Designing fractional factorial split-plot experiments with few whole-plot factors

D. R. Bingham,

University of Michigan, Ann Arbor, USA

E. D. Schoen

TNO TPD, Delft, the Netherlands

and R. R. Sitter

Simon Fraser University, Burnaby, Canada

Summary. When it is impractical to perform the experimental runs of a fractional factorial design in a completely random order, restrictions on the randomization can be imposed. The resulting design is said to have a split-plot, or nested, error structure. Similarly to fractional factorials, fractional factorial split-plot designs can be ranked by using the aberration criterion. Techniques that generate the required designs systematically presuppose unreplicated settings of the whole-plot factors. We use a cheese-making experiment to demonstrate the practical relevance of designs with replicated settings of these factors. We create such designs by splitting the whole plots according to one or more subplot effects. We develop a systematic method to generate the required designs and we use the method to create a table of designs that is likely to be useful in practice.

1. Introduction

Two-level fractional factorial (FF) designs are often used to screen a large number of factors in industrial settings. By varying each of the factors over the selected levels and performing experimental trials or runs, experimenters can construct simple models that describe the process sufficiently adequately to determine which of the factors are influential. Interest in FF designs dates as far back as Yates (1937), Fisher (1942) and Finney (1943).

In some cases, for reasons involving cost, the nature of the process or the experimental venue, it may be impractical to perform experimental runs in a completely random order. Thus, we may want to impose restrictions on the randomization. Such restrictions could also be imposed to increase the precision of effect estimates for certain identified factors. The result is what is typically termed a split-plot structure and in this setting these designs are usually referred to as fractional factorial split-plot (FFSP) designs (Kempthorne (1952), page 318; Box and Jones (1992)).

To aid practitioners, Addelman (1964) presented a catalogue of 36 two-level FFSP designs. These are constructed by confounding interactions between subplot factors with whole plots.

Address for correspondence: R. R. Sitter, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada.
E-mail: sitter@stat.sfu.ca

Addelman called this technique 'split-plot confounding'. In his approach, a blocked design for the subplot factors is merged with a design for whole-plot factors. However, the best way of merging the respective subdesigns, and indeed the choice of these subdesigns, is not made explicit. Bisgaard (2000) presented rules to calculate the defining relation in a split-plot design from the defining relation of the respective subdesigns and the subplot effects confounded with whole plots. The choice of the subdesigns and the confounded effects is, again, left open. Schoen (1999), also dealing with split-plot confounding, gave more explicit guidelines for these practical issues.

None of the attempts by the aforementioned researchers suggest a systematic approach and thus there is no guarantee that a design which is constructed by their methods is optimal in any sense. A connection with optimality according to the aberration criterion (Fries and Hunter, 1980) was given by Huang *et al.* (1998) and Bingham and Sitter (1999, 2001). They also gave algorithms to create minimum aberration (MA) FFSP designs. Loosely speaking, an MA FFSP design collects as much information as possible about the process given the fractionation that is used at the whole-plot level and the subplot level of the design (a more formal definition will be given in what follows). Like FF experiments, one often selects the smallest MA FFSP design that meets the experimental goals. However, unlike FF designs, FFSP designs have two sources of error which are used to assess the significance of the effects. Thus, there are important issues of estimation, precision and the identification of which factors are important that affect the selection of a design and are unique to the split-plot situation. These issues are also considered in the area of response surface methodology. We refer to Letsinger *et al.* (1996) and Goos and Vandebroek (2001). Note that Letsinger *et al.* (1996) used the heading of bi-randomization to refer to the nested error structure.

Despite the recent developments in FFSP designs, many practical issues await a solution. One of these occurs when the existing techniques for creating designs require too many subplots per whole plot, or too few whole plots. Indeed, the algorithms of Bingham and Sitter (1999) or Huang *et al.* (1998) always give designs with unreplicated, and possibly fractional, subdesigns for the whole-plot factors. This implies a limitation in the number of whole plots. Given the total number of subplots in a design, the limitation could result in a number of subplots per whole plot that is physically impossible to attain. Further, the analysis becomes quite difficult when there are just a few whole plots. Thus, we sometimes prefer designs with replicated settings of the whole-plot factors.

In this paper, we present a systematic method to split the whole plots of an FFSP design, resulting in a decrease in the whole-plot size and an increase in the number of whole plots. The algorithm of Bingham and Sitter (1999) is then extended to obtain MA FFSP designs with replicated whole plots. We should note that implicit in this paper is the assumption that all factors are of equal interest. In some cases this is not so. For example, in robust parameter experiments the control factor by noise factor interactions are of increased interest. In such cases the aberration criterion as defined here needs to be altered depending on the specific experimental situation and goals (see Wu and Hamada (2000), chapter 10, Zhu (2000) and Bingham and Sitter (2003)).

In the next section, we shall discuss an example from the cheese-making industry and use it to introduce the notation that is required for describing FFSP designs. In Section 3, we discuss the splitting method and we present a table of useful designs. Section 4 gives an algorithm for the systematic creation of the designs required. Section 5 presents a few related theoretical results. We conclude with a brief discussion.

2.   A cheese-making example

Consider the following example from the food production industry. In a cheese-making factory,

there were problems with some quality characteristics of the cheeses being produced. The production of cheeses consists of two stages. First, milk is processed into a batch of curds. These curds are then processed into the resulting cheese.

Discussion between food technologists revealed nine factors that might be contributing to the problems of quality; $k_1 = 2$ of these factors, denoted A and B, affected the treatment of the milk in the first stage, whereas the remaining $k_2 = 7$ factors, denoted p, q, r, s, t, u and v, were related to processing conditions to generate the curds from which the cheeses are made. It was decided that a designed experiment should be run to investigate the effects of these factors on the resulting quality of the cheeses being produced.

Each of the $k_1 + k_2 = 9$ factors were to be investigated at two levels. For confidentiality, details on the factors themselves are omitted and the levels of the factors are referred to as the '+' and '−' levels for each factor. Because of the high cost of experimentation, it was impossible to make cheeses at all the $2^9 = 512$ possible factor level settings. There was sufficient budget in this industrial setting for producing a total of 32 curds. To reduce the run size of the experiment one could elect to perform a 2^{k-p} FF design.

As milk in a single tank gives rise to several batches of curds, and as there are factors operating on the contents of a storage tank and on processing into curds, this was clearly an ideal situation to run a split-plot experiment, i.e. to treat the milk in a tank under a randomly selected setting of the factors A and B, termed whole-plot factors, and then to apply different randomly selected settings of factors p–v, termed subplot factors, to produce a batch of curds.

In such a scenario, the factors affecting the treatment of the milk, whole-plot factors A and B, when considered alone form a completely randomized design. In contrast, the second-stage processing factors p–v when considered separately form a randomized block design. To construct the design we assign p_1 whole-plot factors and p_2 subplot factors to interactions involving the remaining factors, where $p = p_1 + p_2$. The design matrix for this experimental set-up is identical to a 2^{k-p} FF design, where $k = k_1 + k_2$ and $p = p_1 + p_2$, and is thus called a $2^{k_1+k_2-(p_1+p_2)}$ FFSP design. Although an FFSP design matrix corresponds to an FF design matrix, the randomization of the experiment is different.

The choice of factor level settings to be performed is determined by the *p fractional generators*. For example, suppose that we chose to run the above experiment as a $2^{2+7-(0+4)}$ design. To do this, we must assign no whole-plot factors and four subplot factors to interactions involving the remaining factors. A possible assignment of these factors is $s = ABq$, $t = Apq$, $u = ABpr$ and $v = Aqr$. In this case, the settings of factor s for each run are determined by the level combinations of factors A, B and q, for example. Letting I be a column of +s, the fractional generators for this design are $ABqs$, $Apqt$, $ABpru$ and $Aqrv$, i.e. the words containing only one of the *added factors s*, t, u and v. These four generators imply 11 other relations to form the *defining contrast subgroup*,

$$I = ABqs = Apqt = Aqrv = Bpst = Brsv = prtv = pqrsu = qstuv = Arstu$$
$$= Apsuv = Bqrtu = Bpquv = ABpru = ABtuv = ABpqrstv. \quad (1)$$

In this example, there is no fractionation of the whole-plot factors and thus all words in the defining contrast subgroup contain subplot factors. In general, the whole-plot factors might be fractionated. If so, the defining contrast subgroup would contain some words with only whole-plot factors. In this paper, we shall restrict ourselves primarily to a discussion of cases with no whole-plot fractionation, as this would be the most likely situation for replicated whole plots.

Words of different length in the defining contrast subgroup have different implications. Words with only three factors cause main effects to be aliased with two-factor interactions. Similarly, length 4 words alias main effects with three-factor interactions and two-factor interactions with

two-factor interactions, whereas words of length 5 alias main effects with four-factor interactions and two-factor interactions with three-factor interactions, and so on. In many situations, empirical evidence suggests that interactions involving three or more factors are negligible. We shall assume this to be so for the remainder of the paper. If there are only words of length 5 and higher in the defining contrast subgroup, then all main effects and two-factor interactions are aliased with negligible effects. Consequently, if we assume that all interactions involving three or more factors are negligible, then the run size of the experiment may be reduced without any loss of information on the main effects and two-factor interactions. In conclusion, designs with as few short words as possible are preferred.

Let $A_i(D)$ denote the number of words of length i in the defining contrast subgroup of design D and let

$$W = (A_3(D), A_4(D), A_5(D), \ldots)$$

be the word length pattern of the design (words of length 1 or 2 imply designs which are not useful). The *resolution* of a design is the length of the shortest word in the defining contrast subgroup. So the above $2^{2+7-(0+4)}$ FFSP design has resolution IV and word length pattern $W = (0, 6, 8, 0, 0, 1)$.

Designs with larger resolution are typically said to be better than designs with smaller resolution. However, designs with equal resolution may have different word length patterns and therefore will not be the same. A refinement of the resolution criterion that distinguishes between designs with equal resolution is the aberration criterion (Fries and Hunter, 1980). When applied to FFSP designs (Huang *et al.*, 1998; Bingham and Sitter, 1999), the concept can be summarized as follows.

Suppose that D_1 and D_2 are $2^{k_1+k_2-(p_1+p_2)}$ FFSP designs. If $A_i(D_1) = A_i(D_2)$ for $i = 3, \ldots,$ $r-1$ and $A_r(D_1) < A_r(D_2)$, we say that D_1 has less aberration than D_2. A design is said to be MA if no other design has less aberration.

The aberration criterion provides a good general rule for comparing designs, particularly those designs with equal resolution, if we wish to treat all factors and effects of the same order equally. If we have particular interest in a subset of effects, then some other ranking of designs is advisable (see Bingham and Sitter (2003), where they consider robust parameter design).

The design in expression (1) is one of three non-isomorphic MA $2^{2+7-(0+4)}$ FFSP designs that were given in Bingham and Sitter (2001). Note that this FFSP design allows whole-plot factors in the subplot fractional generators (Kempthorne (1952), page 318). However, the converse is not allowable. If subplot factors appear in whole-plot generators the split-plot nature of the design is destroyed. Generally, using whole-plot factors in the subplot fractional generators is useful as it allows us to reduce the aberration. We should also note that in this case the MA FFSP design matrix is the same as the MA 2^{9-4} FF design matrix, i.e. it is possible in this case to use the MA FF and to run it as a split plot with four whole-plot runs and eight subplot runs within each whole plot. The method for obtaining MA FFSP designs in Huang *et al.* (1998) involved taking the MA FF and attempting to split it for various split-plot situations. If this is possible the resulting FFSP will be MA. This is not always possible as some MA FF designs cannot be split in certain ways (Bingham and Sitter, 1999).

In the cheese-making experiment there are just two whole-plot factors and thus it is natural to consider the above arrangement in four whole plots of eight subplots each. This is in fact the recommendation if we obtain an MA design from Table 4 of Bingham and Sitter (2001). Such a design was not used, however. Firstly, a single supply of milk cannot yield eight curds. Secondly, even if this were possible, as we shall discuss in what follows, an unreplicated design

with only four whole plots would give insufficient information (too few degrees of freedom) to separate active whole-plot effects from inactive effects. This is because the whole-plot effects must be evaluated separately from subplot effects, in view of the different standard errors that are required for the evaluation. If the evaluation involves half-normal (or normal) plotting of effects, separate plots are to be made for the respective sets of effects (Daniel, 1959; Box and Jones, 1992; Bingham and Sitter, 2001). The problem that is encountered if there are few whole plots is that a half-normal (or normal) plot has too few effects to be useful. A minimum of seven is highly recommended (Schoen, 1999). This would also be true of other more sophisticated analysis methods (Schoen and Kaul, 2000; Loeppky and Sitter, 2002).

This reasoning leaves us with options for eight whole plots and for 16 whole plots respectively. The latter, however, would require more time, because the batches of milk can only be handled one after another. Thus, the eight-whole-plot option was settled on, in part for operational and cost reasons and in part to ensure enough degrees of freedom at the whole-plot level of the design to separate active from inactive whole-plot effects.

3. Splitting a $2^{k_1+k_2-(0+p_2)}$ fractional factorial split plot into 2^{k_1+r} whole plots

The eventual design that was used for the cheese-making experiment was adapted *ad hoc* from the MA 2^{9-4} FF design (Chen *et al.*, 1993) by using an approach that was similar to that of Huang *et al.* (1998). The defining relation is identical to that given in expression (1). Thus, in this case the MA FF can be run in $4 = 2^{k_1}$ as well as in $8 = 2^{k_1+1}$ whole plots.

To understand how this can be so, let us look carefully at the design matrix that was used for the cheese-making experiment that is displayed in Table 1. The second to sixth columns are the independent columns whose rows represent all 32 possible level settings of five two-level factors. Factors A, B, p, q and r (termed basic factors) are assigned to these and then factors s, t, u and v (termed added factors) are assigned to interactions between these columns obtained through componentwise multiplication. This is what determines the aliasing relationships that are implied by the defining contrast subgroup in expression (1).

If the experiment were to be run as a completely randomized FF design, then the second to 10th columns of Table 1 would represent the 32 level settings to be run and these would be run in random order. If we wished to run the design as a split plot with four whole plots and eight subplots, we would randomly choose one of the four possible settings of whole-plot factors A and B, $(-,-)$, $(-,+)$, $(+,-)$ and $(+,+)$, fix this setting and keep it fixed while performing the eight factor level settings of the subplot factors p–v indicated at that fixed setting of A and B.

If columns A and B are used to define the whole plots, then the AB-interaction column also is at the whole-plot level of the randomization since if A and B are held fixed then AB is fixed. In this design, the defining contrast subgroup in expression (1) implies the alias relation

$$AB = qs = Bpqt = Bqrv = Apst = Arsv = ABprtv = ABpqrsu$$
$$= ABqstuv = Brstu = Bpsuv = Aqrtu = Apquv = pru = tuv = pqrstv,$$

so that these interactions are all moved to the whole-plot level of the design. In this case, the only effects of interest from this group are $AB = qs$.

One way to run the design matrix in Table 1 in $8 = 2^{k_1+1}$ whole plots each containing four subplot runs is to use the settings of A, B and the $Apqr$ column to split the design, i.e. we randomly choose one of the eight possible combinations of settings in these three columns, and then, while holding this fixed, randomly perform the four runs so indicated. By so running the experiment, effects A, B, $Apqr$, AB, pqr, $ABpqr$ and $Bpqr$ and all effects aliased with them would all now be at the whole-plot level of the design. Of the effects of interest, this would

Table 1. Cheese-making experimental design matrix

Settings	A	B	p	q	r	s	t	u	v	Apqr	y
1	−	−	−	−	−	−	−	+	−	+	y_1
2	+	−	−	−	−	+	+	−	+	−	y_2
3	−	+	−	−	−	+	−	−	−	+	y_3
4	+	+	−	−	−	−	+	+	+	−	y_4
5	−	−	+	−	−	+	+	−	−	−	y_5
6	+	−	+	−	−	+	−	+	+	+	y_6
7	−	+	+	−	−	+	+	+	−	−	y_7
8	+	+	+	−	−	−	−	−	+	+	y_8
9	−	−	−	+	−	+	+	+	+	−	y_9
10	+	−	−	+	−	−	−	−	−	+	y_{10}
11	−	+	−	+	−	−	+	−	+	−	y_{11}
12	+	+	−	+	−	+	−	+	−	+	y_{12}
13	−	−	+	+	−	+	−	−	+	+	y_{13}
14	+	−	+	+	−	−	+	+	−	−	y_{14}
15	−	+	+	+	−	−	−	+	+	+	y_{15}
16	+	+	+	+	−	+	+	−	−	−	y_{16}
17	−	−	−	−	+	−	−	−	+	−	y_{17}
18	+	−	−	−	+	+	+	+	−	+	y_{18}
19	−	+	−	−	+	+	−	+	+	−	y_{19}
20	+	+	−	−	+	−	+	−	−	+	y_{20}
21	−	−	+	−	+	−	+	+	+	+	y_{21}
22	+	−	+	−	+	+	−	−	−	−	y_{22}
23	−	+	+	−	+	+	+	−	+	+	y_{23}
24	+	+	+	−	+	−	−	+	−	−	y_{24}
25	−	−	−	+	+	+	+	−	−	+	y_{25}
26	+	−	−	+	+	−	−	+	+	−	y_{26}
27	−	+	−	+	+	−	+	+	−	+	y_{27}
28	+	+	−	+	+	+	−	−	+	−	y_{28}
29	−	−	+	+	+	+	−	+	−	−	y_{29}
30	+	−	+	+	+	−	+	−	+	+	y_{30}
31	−	+	+	+	+	−	−	−	−	−	y_{31}
32	+	+	+	+	+	+	+	+	+	+	y_{32}

imply that A, B, $AB = qs$, $pv = rt$, qu and su would all be at the whole-plot level of the design. However, since none of the subplot main effects p–v have moved, the design maintains the desired split-plot structure.

The reason why an FF design could not be run in a desired split-plot mode revolves around the aliasing structure that is implied in the defining contrast subgroup, or around the additional effects that are chosen to define the whole plots. If AB would be aliased with a main effect, the above design could not be run in four whole pots without raising that main effect to the whole-plot level of randomization. Thus, the split-plot structure would be destroyed. This will remain so even if the subplot factor is only aliased with a higher order whole-plot interaction that has been assumed negligible. If we would choose A, B and ABp to define whole plots, then main effect p would be raised to the whole-plot level, and the split-plot structure, again, would be destroyed.

In the cheese-making experiment, there are eight subplot treatments for each of the four possible level settings for the whole-plot factors A and B. Because it was desirable to perform only four of the eight subplot treatments with the whole-plot factors fixed, we must decide which four subplot treatments to group together. This was done by using the $Apqr$-interaction, i.e. the subplot treatments, for fixed whole-plot level settings, were divided into two groups based on

the sign of the $Apqr$-interaction. This amounts to blocking the subplot treatments within each fixed whole-plot level setting.

More formally, a *splitting factor* ρ can be used to assign subplot treatments into two groups to be run together for fixed whole-plot level settings. In the previous example, the *splitting generator* $I = Apqr\rho$ can be crossed with an FFSP's defining contrast subgroup, and those effects that are aliased with whole-plot main effects, ρ, interactions involving only whole-plot factors or interactions involving whole-plot main effects and ρ are moved to the whole-plot level of the design.

In general, if we wish to split the subplot treatments, for each fixed whole-plot level setting, into 2^r groups, then r splitting factors are required. In this case, the splitting generators, which are similar to blocking generators in blocked FF designs (for example, see Bisgaard (1994) and Sitter *et al.* (1997)), have an associated defining contrast subgroup. The defining contrast subgroup of the FFSP design is used to identify which effects are aliased. The combined defining contrast subgroup of the FFSP design and the splitting generators identifies which effects are tested at the whole-plot and subplot levels respectively. In what follows, we shall introduce and discuss a systematic method for using splitting factors to obtain MA FFSP designs with more whole-plot runs than the number of possible whole-plot factor settings. Before doing so, however, we use the method to generate some useful designs.

Table 2 lists MA $2^{k_1+k_2-(0+p_2)}$ FFSP designs with 2^{k_1+r} whole-plot runs denoted $k_1.k_2.r.p_2$ for 88 16-, 32- and 64-run cases of interest. Note that large FFSP designs are more commonly used than large FF designs because the cost of the experiment is often driven primarily by the number of whole plots and less so by the total number of runs (see Bingham and Sitter (2001) for further discussion). We use a similar representation to that used by Bingham and Sitter (1999) (see also Chen *et al.* (1993) and Sitter *et al.* (1997)). For each run size–number of whole-plot run combination, we arrange the columns in Yates order as given in Table 3 with the independent columns in bold italics. To illustrate, consider the cheese-making experiment which has $k_1 = 2$ whole-plot factors and $k_2 = 7$ subplot factors, and the experimenter wished to use a 32-run design with eight whole plots and four subplots within each whole plot, i.e. a $2^{2+7-(0+4)}$ FFSP design in 2^{2+1} whole plots: design 2.7.1.4 in Table 2. In this 32-run case (i.e. $32 = 2^5$) only the first five rows of Table 3 are used. The first two rows are labelled by the whole-plot factors A and B, whereas the third to fifth rows are labelled by the basic subplot factors p, q and r. From entry 2.7.1.4 in Table 2, we see that the MA design assigns the splitting factor ρ to column 29 of Table 3 and added subplot factors s, t, u and v to columns 11, 13, 23 and 25 respectively. Since column 29 has 1s in rows A, p, q and r, $\rho = Apqr$. Similarly, there are 1s in rows A, B and q of column 11, so $s = ABq$, and so on with the other added factors. It follows that the MA FFSP that is so obtained is exactly the design that was used in the cheese-making experiment. For brevity, Table 2 is not comprehensive. We did not include designs which were equivalent in terms of the MA but were still not isomorphic.

Addelman (1964) gave a catalogue of 36 designs. If we compare this catalogue with the 88 designs in Table 2 only two common cases are considered, both with 64 runs in 16 whole plots and four subplots: designs 3.4.1.1 and 3.5.1.2.

In the second case (3.5.1.2), Addelman recommended the same design as in Table 2. In the first case (3.4.1.1), Addelman's design has defining contrast subgroup

$$I = ABCrs = \rho pqr = \rho ABCpqs,$$

whereas the design that is given in Table 2 has defining contrast subgroup

$$I = ABCpqrs = \rho ABpq = \rho Crs.$$

Table 2. MA FFSP designs using splitting factors to create additional whole plots

Design†	Splitting columns	Subplot columns	Word length pattern
16-run designs with 8 whole plots and 2 subplots			
1.4.2.1	6 10	15	0 0 1
1.5.2.2	6 10	3 15	1 1 1
1.6.2.3	6 10	3 5 14	2 3 2
1.7.2.4	6 10	3 5 9 14	3 7 4 0 1
1.8.2.5	6 10	3 5 9 14 15	4 14 8 0 4 1
2.3.1.1	12	7	0 1
2.4.1.2	12	7 11	0 3
2.5.1.3	12	5 6 11	2 3 2
2.6.1.4	12	5 6 9 10	4 5 4 2
2.7.1.5	12	5 6 7 9 10	6 10 8 4 2 1
2.8.1.6	12	5 6 7 9 10 11	8 18 16 8 8 5
32-run designs with 8 whole plots and 4 subplots			
1.5.2.1	10 18	31	0 0 0 1
1.6.2.2	10 18	15 30	0 1 2
1.7.2.3	10 18	13 23 30	0 3 4
1.8.2.4	10 18	15 21 27 28	0 6 8 0 0 1
1.9.2.5	10 18	13 14 23 28 29	0 10 16 0 0 5
1.10.2.6	10 18	13 14 21 22 26 31	0 25 0 27 0 10 0 1
1.11.2.7	10 18	13 14 21 22 26 28 31	0 38 0 52 0 33 0 4
1.12.2.8	10 18	7 13 14 21 22 26 28 31	0 55 0 96 0 87 0 16 0 1
1.13.2.9	6 24	3 5 9 14 15 17 22 26 28	5 55 45 96 106 87 82 16 17 1 1
1.14.2.10	6 24	3 5 9 14 15 17 22 23 26 28	6 77 62 168 188 203 188 56 62 7 6
2.4.1.1	20	31	0 0 0 1
2.5.1.2	20	25 30	0 1 2
2.6.1.3	20	15 21 27	0 3 4
2.7.1.4	20	15 21 22 27	0 6 8 0 0 1
2.8.1.5	20	7 11 19 29 30	0 10 16 0 0 5
2.9.1.6	20	7 13 19 21 22 26	0 25 0 27 0 10 0 1
2.10.1.7	20	11 13 19 21 22 25 26	0 38 0 52 0 33 0 4
2.11.1.8	28	7 11 13 14 19 21 22 25	0 55 0 96 0 87 0 16 0 1
2.12.1.9	28	7 11 13 14 19 21 22 25 26	0 77 0 168 0 203 0 56 0 7
2.13.1.10	28	5 6 11 12 15 19 20 23 25 26	6 77 62 168 188 203 188 56 62 7 6
32-run designs with 16 whole-plots and 2 subplots			
1.5.3.1	3 5 9	31	0 0 0 1
1.6.3.2	3 5 9	7 27	0 1 2
1.7.3.3	6 10 18	14 22 27	0 3 4
1.8.3.4	6 10 18	14 22 27 29	0 6 8 0 0 1
1.9.3.5	13 14 21	7 11 19 29 30	0 10 16 0 0 5
1.10.3.6	3 5 9	7 11 13 19 21 25	0 25 0 27 0 10 0 1
1.11.3.7	15 23 27	7 11 13 14 19 21 25	0 38 0 52 0 33 0 4
1.12.3.8	15 23 27	7 11 13 14 19 21 22 25	0 55 0 96 0 87 0 16 0 1
1.13.3.9	15 23 27	7 13 14 19 22 25 26 30 31	0 77 0 168 0 203 0 56 0 7
1.14.3.10	6 10 18	9 14 15 17 22 23 26 27 28 29	6 77 62 168 188 203188 56 62 7 6
2.4.2.1	12 20	31	0 0 0 1
2.5.2.2	12 20	11 29	0 1 2
2.6.2.3	12 20	11 19 29	0 3 4
2.7.2.4	12 22	7 11 19 29	0 6 8 0 0 1

<div align="right">(continued)</div>

Table 2 (*continued*)

Design†	Splitting columns	Subplot columns	Word length pattern
2.8.2.5	12 20	7 11 19 29 30	0 10 16 0 0 5
2.9.2.6	12 20	7 11 19 29 30 31	2 14 22 8 6 9 2
2.10.2.7	12 20	5 7 9 10 18 19 31	4 20 32 22 20 19 8 2
2.11.2.8	12 20	5 6 9 10 17 18 28 31	6 29 46 46 50 41 26 10 0 1
2.12.2.9	12 20	5 6 9 10 17 18 19 30 31	8 42 64 85 112 85 64 42 8 0 0 1
2.13.2.10	12 24	5 6 7 9 11 17 18 29 30 31	10 60 90 141 212 193 164 98 34 18 2 1
3.3.1.1	24	15	0 0 1
3.4.1.2	24	11 23	0 1 2
3.5.1.3	24	11 13 23	0 3 4
3.6.1.4	24	11 13 14 23	0 7 7 0 0 0 1
3.7.1.5	24	11 13 14 19 21	0 16 0 12 0 3
3.8.1.6	24	11 13 14 19 21 22	0 26 0 24 0 13
3.9.1.7	24	9 10 12 15 19 21 22	3 26 22 24 28 13 10 0 1
3.10.1.8	24	9 10 13 15 19 20 21 22	6 28 51 42 42 51 28 6 0 0 1
3.11.1.9	24	9 10 12 15 17 18 21 22 23	9 37 73 84 94 103 70 28 9 3 1
3.12.1.10	24	9 10 11 12 15 17 18 21 22 23	12 51 102 144 192 207 164 96 36 13 6

64-run designs with 16 whole plots and 4 subplots

Design†	Splitting columns	Subplot columns	Word length pattern
1.6.3.1	6 10 48	63	0 0 0 0 1
1.7.3.2	6 10 48	15 51	0 0 2 1
1.8.3.3	6 10 18	14 27 54	0 1 4 2
1.9.3.4	6 10 18	14 23 43 54	0 2 8 4 0 1
1.10.3.5	6 10 18	14 22 27 39 58	0 4 14 8 0 3 2
1.11.3.6	6 10 18	14 22 27 39 58 61	0 6 24 16 0 9 8
1.12.3.7	6 10 48	15 23 25 38 42 52 56	0 14 28 24 24 17 12 8
1.13.3.8	40 54 58	15 29 35 38 42 49 52 56	0 22 40 36 56 49 24 20 8
1.14.3.9	40 54 58	15 29 35 38 42 49 52 56 63	0 30 60 60 105 105 60 60 30 0 0 1
2.5.2.1	28 44	63	0 0 0 0 1
2.6.2.2	44 52	15 51	0 0 2 1
2.7.2.3	12 56	7 27 45	0 1 4 2
2.8.2.4	28 56	7 27 43 53	0 2 8 4 0 1
2.9.2.5	24 36	7 11 29 45 51	0 4 14 8 0 3 2
2.10.2.6	20 36	15 28 44 53 54 59	0 6 24 16 0 9 8
2.11.2.7	12 20	19 29 30 35 44 52 59	0 14 28 24 24 17 12 8
2.12.2.8	52 56	7 11 19 30 37 41 49 60	0 22 40 36 56 49 24 20 8
2.13.2.9	52 56	7 11 19 30 37 41 49 60 63	0 30 60 60 105 105 60 60 30 0 0 1
3.4.1.1	27	63	0 0 0 0 1
3.5.1.2	61	15 45	0 0 2 1
3.6.1.3	52	27 45 56	0 1 4 2
3.7.1.4	48	27 29 46 56	0 2 8 4 0 1
3.8.1.5	54	11 13 23 39 57	0 4 14 8 0 3 2
3.9.1.6	48	11 13 23 39 57 62	0 6 24 16 0 9 8
3.10.1.7	24	11 13 19 21 42 54 60	0 14 28 24 24 17 12 8
3.11.1.8	56	23 25 26 28 39 41 42 44	0 22 40 36 56 49 24 20 8
3.12.1.9	48	23 25 26 28 39 41 42 44 63	0 30 60 60 105 105 60 60 30 0 0 1

†*a.b.c.d* denotes a design with *a* whole-plot factors, *b* subplot factors, *c* splitting factors and *d* subplot fractional generators.

Table 3.   Matrix for 16-, 32- and 64-run designs†

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0
0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1
0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

22	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*	*41*	*42*
0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0
1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1
1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0
0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1
1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1

43	*44*	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*	*56*	*57*	*58*	*59*	*60*	*61*	*62*	*63*
1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

†The independent columns are in bold italics and numbered 1, 2, 4, 8, 16 and 32. The 16-run designs use the first four rows, the 32-run designs use the first five rows and the 64-run designs use all six rows.

It is clear from the first word of each design that the design in Table 2 has smaller aberration. However, if we truly believe that three-factor interactions are negligible and can be ignored, Addelman's design can be viewed as being superior in this case. To see this, note that both designs have resolution greater than 5 and thus have all main effects and two-factor interactions estimable assuming that three-factor interactions and higher are negligible. Addelman's design has only one subplot two-factor interaction at the whole-plot level, whereas the design in Table 2 has two. This illustrates that no single criterion, including MA, is the best for every experimental situation.

As a final point, note that the splitting columns in Table 2 form a group. By this virtue, they could also be used to block the whole plots of the corresponding designs. However, this would result in a loss of degrees of freedom for the whole-plot effects. Also, as the blocking is performed according to effects involving subplot basic factors, losing these effects could be unattractive. Clearly, further study is required to develop this issue more comprehensively.

4.   A systematic method for generating designs

In this section we describe the proposed systematic method in detail for the general situation where the experimenter desires more whole plots than the number of whole-plot factor level combinations, 2^{k_1}, i.e. where it is desirable to replicate the whole-plot settings.

4.1.   The procedure
It turns out that we can construct such designs by using a very simple method in combination

with the algorithm of Bingham and Sitter (1999), which efficiently finds all non-isomorphic FFSP designs given a particular split and fractionation, $2^{k_1+k_2-(p_1+p_2)}$. We state the method as it would be applied to the situation where we wish to find the MA $2^{k_1+k_2-(0+p_2)}$ FFSP design run in 2^{k_1+r} whole plots, denoted $2^{k_1+[k_2+r]-(0+[p_2+r])}$, with the r emphasizing that the r splitting factors are initially treated as additional subplot factors.

Step 1: let $\rho_1, ..., \rho_r$ be splitting factors that are used for construction but which will have no factors assigned to them when running the experiment.

Step 2: use the algorithm of Bingham and Sitter (1999) to obtain the set of non-isomorphic $2^{k_1+[k_2+r]-(0+[p_2+r])}$ FFSP designs. Do this by treating the r splitting factors as if they were additional subplot factors at this point. In this setting, two designs are isomorphic if one can be obtained from the other by relabelling between whole-plot factors, between subplot factors and between splitting factors.

Step 3: remove designs which do not have the correct split-plot structure by eliminating all designs with at least one word in the defining contrast subgroup that contains only one subplot factor and all designs containing words with only splitting factors. The term subplot factor does not include the splitting factors here.

Step 4: reduce the defining contrast subgroups to reflect the fact that the splitting factors only indicate the randomization structure and are not true factors by eliminating all words containing splitting factors from the defining contrast subgroup of the remaining set of non-isomorphic designs and choosing those with least aberration.

When actually running the experiment we treat the splitting factor columns as whole-plot factors for the purposes of randomization only.

This procedure is general and applies to designs with one or more splitting generators. To elaborate on the need to avoid words with only splitting factors, consider the case where $r = 2$, i.e. there are two splitting generators. Suppose that $\rho_1 = pq$ and $\rho_2 = pr$, or equivalently $I = pq\rho_1 = pr\rho_2$. These relationships imply a third ($I = qr\rho_1\rho_2$) and thus the splitting generators have an associated defining contrast subgroup of their own. If we wanted to add a third splitting generator, we could not select $I = qr\rho_3$ because this would imply that $\rho_3 = \rho_2\rho_1$ and the desired split would not occur.

4.2.   A simple illustrating example

Suppose that we wish to run a $2^{1+3-(0+1)}$ FFSP in four whole-plot runs. We let A denote the whole-plot factor, p, q and r the subplot factors and ρ the splitting factor and use the search table algorithm of Bingham and Sitter (1999) to obtain the four non-isomorphic $2^{1+[3+1]-(0+[1+1])} = 2^{1+4-(0+2)}$ FFSP designs (i.e. steps 1 and 2): design D_1,

$$I = Apr = Aq\rho = pqr\rho;$$

design D_2,

$$I = Apr = pq\rho = Aqr\rho;$$

design D_3,

$$I = pqr = Ap\rho = Aqr\rho;$$

design D_4,

$$I = Apqr = Ap\rho = qr\rho.$$

Table 4. MA $2^{1+3-(0+1)}$ FFSP run in four whole plots

Whole-plot run	Subplot run	$A = pr$	$\rho = pq$	$A\rho = qr$	q	Aq	$p = Ar$	$r = Ap$
1	1	−	−	+	−	+	+	−
	2	−	−	+	+	−	−	+
2	1	+	−	−	−	−	+	+
	2	+	−	−	+	+	−	−
3	1	−	+	−	−	+	−	+
	2	−	+	−	+	−	+	−
4	1	+	+	+	−	−	−	−
	2	+	+	+	+	+	+	+

Designs D_2, D_3 and D_4 would be considered isomorphic if p, q, r and ρ were treated equally. This illustrates the point that was made in step 2 that isomorphism is assessed by relabelling within the three types of factors, whole plot, subplot and splitting. So designs D_2, D_3 and D_4 differ in the placement of the splitting factor ρ. The ρ-column splits each subplot into two. Since ρ is not a real factor, when considering aberration and resolution words containing ρ do not count. However, ρ identifies which effects are moved to the whole-plot level of the design. So, any effect which is aliased with ρ or with an interaction involving whole-plot factors and/or ρ are tested against whole-plot error. This is the reason for step 3 in the method. We must remove any design with defining contrast subgroup having a word with only a single subplot factor. We see that designs D_1, D_3 and D_4 all fall into this category. For example, in D_1, $q = A\rho$ and thus q does not vary if the levels of factor A and splitting factor ρ are held fixed, i.e. q is moved to the whole-plot level of the design. Thus, the MA FFSP design under the stated restrictions is D_2.

Consider design D_2 carefully. It is an eight-run design which is run in four whole plots as depicted in Table 4. In Table 4, we have kept only main effects, two-factor interactions and the splitting factor ρ. We see that the splitting factor ρ is used together with A to determine the whole-plot structure. We can also see that pr, pq and qr are all at the whole-plot level of the design, since they are aliased with A, ρ and $A\rho$ respectively.

4.3. The cheese-making experiment revisited
Now let us reconsider the cheese-making experiment by applying the algorithm to this situation. To do so, we run the algorithm of Bingham and Sitter (1999) on the case $2^{2+[7+1]-(0+[4+1])}$ to obtain all non-isomorphic designs with the extended definition of isomorphism in steps 2 and 3. Ranking the non-isomorphic split FFSP designs in terms of the aberration of the fractional factorial portion of the design yielded seven non-isomorphic MA designs. The designs are listed in Table 5. Indeed, the seven designs can be viewed as splitting the whole-plot runs for two of the three MA $2^{2+7-(0+4)}$ designs in Bingham and Sitter (2001). In the next section we shall demonstrate that the third design cannot be split and thus was discarded in step 3 of the algorithm.

Consider the extended aberration criterion of Bingham and Sitter (2000) which ranks designs with fewer subplot-by-subplot two-factor interactions at the whole-plot level as better. The numbers of these two-factor interactions associated with the designs in Table 5 vary in conjunction with the fractional generators; they do not vary with the splitting generators that are used for

Table 5. MA $2^{2+7-(0+4)}$ FFSP run in eight whole plots

Fractional generators	Split generator	Number of subplot 2-factor interactions with whole-plot error
$ABqs$, $Apqt$, $ABpru$, $Aqrv$	$Apqr\rho$	5
$ABqs$, $Apqt$, $ABpru$, $Aqrv$	$Bpqr\rho$	5
$ABqs$, $Apqt$, $ABpru$, $Aqrv$	$ABpqr\rho$	5
$ABqs$, $Apqt$, $ABpru$, $Aqrv$	$pqr\rho$	5
$ABps$, $ABqt$, $ABru$, $Apqrv$	$pq\rho$	9
$ABps$, $ABqt$, $ABru$, $Apqrv$	$Apq\rho$	9
$ABps$, $ABqt$, $ABru$, $Apqrv$	$ABpq\rho$	9

each of the two fractional generators. By choosing the first design in Table 5, the experimenters used an MA design with the fewest two-factor interactions tested at the whole-plot level. It is MA and has the fewest two-factor interactions tested as whole-plot effects, i.e. extended MA.

The selection of the design for the cheese-making experiment raises an interesting issue with respect to the true number of different split FFSP designs. In this example, there are seven non-isomorphic designs for the experimenter to choose from. The practical issue facing the experimenter is to select the best extended MA design. The first four designs in Table 5 are the non-isomorphic extended MA designs. But are they different?

The short answer to this question is no. To see why, we first note that the defining contrast subgroups for the fractional generators of the first four designs are the same. The only distinguishing feature between the designs is the splitting generators. The effect of the splitting generator is to move two-factor interactions between subplot factors only, to the whole-plot level of the design. Therefore, whether the two-factor interaction is aliased with ρ or $A\rho$ is not relevant. Thus, the generalized interaction between the subplot factors is the only real distinguishing feature. The simplicity of this argument is a direct consequence of assuming that interactions involving three or more factors are negligible.

5. Some theoretical results

This section presents two theoretical results to help to discover cases where an FFSP design cannot be split. These cases precisely motivate the algorithm proposed. Our first result states a simple case where this can occur for $r = 1$. This is presented for illustration. Result 2 generalizes the result to $r \geqslant 1$.

Result 1. If all generalized interactions between basic subplot factors appear in the generators of the FFSP design the design cannot be run in 2^{k_1+1} whole plots (i.e. $r = 1$) without destroying the original split-plot structure.

For a proof of this result see Appendix A.

To see an example of this, let us again reconsider the cheese-making experiment. The experimenters wished to run a $2^{2+[7+1]-(0+[4+1])}$ FFSP design. There are three non-isomorphic $2^{2+7-(0+4)}$ FFSP designs (Bingham and Sitter, 2001). The algorithm of the previous section yielded non-isomorphic splits of two of these (Table 5). Interestingly, the best $2^{2+7-(0+4)}$ design, according to the extended aberration criterion, could not be split and was discarded in step 3 of the algorithm. The design which was actually used, as given in expression (1), is a split of the second best design. The best $2^{2+7-(0+4)}$ design has generators $Apqs$, $Aprt$, $ABqru$ and

$Bpqrv$. We see then that all four possible generalized interactions, pq, pr, qr and pqr, appear in a generator and result 1 applies. To illustrate, suppose that we try to split the whole plots by using, say, $I = Bpq\rho$. The product of the split generator and the first fractional generator is $I = ABs\rho$. Thus s is aliased with $AB\rho$ and is constant within whole plots. Therefore, s is a whole-plot factor and the desired structure has been destroyed. In this case, the design cannot be split.

Result 2 below generalizes the principle in result 1. The idea is that to split the subplot level settings, for fixed whole-plot factor settings, into 2^r groups, while preserving the original structure of the design, enough degrees of freedom must be available at the subplot level of the design. Recall that there are $2^{k_2-p_2} - k_2 + p_2 - 1$ generalized interactions between basic subplot factors.

Result 2. If the number of generalized interactions between basic subplot factors appearing in the generators of the FFSP design is greater than $2^{k_2-p_2} - k_2 + p_2 - 1 - r$ the design cannot be run in 2^{k_1+r} whole plots without destroying the original split-plot structure.

The proof of result 2 is similar to that for result 1 except that we must identify which of the r splitting generators will destroy the original structure of the design.

Result 2 states a condition where we know that the whole-plot trials of a design cannot be split. However, it does not provide a sufficient condition for when a split does exist. Indeed, similar to blocking, for the runs of the FFSP to be grouped in the desired fashion, contained in the remaining generalized interactions between basic subplot factors must be a group with at least $2^r - 1$ elements.

6. Conclusion

In this paper, we have introduced systematic methodology for redistributing the degrees of freedom for an FF design with a nested error structure. It amounts to splitting the runs of an FFSP design within fixed whole-plot settings. The methodology is easy to implement and gives the experimenter flexibility when deciding how to analyse the experiment. To aid practitioners, a table of MA designs that is obtained by using the proposed method for various situations is provided.

Acknowledgements

Appendix A: Proof of result 1

We prove result 1 by contradiction. Let D be a $2^{k_1+k_2-(0+p_2)}$ FFSP design and assume that the design can be run in 2^{k_1+1} whole plots without destroying the original split-plot structure. Suppose that each of the $2^{k_2-p_2} - k_2 + p_2 - 1$ generalized interactions between basic subplot factors is contained in at least one of the p_2 subplot generators. Let S be the splitting generator. By definition, S is the generalized interaction between at least two basic subplot factors, possibly some number of whole-plot factors and ρ. Because all generalized interactions between basic subplot factors are contained in at least one subplot generator, there is a generator g_1 where the product of S and g_1 contains exactly one subplot added factor. Thus this added subplot factor is aliased with ρ and whole-plot factors. This is a contradiction since the added subplot factor is now a whole-plot factor and the original structure is destroyed.

References

Addelman, S. (1964) Some two-level factorial plans with split-plot confounding. *Technometrics*, **6**, 253–258.

Bingham, D. and Sitter, R. R. (1999) Minimum aberration two-level fractional factorial split-plot designs. *Technometrics*, **41**, 62–70.

Bingham, D. and Sitter, R. R. (2000) Fractional factorial split-plot designs for robust parameter experiments. *Proc. Phys. Engng Sci. Sect. Am. Statist. Ass.*, 12–21.

Bingham, D. and Sitter, R. R. (2001) Design issues for fractional factorial split-plot experiments. *J. Qual. Technol.*, **32**, 2–15.

Bingham, D. and Sitter, R. R. (2003) Fractional factorial split-plot designs for robust parameter experiments. *Technometrics*, **45**, 80–89.

Bisgaard, S. (1994) A note on the definition of resolution for blocked 2^{k-p} designs. *Technometrics*, **36**, 308–311.

Bisgaard, S. (2000) The design and analysis of $2^{k-p} \times 2^{q-r}$ split-plot experiments. *J. Qual. Technol.*, **32**, 39–56.

Box, G. E. P. and Jones, S. (1992) Split-plot designs for robust product experimentation. *J. Appl. Statist.*, **19**, 3–26.

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

Daniel, C. (1959) Use of half-normal plots in interpreting factorial two-level experiments. *Technometrics*, **1**, 311–340.

Finney, D. J. (1943) The fractional replication of factorial arrangements. *Ann. Eugen.*, **12**, 291–301.

Fisher, R. A. (1942) The theory of confounding in factorial experiments in relation to the theory of groups. *Ann. Eugen.*, **11**, 341–353.

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

Goos, P. and Vandebroek, M. (2001) Optimal split-plot designs. *J. Qual. Technol.*, **33**, 436–450.

Huang, P., Chen, D. and Voelkel, J. (1998) Minimum aberration two-level split-plot designs. *Technometrics*, **40**, 314–326.

Kempthorne, O. (1952) *The Design and Analysis of Experiments*. New York: Wiley.

Letsinger, J. D., Myers, R. H. and Lentner, M. (1996) Response surface methods for bi-randomization structures. *J. Qual. Technol.*, **28**, 381–397.

Loeppky, J. L. and Sitter, R. R. (2002) Analyzing unreplicated blocked or split-plot fractional factorial designs. *J. Qual. Technol.*, **34**, 229–243.

Schoen, E. D. (1999) Designing fractional two-level experiments with nested error structures. *J. Appl. Statist.*, **26**, 495–508.

Schoen, E. D. and Kaul, E. A. A. (2000) Three robust scale estimators to judge unreplicated experiments. *J. Qual. Technol.*, **32**, 276–283.

Sitter, R. R., Chen, J. and Feder, M. (1997) Fractional resolution and minimum aberration in blocked 2^{n-k} designs. *Technometrics*, **39**, 382–390.

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

Yates, F. (1937) The design and analysis of factorial experiments. *Technical Communication 35*. Imperial Bureau of Soil Science, Harpenden.

Zhu, Y. (2000) A theory of experimental design for multiple groups of factors. *PhD Thesis*. Department of Statistics, University of Michigan, Ann Arbor.