded with blocks, i.e., $\frac{1}{2}n(n-1) - (3A_{3,0} + B_2)$ is maximized and that these interactions are distributed very uniformly over the alias sets. Once the former is maximized, i.e., $3A_{3,0} + B_2$ is minimized, the latter can be achieved by making $A_{4,0}$ small. This leads to the first two components of W_{CC}.

Under a regular fractional factorial design with orthogonal blocking, a model with fixed block effects is either not estimable, and thus has a singular information matrix with zero determinant or all its parameters can be estimated with the highest efficiency. Let δ_k be the number of candidate models containing all the main effects and k two-factor interactions. Therefore $E_k(d)/\delta_k$ can be regarded as the average efficiency of d over the candidate models. On the other hand, when the block effects are random, or under a randomization model, different estimable models may be estimated with different efficiencies depending on how many factorial effects in the model are confounded with or are orthogonal to the blocks. We use the average efficiency over all the candidate models, called the information capacity by D. X. Sun, in his unpublished 1993 Ph.D thesis from The University of Waterloo, and Li & Nachtsheim (2000), to measure model robustness of a design. The use of the average efficiency over a set of candidate models to select model-robust factorial designs has also been proposed by Tsai et al. (2000) in a different context.

We assume that no main effect is either a defining effect or confounded with blocks, and that no two main effects are aliased with each other. Then they are estimated with the same precision. Thus we shall concentrate on the efficiencies of the estimators of two-factor interactions, measured by $D^{1/k}$, where D is the determinant of the information matrix for the two-factor interactions, and k is the number of two-factor interactions in the model. For simplicity we normalize the interaction contrasts so that their estimators are of the form $l'y$ with $l'l = 1$. For a given design d, we define the information capacity $I_k(d)$ as the average of $D^{1/k}$ over the set of models containing all the main effects and k two-factor interactions. We seek a design that has large I_k. Clearly, under models with fixed block effects, maximizing $I_k(d)$ is equivalent to maximizing $E_k(d)$ as considered in Chen & Cheng (1999).

The $2^n - 2^p$ treatment factorial effects other than those in the defining contrast subgroup are divided into $g \equiv 2^{n-p} - 1$ alias sets each of size 2^p. The effects in $2^q - 1$ of these alias sets are confounded with blocks. Let $f = g - n$. Without loss of generality, suppose that the effects in each of the first h alias sets are confounded with blocks but are not aliased with main effects, that those in the next $f - h$ alias sets are neither confounded with blocks nor aliased with main effects, and that each of the last n alias sets contains one main effect. When no main effect is confounded with blocks, as assumed in this section, we have $h = 2^q - 1$. For each $i = 1, \ldots, g$, let m_i be the number of two-factor interactions contained in the ith alias set. In the Appendix, we show that I_k is large if

$$\begin{array}{ll} (i) & r^{1/k}\sum_{i=1}^{h} m_i + \sum_{i=h+1}^{f} m_i \text{ is as large as possible, and} \\ (ii) & r^{1/k}\, m_1, \ldots, r^{1/k} m_h, m_{h+1}, \ldots, m_f \text{ are as equal as possible.} \end{array} \qquad (1)$$

Let $u = \frac{1}{2}n(n-1) - \sum_{i=h+1}^{f} m_i$ and $v = \sum_{i=1}^{h} m_i$. Then u is the number of two-factor interactions that are aliased with main effects or are confounded with blocks, and v is the number of two-factor interactions that are not aliased with main effects but are confounded with blocks. To make $r^{1/k}\sum_{i=1}^{h} m_i + \sum_{i=h+1}^{f} m_i$ large is the same as to make $u - r^{1/k}v$ small. On the other hand, a good surrogate for (ii) is to have small $r^{2/k}\sum_{i=1}^{h} m_i^2 + \sum_{i=h+1}^{f} m_i^2$.

Under the assumption that no main effect is confounded with blocks, we have $u = 3A_{3,0} + B_2$ and $v = B_2$; therefore, $u - r^{1/k}v = 3A_{3,0} + (1 - r^{1/k})B_2$. Since $A_{4,0} = \frac{1}{6}\{\sum_{i=1}^{g} m_i^2 - n(n-1)/2\}$, see (2.2) of Cheng, et al. (1999), $\sum_{i=1}^{g} m_i^2$ is small if $A_{4,0}$ is small. Then $r^{2/k}\sum_{i=1}^{h} m_i^2 + \sum_{i=h+1}^{f} m_i^2$ would also be small. Thus a good and simple surrogate for maximizing I_k is to minimize $3A_{3,0} + (1 - r^{1/k})B_2$, followed by the minimization of $A_{4,0}$. We call this two-stage procedure, which replaces B_2 in W_{CC} with $(1 - r^{1/k})B_2$, the W_k^r-criterion. The weight $1 - r^{1/k}$ is imposed to penalize confounding with blocks. The heaviest penalty is imposed when the block effects are fixed, $r = 0$, since in this case confounding with blocks leads to total loss of information. Indeed W_k^0-optimality is the same as W_{CC}-optimality. The penalty decreases as r increases. For $r = 1$, W_k^1 successively minimizes $A_{3,0}$ and $A_{4,0}$, as in the usual minimum-aberration criterion for unblocked factorial designs. We shall denote this criterion by W_{MA}, partly to emphasize that it does not depend on k. The following result provides a useful tool for studying W_k^r-optimal designs.

THEOREM 1. *If a design is both W_k^0- and W_k^1-optimal, i.e., it is W_{CC}- and W_{MA}-optimal, then it is W_k^r-optimal for all k and all $0 \le r \le 1$. The same conclusion also holds for a design that is both W_{CC}- and W_1- optimal.*

Proof. Suppose d is both W_k^0- and W_k^1-optimal. For any other design d', we need to show that $3A_{3,0}(d) + (1 - r^{1/k})B_2(d) \le 3A_{3,0}(d') + (1 - r^{1/k})B_2(d')$, and, if $3A_{3,0}(d) + (1 - r^{1/k})B_2(d) = 3A_{3,0}(d') + (1 - r^{1/k})B_2(d')$, then $A_{4,0}(d) \le A_{4,0}(d')$.

We have $3A_{3,0}(d') + (1 - r^{1/k})B_2(d') - \{3A_{3,0}(d) + (1 - r^{1/k})B_2(d)\} = (1 - r^{1/k})\,[\{3A_{3,0}(d') + B_2(d')\} - \{3A_{3,0}(d) + B_2(d)\}] + 3r^{1/k}\{A_{3,0}(d') - A_{3,0}(d)\}$, which is nonnegative since both $\{3A_{3,0}(d') + B_2(d')\} - \{3A_{3,0}(d) + B_2(d)\}$ and $A_{3,0}(d') - A_{3,0}(d)$ are nonnegative. Now suppose the above difference is zero. If $0 \le r < 1$, then $3A_{3,0}(d') + B_2(d')$ must be equal to $3A_{3,0}(d) + B_2(d)$, and, for $r = 1$, we have $A_{3,0}(d') = A_{3,0}(d)$. In both cases it follows from the W_k^0- and W_k^1-optimality of d that $A_{4,0}(d) \le A_{4,0}(d')$. □

Minimizing $A_{4,0}$ is a good, albeit not the best, surrogate for part (ii) of (1), since the role played by r is ignored. A more accurate surrogate for maximizing I_k is to maximize $r^{1/k} \sum_{i=1}^{h} m_i + \sum_{i=h+1}^{f} m_i$, followed by the minimization of $r^{2/k} \sum_{i=1}^{h} m_i^2 + \sum_{i=h+1}^{f} m_i^2$. We call this the $\tilde{W}_k^r$-criterion. We note that $r^{2/k} \sum_{i=1}^{h} m_i^2 + \sum_{i=h+1}^{f} m_i^2$ can be expressed in terms of some modified and more complicated versions of wordlengths; the details are omitted since the m_i values can be determined in a straightforward manner. One can use $\tilde{W}_k^r$ as a primary criterion or as a secondary criterion when W_k^r cannot produce a unique optimal design. The advantage of W_k^r is its simplicity, and, as discussed earlier, it is nicely related to the existing criteria. Our experience indicates that W_k^r works well at least for up to 64-run designs.

The following result similar to Theorem 1 can easily be established.

THEOREM 2. *If a design is both $\tilde{W}_k^0$- and $\tilde{W}_k^1$-optimal, then it is $\tilde{W}_k^r$-optimal for all k and all* $0 \le r \le 1$.

We note that both $\tilde{W}_k^0$ and $\tilde{W}_k^1$ do not depend on k. By Theorem 2, to show that a design is $\tilde{W}_k^r$-optimal for all k and all $0 \le r \le 1$, it suffices to verify that (i) it maximizes $\sum_{i=1}^{f} m_i$, and minimizes $\sum_{i=1}^{f} m_i^2$ among those which maximize $\sum_{i=1}^{f} m_i$, and (ii) it maximizes $\sum_{i=h+1}^{f} m_i$, and minimizes $\sum_{i=h+1}^{f} m_i^2$ among those which maximize $\sum_{i=h+1}^{f} m_i$. Requirement (i) ensures that the design has the largest number of two-factor interactions that are not aliased with main effects and that these two-factor interactions are very uniformly distributed over the alias sets.

When there are two strata, (ii) further requires that a large number of the two-factor interactions that are not aliased with main effects be estimated in the more precise intrablock stratum, and that these two-factor interactions be also very uniformly distributed over the alias sets.

The practical significance of Theorems 1 and 2 is that one can establish a strong optimality result that holds for all k and r by checking two simple cases that do not involve k. Furthermore, one can use W_{CC} and W_1, or W_{CC} and W_{MA}, or $\tilde{W}_k^0$ and $\tilde{W}_k^1$, together to screen out inferior designs. If a design d_1 is at least as good as another design d_2 under both criteria, and is better than d_2 under at least one of the two criteria, then we say that d_1 dominates d_2 with respect to the two criteria. In this case d_2 can be dropped from consideration, and we say that it is inadmissible. Typically there may be only a small number of admissible designs. Then we can carry out the simple task of comparing them directly for given k and r. When comparing the admissible designs, we suggest using the $I_k(d)$-criterion directly. This is because the range of r values for a design to be better than another under $I_k(d)$ may not coincide with that determined by simple surrogates such as W_k^r.

Example 1. Consider the case $n = 13, p = 8$ and $q = 3$, i.e., 32-run designs for 13 factors with the 32 runs grouped into 8 blocks of size 4 each. Let d_1, d_2 and d_3 be three designs with the following sets of independent defining words:

$$d_1 : \{1236, 1247, 1348, 2349, 125t_{10}, 135t_{11}, 235t_{12}, 145t_{13}, 12b_1, 13b_2, 14b_3\},$$

$$d_2 : \{1236, 1247, 1348, 2349, 125t_{10}, 135t_{11}, 235t_{12}, 145t_{13}, 13b_1, 14b_2, 15b_3\},$$

$$d_3 : \{126, 137, 148, 2349, 1234t_{10}, 235t_{11}, 245t_{12}, 345t_{13}, 23b_1, 24b_2, 15b_3\},$$

where $t_{10}, \ldots, t_{13}$ are treatment factors $10, \ldots, 13$, and b_1, b_2, b_3 are block factors. Then $A_{3,0}(d_1) = 0$, $A_{4,0}(d_1) = 55$, $B_2(d_1) = 38$; $A_{3,0}(d_2) = 0$, $A_{4,0}(d_2) = 55$, $B_2(d_2) = 36$; $A_{3,0}(d_3) = 4$, $A_{4,0}(d_3) = 39$, $B_2(d_3) = 22$.

In view of these wordlengths, d_1 is worse than d_2 with respect to W_{CC} and the two designs are tied under W_{MA}. We note that d_1 is also worse than d_2 with respect to $\tilde{W}_k^0$ and the two designs are tied under $\tilde{W}_k^1$. By the proofs of Theorems 1 and 2, d_1 is worse than d_2 with respect to W_k^r and $\tilde{W}_k^r$ for all k and all $0 \leq r < 1$, and they are tied for $r = 1$. These two designs do have identical information capacities when $r = 1$, and, for $r < 1$, $I_k(d_1) < I_k(d_2)$ for all the k and r values we have examined. On the other hand, d_2 is better than d_3 with respect to W_{MA} and is worse than d_3 with respect to W_{CC}. Therefore d_2 is better than d_3 with respect to W_k^r except when r is small. Likewise, d_2 is better than d_3 with respect to $\tilde{W}_k^1$ and is worse than d_3 with respect to $\tilde{W}_k^0$. A complete search of all the designs with $n = 13, p = 8, q = 3$ shows that d_2 and d_3 are the only two admissible designs with respect to W_{MA} and W_{CC}. They are also the only admissible designs with respect to W_1 and W_{CC} as well as $\tilde{W}_k^0$ and $\tilde{W}_k^1$. It follows from Theorems 1 and 2 that d_3 is W_k^r- and $\tilde{W}_k^r$-optimal for small r values and d_2 is W_k^r- and $\tilde{W}_k^r$-optimal for larger r values. A comparison of the information capacities of these two designs shows that $I_k(d_2) > I_k(d_3)$ except when r is small.

Xu (2006) and Xu & Lau (2006) constructed all minimum-aberration blocked two-level fractional factorial designs with 8, 16, 32 runs and those with 64 runs and up to 32 factors with respect to several criteria. They found that in most cases these criteria lead to the same minimum-aberration design. In particular, for 16-run designs, except for $(n, p, q) = (5, 1, 1)$ and $(5, 1, 2)$, W_1 and W_{CC} produce the same minimum-aberration designs. These designs are W_k^r- and $\tilde{W}_k^r$-optimal for all k and all $0 \leq r \leq 1$. For both of the two cases $(n, p, q) = (5, 1, 1)$ and $(5, 1, 2)$, the W_{CC}-optimal design is W_k^r- and $\tilde{W}_k^r$-optimal for small r values, and the W_1-optimal design is W_k^r- and $\tilde{W}_k^r$-optimal for larger r values, and these two designs are the only admissible designs with respect to W_1 and W_{CC} as well as $\tilde{W}_k^0$ and $\tilde{W}_k^1$. For 32-runs designs, W_1 and W_{CC} produce the same minimum-aberration designs except when $(n, p, q) = (6, 1, 2)$, $(7, 2, 2)$, $(8, 3, 3)$,

$(9,4,2)$, $(9,4,3)$, $(10,5,1)$, $(10,5,2)$, $(10,5,3)$, $(11,6,3)$, $(12,7,3)$, $(13,8,3)$, $(21,16,1)$ and $(21,16,2)$. These are also the only cases of 32-run designs where there are more than one admissible design with respect to W_1 and W_{CC} as well as $\tilde{W}_k^0$ and $\tilde{W}_k^1$. In all the twelve cases except $(n,p,q)=(9,4,2)$, the W_1- and W_{CC}-optimal designs are the only admissible designs. For $(n,p,q)=(9,4,2)$, there is a third admissible design with independent defining effects 12456, 2359, 2578, 2689, $14b_1$ and $15b_2$.

The numerical computations we have done demonstrate that the criteria introduced in this article are good surrogates for maximum information capacity, but we should keep in mind that they are only surrogates. We have examined all 16-run designs to study the consistency of rankings under $\tilde{W}_k^0$, $\tilde{W}_k^1$, W_1 and W_{CC}. Treating those which perform equally well under all the four criteria as the same design, we examined 388 pairs of 16-run designs with the same numbers of treatment factors and blocks. They fall into the following four categories: (i) the two designs perform equally well under both $\tilde{W}_k^0$ and $\tilde{W}_k^1$, but one of them dominates the other with respect to W_1 and W_{CC}, two pairs; (ii) one design dominates the other with respect to $\tilde{W}_k^0$ and $\tilde{W}_k^1$, but neither dominates the other with respect to W_1 and W_{CC}, three pairs; (iii) one design dominates the other with respect to $\tilde{W}_k^0$ and $\tilde{W}_k^1$ as well as W_1 and W_{CC}, 374 pairs; (iv) neither design dominates the other with respect to $\tilde{W}_k^0$ and $\tilde{W}_k^1$ or W_1 and W_{CC}, nine pairs. In category (i), the two designs have identical information capacities for all r and k, and in category (ii), the better design under $\tilde{W}_k^0$ and $\tilde{W}_k^1$ dominates the other with respect to the information capacity criterion for all the k and r values that we have examined. These results confirm our expectation that $\tilde{W}_k^0$ and $\tilde{W}_k^1$ give more accurate comparisons than W_1 and W_{CC}. In 336 of the 374 pairs in category (iii), the better design with respect to $\tilde{W}_k^0$ and $\tilde{W}_k^1$, or W_1 and W_{CC}, dominates the other under the information capacity criterion for all the k and r values that we have examined. In the 38 exceptional pairs, the better design under $\tilde{W}_k^0$ and $\tilde{W}_k^1$ fails to have larger $I_k(d)$ only for larger

k's. When k is large, the information capacity criterion may favour designs with smaller A_4. This is because smaller A_4 implies fewer pairs of aliased two-factor interactions, and thus possibly more estimable models containing a larger number of two-factor interactions. In category (iv), neither design dominates the other with respect to the information capacity criterion for all r and k.

4. SPLIT-PLOT DESIGNS

The principle we used to derive W_k^r and $\tilde{W}_k^r$ is applicable to split-plot designs since they are a special kind of blocked fractional factorial design. Again we assume that the two-factor interactions are equally important. In this case the split-plot structure arises purely from some practical needs; for example, it is difficult to change the levels of certain factors or some factors require larger experimental units. This is different from the situation in which there are what Cox (1958, p.??) called classification factors, which are included in the experiment primarily to examine their possible interactions with other factors, such as in robust product experiments. A different approach needs to be developed for such experiments and will not be treated here.

Suppose that there are n_1 whole-plot factors and n_2 subplot factors, each at two levels, and that there are $2^{n_1-p_1}$ whole plots, each containing $2^{n_2-p_2}$ subplots. It is convenient to consider the design as a 2^{n-p} design with $n = n_1 + n_2$ and $p = p_1 + p_2$. We use capital letters $A, B, \ldots$ to denote whole-plot factors and use lower case letters $p, q, \ldots$ to denote subplot factors. We also assume that $p_1 \geq 0$, but our approach can also be applied to the case where the whole-plot factors are replicated, as discussed in Bingham et al. (2004).

Under a split-plot design, main effects of the whole-plot factors are estimated with the same precision, and main effects of the subplot factors are also estimated with the same precision. It follows that all the designs have the same overall efficiency for the estimators of the main effects of treatment factors. We can therefore concentrate on the efficiencies of the estimators of two-factor interactions, and define the information capacity I_k and derive its surrogates in the same way as in §3 and the Appendix. As in the paragraph following (1), the first step of a good surrogate for maximizing I_k is to minimize $u - r^{1/k}v$, where in the present context u is the number of two-factor interactions that are aliased with main effects or are estimated in the whole-plot stratum, and v is the number of two-factor interactions that are not aliased with main effects and are estimated in the whole-plot stratum. Since two-factor interactions that are estimated in the whole-plot stratum may be aliased with main effects, the computation of u and v cannot be done as in §3, and a correct form of the W_k^r-criterion can no longer be expressed purely in terms of the wordlengths. On the other hand, the $\tilde{W}_k^r$-criterion can be carried over without any change, and is a more accurate surrogate than W_k^r for the information capacity criterion. In fact, there are seven cases of 32-run split-plot designs in which W_k^r-optimal designs are not unique but can be discriminated by $\tilde{W}_k^r$. To save space, in this section we shall concentrate on $\tilde{W}_k^r$.

As in §3, the $\tilde{W}_k^r$-criterion first maximizes $r^{1/k}\sum_{i=1}^h m_i + \sum_{i=h+1}^f m_i$, and then minimizes $r^{2/k}\sum_{i=1}^h m_i^2 + \sum_{i=h+1}^f m_i^2$. Here the first sum $\sum_{i=1}^h$ is over the alias sets of effects which are not aliased with main effects and are estimated in the whole-plot stratum, and the second sum $\sum_{i=h+1}^f$ is over those which are not aliased with main effects and are estimated in the subplot stratum. Furthermore, Theorem 2 still holds. Therefore, as discussed in the paragraph preceding Example 1, we can use $\tilde{W}_k^1$ and $\tilde{W}_k^0$ to screen out inadmissible designs.

Example 2. Consider the case $n_1 = 5$, $n_2 = 2$, $p_1 = p_2 = 1$. Table 4 of Bingham & Sitter (2001) listed two designs which are the best and are equally good according to their approach:

$$d_1 : I = ABCDE = ABpq = CDEpq,$$

$$d_2 : I = ABCE = ABDpq = CDEpq.$$

We point out that these two designs also cannot be distinguished by the proper form of W_k^r for split-plot designs. As to $\tilde{W}_k^r$, under both d_1 and d_2, there are 10 alias sets of effects which are estimated in the whole-plot stratum and are not aliased with main effects; the numbers of two-factor interactions in these alias sets are nine 1's and one 2 for d_1, and three 2's, five 1's and two 0's for d_2. For the 14 alias sets of effects that are estimated in the subplot stratum and are not aliased with main effects, we have two 2's, six 1's and six 0's for d_1, and ten 1's and four 0's for d_2. The values of $\sum_{i=1}^{f} m_i$, $\sum_{i=h+1}^{f} m_i$, $\sum_{i=1}^{f} m_i^2$ and $\sum_{i=h+1}^{f} m_i^2$ for d_1 are 21, 10, 27 and 14, respectively, and those for d_2 are 21, 10, 27 and 10. It follows from Theorem 2 that d_2 is $\tilde{W}_k^r$-better than d_1 for all k and all $0 \leq r < 1$, and ties with d_1 for $r = 1$. Essentially d_2 beats d_1 because it has a more uniform distribution, over the alias sets, of the two-factor interactions that are not aliased with main effects and are estimated in the more precise subplot stratum. In fact, it can be verified that, for $n_1 = 5$, $n_2 = 2$ and $p_1 = p_2 = 1$, d_2 is $\tilde{W}_k^r$-optimal for all k and all $0 \leq r \leq 1$. One can also see that d_2 is superior in the sense that all the aliasing between two-factor interactions occurs in the less precise whole-plot stratum. Comparison of the information capacities of d_1 and d_2 confirms that d_2 is indeed a better design.

Bingham & Sitter (2001) used the usual minimum-aberration criterion for unblocked designs to compare split-plot designs, but since it often leads to more than one nonisomorphic minimum-aberration design, for further discrimination, they proposed the secondary criterion of minimizing the number of subplot two-factor interactions that are estimated in the whole-plot stratum. This approach leads to a unique optimal design in each of the 16-run and all but 12 of the 32-run cases they had studied. The results are tabulated in Table 3, for 16-run designs, and Table 4, for 32-run designs, of their paper.

Except for one 16-run case, with $(n_1, n_2, p_1, p_2) = (4, 2, 1, 1)$, and two 32-run cases, with $(n_1, n_2, p_1, p_2) = (7, 3, 3, 2)$ and $(8, 2, 4, 1)$, the best design or one of the best designs Bingham & Sitter (2001) found is $\tilde{W}_k^r$-optimal for all k and all $0 \le r \le 1$. In each of the exceptional cases, our approach found another admissible design, with respect to $\tilde{W}_k^1$ and $\tilde{W}_k^0$, which has larger I_k when r is small or k is large. These additional admissible designs have the following independent defining effects: $ABD, ACpq$ for $(n_1, n_2, p_1, p_2) = (4, 2, 1, 1)$; $ABE, ACF, BCDG, BCpq, ADpr$ for $(n_1, n_2, p_1, p_2) = (7, 3, 3, 2)$; and $ABE, ACF, ADG, BCDH, ABCDpq$ for $(n_1, n_2, p_1, p_2) = (8, 2, 4, 1)$.

Implicitly, our criteria also take into account the number of subplot two-factor interactions that are estimated in the whole-plot stratum, but those which are aliased with main effects are excluded. This is because, once a two-factor interaction is aliased with main effects, according to the hierarchical principle, it cannot be estimated in a model in which all the main effects need to be estimated. Consequently such a two-factor interaction does not play a role in the computation of information capacities. For example, for $(n_1, n_2, p_1, p_2) = (1, 8, 0, 5)$, the first design listed in Bingham & Sitter's (2001) Table 3 has one subplot two-factor interaction that is estimated in the whole-plot stratum, while the second design has four such two-factor interactions. However, since all these two-factor interactions are aliased with whole-plot main effects, both designs have the same number, 0, of subplot two-factor interactions that are not aliased with main effects and are estimated in the whole-plot stratum. It follows that they perform equally well under the $\tilde{W}_k^r$-criteria and have exactly the same information capacities for all k and r. In addition to $(n_1, n_2, p_1, p_2) = (1, 8, 0, 5)$, there are three other 16-run cases, with $(n_1, n_2, p_1, p_2) = (1, 9, 0, 6)$, $(1, 10, 0, 7)$, $(3, 10, 1, 8)$, where the two top designs listed in Bingham & Sitter's (2001) Table 3 are $\tilde{W}_k^r$-optimal, and have identical information capacities, for all k and r.

However, this phenomenon does not occur in the case of 32-run designs. Furthermore, in seven of the twelve 32-run cases, including the one discussed in Example 2, where Bingham &

Table 1. *Seven optimal 32-run split-plot designs*

$n_1.n_2.p_1.p_2$	Independent defining words
3.4.0.2	$ABpr,\ ACpqs$
5.2.1.1	$ABCE,\ ABDpq$
3.5.0.3	$ABpr,\ ABqs,\ ACpqt$
4.4.0.3	$ABpq,\ ACDpr,\ BCDps$
5.3.1.2	$ABCE,\ ABpq,\ ACDpr$
3.6.0.4	$ABpr,\ ABqs,\ ACpqt,\ BCpqu$
5.4.1.3	$ABCE,\ ABpq,\ ACDpr,\ BCDps$

Sitter's (2001) secondary criterion, and the counterpart of the W_k^r criterion for split-plot designs, failed to produce unique optimal designs, we are able to find a unique $\tilde{W}_k^r$-optimal design. These seven cases and the corresponding optimal designs are provided in Table 1. These designs are better than the other designs that are also optimal under Bingham & Sitter's (2001) secondary criterion, because they have more uniform distributions, over the alias sets, of the two-factor interactions that are not aliased with main effects and are estimated in the subplot stratum. In each of the other five cases, namely $(n_1, n_2, p_1, p_2) = (1,6,0,2)$, $(2,5,0,2)$, $(1,7,0,3)$, $(2,6,0,3)$ and $(1,8,0,4)$, there are two optimal designs which have identical information capacities for all k and r.

ACKNOWLEDGEMENT

Ching-Shui Cheng's research was supported by a grant from the U.S. National Science Foundation and Pi-Wen Tsai's research was supported by a grant from the Taiwan National Science Council. The authors would like to thank the referees for their helpful comments.

APPENDIX

Derivation of a good surrogate for maximizing $I_k(d)$

When a model is not estimable, the determinant of the information matrix is zero. Therefore, if $k > f$, then, since not all the k two-factor interactions can be estimated, $I_k(d)$ must be equal to zero. For $k \le f$, since a normalized estimable contrast is estimated with variance ξ_2 if it is orthogonal to blocks and with variance $\xi_1 = \xi_2/r$ if it is confounded with, blocks, for each estimable model containing all the main effects and k two-factor interactions, if t of these two-factor interactions are confounded with blocks, then the determinant of the information matrix for the two-factor interactions is equal to r^t. The total number of models in which t two-factor interactions are confounded with blocks and the other $k-t$ are orthogonal to blocks is given by

$$\left(\sum_{1\le i_1<\cdots<i_t\le h}\prod_{u=1}^{t} m_{i_u}\right)\left(\sum_{h+1\le j_1<\cdots<j_{k-t}\le f}\prod_{v=1}^{k-t} m_{j_v}\right).$$

Thus $I_k(d)$ is equal to

$$\sum_{t=\max(0,k-f+h)}^{\min(k,h)} r^{t/k}\cdot\left(\sum_{1\le i_1<\cdots<i_t\le h}\prod_{u=1}^{t} m_{i_u}\right)\left(\sum_{h+1\le j_1<\cdots<j_{k-t}\le f}\prod_{v=1}^{k-t} m_{j_v}\right)\Big/\binom{n(n-1)/2}{k}$$

$$=\sum_{t=\max(0,k-f+h)}^{\min(k,h)}\left(\sum_{1\le i_1<\cdots<i_t\le h}\prod_{u=1}^{t}(r^{1/k}m_{i_u})\right)\left(\sum_{h+1\le j_1<\cdots<j_{k-t}\le f}\prod_{v=1}^{k-t} m_{j_v}\right)\Big/\binom{n(n-1)/2}{k}$$

$$=E_k(r^{1/k}m_1,\ldots,r^{1/k}m_h,m_{h+1},\ldots,m_f)\Big/\binom{n(n-1)/2}{k}, \tag{A1}$$

where

$$E_k(x_1,\ldots,x_f)=\sum_{1\le i_1<\cdots<i_k\le f}\prod_{j=1}^{k} x_{i_j},$$

as in (3.1) of Cheng et al. (1999). It follows from the same argument as was employed there that the quantity in (A1) is large if $\sum_{i=1}^{h} r^{1/k}m_i+\sum_{i=h+1}^{f} m_i$ is large and $r^{1/k}m_1,\ldots,r^{1/k}m_h,m_{h+1},\ldots,m_f$ are as equal as possible.

REFERENCES

BAILEY, R. A. (1981). A unified approach to design of experiments. *J. R. Statist. Soc.* A **144**, 214–23.

BINGHAM, D. & SITTER, R. R. (1999a). Minimum aberration two-level fractional factorial split-plot designs. *Technometrics* **41**, 62–70.

BINGHAM, D. & SITTER, R. R. (1999b). Some theoretical results for fractional factorial split-plot designs. *Ann. Statist.* **27**, 1240–55.

BINGHAM, D. R., SCHOEN, E. D. & SITTER, R. R. (2004). Designing fractional factorial split-plot experiments with few whole-plot factors. *Appl. Statist.* **53**, 325–39.

CHEN, H. & CHENG, C. S. (1999). Theory of optimal blocking of 2^{n-m} designs. *Ann. Statist.* **27**, 1948–73.

CHENG, C. S., STEINBERG, D. M. & SUN, D. X. (1999). Minimum aberration and model robustness for two-level factorial designs. *J. R. Statist. Soc.* B **61**, 85–93.

CHENG, S. W. & WU, C. F. J. (2002). Choice of optimal blocking schemes in two-level and three-level designs. *Technometrics* **44**, 269–77.

COX, D. R. (1958). *Planning of Experiments*. New York: Wiley.

FRIES, A. & HUNTER, W. G. (1980). Minimum aberration 2^{k-p} designs. *Technometrics* **22**, 601–8.

HUANG, P., CHEN, D. & VOELKEL, J. O. (1998). Minimum-aberration two-level split-plot designs. *Technometrics* **40**, 314–26.

LI, W. & NACHTSHEIM, C. J. (2000). Model-robust factorial designs. *Technometrics* **42**, 245–52.

MCLEOD, R. G. & BREWSTER, J. F. (2004). The design of blocked fractional factorial split-plot experiments. *Technometrics* **46**, 135–46.

MILLER, A. (1997). Strip-plot configurations of fractional factorials. *Technometrics* **39**, 153–61.

MUKERJEE, R. & FANG, K. T. (2002). Fractional factorial split-plot designs with minimum aberration and maximum estimation capacity. *Statist. Sinica* **12**, 885–903.

NELDER, J. A. (1965). The analysis of randomized experiments with orthogonal block structure. I. block structure and the null analysis of variance. *Proc. R. Soc. Lond.* A **283**, 147–62.

SITTER, R. R., CHEN, J. & FEDER, M. (1997). Fractional resolution and minimum aberration in blocking factorial designs. *Technometrics* **39**, 382–90.

TSAI, P. W., GILMOUR, S. G. & MEAD, R. (2000). Projective three-level main effects designs robust to model uncertainty. *Biometrika* **87**, 467–75.

XU, H. (2006). Blocked regular fractional factorial designs with minimum aberration. *Ann. Statist.* **34**, 2534–53.

XU, H. & LAU, S. (2006). Minimum aberration blocking schemes for two- and three-level fractional factorial designs. *J. Statist. Plan. Infer.* **136**, 4088–118.