Construction of Row-Column Factorial Designs

J.D.Godolphin

Department of Mathematics, University of Surrey, UK, GU2 7XH.

Summary. The arrangement of 2^n factorials in row-column designs to estimate main effects and two factor interactions is investigated. Single replicate constructions are given which enable estimation of all main effects and maximise the number of estimable two factor interactions. Constructions and guidance are given for multi-replicate designs in single arrays and in multiple arrays. Consideration is given to constructions for 2^{n-t} fractional factorials.

Keywords: confounding, fractional, generator, interaction, orthogonal, replicate

1. Introduction

When conducting factorial experiments it is often necessary to incorporate blocking factors into the design. In experiments involving animals, litters are commonly used as blocks. In industrial experiments, typical blocking factors include: site; batches of raw material and time period. Accommodating small blocks is especially challenging and it is not always possible to set up a design without confounding some effects of interest.

This work concerns the construction of 2^n factorial designs with double confounding, that is, with two forms of blocking. Designs can be represented as one or more rectangular arrays with rows and columns corresponding to blocking factors. Such an experiment with batches of raw material and machines as blocking factors is described in chapter 9 of Hinkelmann and Kempthorne (2005). Recent applications are found in Dash *et al.* (2013) who describe microarray experiments with dye and microarray as blocking factors, and Datta *et al.* (2017) who outline doubly confounded factorial designs for experiments involving livestock.

Early work on factorial designs with double confounding includes Yates (1937a), Yates (1937b) and Rao (1946), with constructions involving quasi-Latin squares. However, despite this early interest, there has been relatively little work on factorial designs with two forms of blocking. By contrast, factorial designs with one form of blocking have been extensively investigated. In particular, a number of papers in the literature formulate confounding schemes in terms of finite Abelian groups. See for example Fisher (1942), Bailey (1977), Dean and Lewis (1980), Bailey (1985) and Kobilinsky (1985).

Designs for two factors, based on Youden squares and Græco Latin squares, are given in Preece (1966) and Preece (1971). Other work focusing on two factors includes: Dean and Lewis (1992), who study designs with two or more blocking factors; Williams and John (1996), who adapt the simulated annealing algorithm of John and Whitaker (1993) to obtain an algorithm that produces designs in multiple arrays; and Gopinath *et al.* (2017), who give some designs which enable main effects to be estimated with full information.

More recent work specific to designs with factorial treatment structure not limited to two factors, and with double confounding, includes John and Lewis (1983), who use a cyclic approach to arrange a factorial replicate in a row-column array. Here attention is not restricted to factors at two levels. Choi and Gupta (2008) produce several designs by selecting subsets of factorial effects to be confounded with blocks. For $n \leq 9$, Dash *et al.* (2013) give designs for s replicates of a 2^n factorial in a $2 \times 2^{sn-1}$ array that provide estimates of all main effects and all two factor interactions. Cheng and Tsai (2013) build on work of Patterson and Bailey (1978) to give the structure for a design key for a 2^n factorial arranged in a rectangular array. Features of the design key necessary for estimation of main effects are highlighted. Wang (2017) gives constructions for a single replicate of a 2^n factorial in a rectangular array which enable estimation of all main effects. Guidance is given for construction of multiple array designs to obtain all two factor interactions. These require a relatively large number of replicates.

In this work, constructions are developed for designs comprising 2^n factorials arranged in $2^p \times 2^q$ rectangular arrays with $p + q \geq n$, from which all main effects and the maximum number of two factor interactions are estimable. It is assumed that both forms of blocking are incomplete, that is, $n < \max\{p, q\}$. A systematic construction approach uses generating matrices, which essentially contain generating sets for two Abelian groups. These generating matrices are design keys for producing rectangular arrays containing 2^{p+q-n} replicates of a 2^n factorial. The approach used places the emphasis on which main effects and two factor interactions are estimable, rather than on the selection of specific interactions to be confounded with rows and columns. The work expands on that of Cheng and Tsai (2013), providing a body of templates for generating matrices to give designs covering p, q, n combinations likely to be of practical use.

The paper is structured with notation and preliminary work on generating matrices in §2. In §3 constructions are developed for single replicate arrays which provide estimates of all main effects and of the largest possible number of two factor interactions. §4 covers designs comprising multiple arrays each containing a single replicate. In those cases where not all main effects and two factor interactions can be estimated from one array, two or more arrays are used to produce a design in which each array contains some confounded effects of interest, with every required effect being estimable from at least one array. Arrays containing more than one replicate are investigated in §5. A brief investigation into constructions involving fractional factorials is contained in §6. There is some flexibility in many of the the constructions, enabling practitioners to tailor a design to give greater emphasis to specific effects. Also, guidance is given on adaptation of generating matrices to accommodate the situation in which only a subset of the two factor interactions are of interest. Comparisons are made between designs obtained by the generator matrices developed and those available elsewhere in the literature.

2. Preliminaries

Basic notation and concepts are consistent with Chapters 7-9 and 13 of Hinkelmann and Kempthorne (2005). The factors of a 2^n experiment are denoted $A_1, A_2, \dots, A_n$. Factor A_i has levels $x_i \in \{0, 1\}$. If $x_i = 0$, then A_i is described as being at *low* level

and, if $x_i = 1$, A_i is described as being at *high* level. In examples where n is given, the factors are labelled $A, B, \dots$ for simplicity, rather than $A_1, A_2, \dots$.

A treatment combination is represented as $a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n}$, where $a_i^0 = 1$ and $a_i^1 = a_i$, and equivalently as the vector $(x_1, x_2, \dots, x_n)$. By convention, $a_1^{x_1} a_2^{x_2} \cdots a_n^{x_n}$ also denotes the true response. The treatment combination with all factors low is given by (1) and $(0, 0, \dots, 0)$. A replicate involves a run in each of the 2^n treatment combinations. The vector representations of the treatment combinations are exactly the 2^n vectors in n -dimensional Euclidean space over $GF(2)$, the Galois Field of order 2. Two treatment combinations $(x_{1,1}, \dots, x_{1,n})$ and $(x_{2,1}, \dots, x_{2,n})$, say, can be *added* to give $(x_{1,1} + x_{2,1}, \dots, x_{1,n} + x_{2,n})$.

A factorial effect is a contrast which partitions the treatment combinations into two sets of size 2^{n-1} . Combinations in the first set have an odd number of factors of the effect at low level and combinations in the second set have an even number at low level. The factorial effect is the subtraction of the average of the true responses of the first set from the average of the true responses of the second. For example, for a 2^3 factorial, with factors A, B, C , the linear combinations corresponding to the main effect of A , also denoted A , and the interaction between A and B , denoted AB , are:

$$\begin{aligned} A &= \frac{1}{4}[-(1) + a - b + ab - c + ac - bc + abc] \\ AB &= \frac{1}{4}[(1) - a - b + ab + c - ac - bc + abc]. \end{aligned}$$

The $2^n - 1$ factorial effects comprise a set of mutually orthogonal contrasts in the 2^n true responses. This work concerns estimation of the n main effects and $n(n-1)/2$ two factor interactions. Unless otherwise stated, interactions of three or more factors will be assumed negligible and, for brevity, *interaction* will be taken to mean a two factor interaction.

2.1. Generating sets

Consider a set of treatment combinations, $\mathcal{T} = \{\tau_1, \dots, \tau_t\}$, not necessarily all distinct, where $\tau_i = (x_{i,1}, \dots, x_{i,n})$. Adding members of all 2^t subsets of $\mathcal{T}$ generates 2^t treatment combinations, again not necessarily distinct, each of form $(\sum_{i=1}^t \phi_i x_{i,1}, \dots, \sum_{i=1}^t \phi_i x_{i,n})$, where $\phi_i \in \{0, 1\}$ and calculations are conducted in $GF(2)$. The set $\mathcal{T}$ can be summarised by a $t \times n$ generating matrix:

$$G = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ x_{t,1} & x_{t,2} & \cdots & x_{t,n} \end{pmatrix}.$$

The distinct treatment combinations generated form an Abelian group with cardinality 2^ρ , where $\rho = \text{Rank}(G)$. In particular, for $t \geq n$, the 2^t treatment combinations comprise 2^{t-n} complete replicates of the 2^n factorial *iff* $\rho = n$. To illustrate these properties, we consider two sets of four treatment combinations from a 2^3 factorial.

Set I $\tau_1 = (1, 1, 0)$, $\tau_2 = (1, 1, 0)$, $\tau_3 = (1, 1, 1)$, $\tau_4 = (0, 1, 1)$.

Set II $\tau_1 = (1, 1, 0)$, $\tau_2 = (1, 1, 0)$, $\tau_3 = (1, 1, 1)$, $\tau_4 = (0, 0, 1)$.

We denote the generating matrices for the sets by G_I and G_{II} respectively,

$$G_I = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad G_{II} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\text{Rank}(G_I) = 3$ and $\text{Rank}(G_{II}) = 2$. Thus, Set I generates two complete replicates of the 2^3 factorial. The 16 treatment combinations generated by Set II constitute four copies of the principal half replicate with defining contrast AB , namely four copies of each of $(1), ab, abc$ and c .

2.2. Row-column designs

Treatment combinations of a 2^n factorial are allocated to the 2^{p+q} experimental units of a $2^p \times 2^q$ array, where p, q are positive integers with $p + q \geq n$, to give a row-column design. Without loss of generality, $p \leq q$. It will be assumed throughout that $q < n$, so both forms of blocking are necessarily incomplete. The single array model is:

$$y_{ijk} = \mu + \tau_i + \rho_j + \gamma_k + \epsilon_{ijk}. \quad (2.1)$$

Here, y_{ijk} is the observation resulting from application of the i th treatment combination to the experimental unit in the j th row and k th column. The overall mean is μ and τ_i , ρ_j and γ_k are the effects of the i th treatment combination, the j th row and the k th column. The error terms ϵ_{ijk} are assumed to be uncorrelated, all with variance σ^2 .

In the constructions developed, treatment combinations are allocated to experimental units by means of generating sets. The first column and first row of an array will be termed the *principal column* and *principal row*. These contain the treatment combinations generated by sets of p and q treatment combinations with generating matrices $\mathcal{G}_c$ and $\mathcal{G}_r$. Treatment combination (1) is allocated to the experimental unit in the intersection of the principal column and row, that is, in the first column of the first row. The remaining treatment combinations from $\mathcal{G}_c$ and $\mathcal{G}_r$ are allocated to the principal column and principal row in any order. For $2 \leq j \leq 2^p$ and $2 \leq k \leq 2^q$, the treatment combination in row j and column k of the array is obtained by adding the vectors for treatment combinations in row j and column k of the principal column and row. In every case, once the array has been constructed, rows and columns are randomised before the design is used.

With this construction process, the 2^{p+q} entries in the array are exactly those generated by the set of $p + q$ treatment combinations in $\mathcal{G}_c$ and $\mathcal{G}_r$. The array can be represented by a partitioned matrix $\mathcal{G}$ of order $(p + q) \times n$, termed an *array generator*

matrix (AGM):

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \begin{pmatrix} x_{1,1} & x_{1,2} & \cdots & x_{1,n} \\ x_{2,1} & x_{2,2} & \cdots & x_{2,n} \\ \vdots & \vdots & & \vdots \\ x_{p,1} & x_{p,2} & \cdots & x_{p,n} \\ x_{p+1,1} & x_{p+1,2} & \cdots & x_{p+1,n} \\ x_{p+2,1} & x_{p+2,2} & \cdots & x_{p+2,n} \\ \vdots & \vdots & & \vdots \\ x_{p+q,1} & x_{p+q,2} & \cdots & x_{p+q,n} \end{pmatrix}.$$

2.3. Estimability conditions on $\mathcal{G}$

From discussion in §2.1 and §2.2, the $2^p \times 2^q$ array obtained from $\mathcal{G}$ contains 2^{p+q-n} replicates of the 2^n factorial *iff* $\text{rank}(\mathcal{G}) = n$. It is desirable for a design to be binary with regards to both blocking factors. This occurs *iff* the 2^p treatment combinations of the principal column are distinct and the 2^q treatment combinations of the principal row are distinct. The condition on $\mathcal{G}$ to achieve these properties is:

Condition 1: The AGM has $\text{Rank}(\mathcal{G}) = n$, $\text{Rank}(\mathcal{G}_c) = p$ and $\text{Rank}(\mathcal{G}_r) = q$.

Unless stated otherwise, all matrix manipulations are conducted in GF(2). The rank of an AGM, or sub-matrix, is determined by reduction of the matrix to row echelon form, working modulo 2.

We now examine properties of three designs, each for a replicate of a 2^5 factorial in a 4×8 array, generated using different AGMs. The purpose is to highlight AGM features which relate to estimability of main effects and interactions. These features will inform a construction approach for arrays with various p, q, n combinations.

EXAMPLE 1. Designs $\mathcal{D}1$, $\mathcal{D}2$ and $\mathcal{D}3$ are constructed from AGMs $\mathcal{G}_1$, $\mathcal{G}_2$ and $\mathcal{G}_3$:

$$\mathcal{G}_1 = \left(\frac{\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{pmatrix}}{\begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}} \right) \quad \mathcal{G}_2 = \left(\frac{\begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}}{\begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}} \right) \quad \mathcal{G}_3 = \left(\frac{\begin{pmatrix} 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}}{\begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}} \right).$$

The AGMs are all full rank. Therefore, each satisfies Condition 1 and yields a complete replicate of the 2^5 factorial in a $2^2 \times 2^3$ array. From the first two rows of $\mathcal{G}_1$, the principal column of $\mathcal{D}1$ is generated by $abcd$ and abe and so contains treatment combinations (1), $abcd$, abe and cde . Likewise, from the last three rows of $\mathcal{G}_1$, the principal row contains (1), acd , bce , $abde$, be , $abcde$, c and ad , giving the array:

$$\mathcal{D}1 = \begin{array}{cccccccc} (1) & acd & bce & abde & be & abcde & c & ad \\ abcd & b & ade & ce & acde & e & abd & bc \\ abe & bcde & ac & d & a & cd & abce & bde \\ cde & ae & bd & abc & bcd & ab & de & ace \end{array}$$

Designs $\mathcal{D}2$ and $\mathcal{D}3$ are obtained from $\mathcal{G}_2$ and $\mathcal{G}_3$ in the same way.

$$\mathcal{D}2 = \begin{array}{cccccccc} (1) & ace & bcd & abde & bce & ab & de & acd \\ abcd & bde & a & ce & ade & cd & abce & b \\ abe & bc & acde & d & ac & e & abd & bcde \\ cde & ad & be & abc & bd & abcde & c & ae \end{array}$$

$$\mathcal{D}3 = \begin{array}{cccccccc} (1) & ac & abd & bcd & bce & abe & acde & de \\ ade & cde & be & abce & abcd & bd & c & a \\ bde & abcde & ae & ce & cd & ad & abc & b \\ ab & bc & d & acd & ace & e & bcde & abde \end{array}$$

Inspecting the design array for $\mathcal{D}1$, each factor is high in exactly half the treatment combinations in each row and column. Thus all main effects are estimable. Next consider the interaction AB . Half the treatment combinations in each row have an even number of A, B high. However, each column either has all treatment combinations with an even number of A, B high or all with an odd number of A, B high. The same property is noted for CD . Thus AB and CD are confounded with columns. In a similar way, AD and BE are confounded with rows. The other interactions have half the treatment combinations in each row and each column with an even number of the two factors high. Thus AC , AE , BC , BD , CE and DE are estimable but AB , AD , BE and CD are not.

These properties can be deduced directly from $\mathcal{G}_1 = [\mathcal{G}_{1c}^T | \mathcal{G}_{1r}^T]^T$, where M^T denotes the transpose of matrix M . Each column of $\mathcal{G}_{1c}$ and of $\mathcal{G}_{1r}$ contains at least one element of unity, with consequence that each treatment is high in exactly half the experimental units in the principal column and in the principal row and therefore in each column and row of the array. Thus, in the linear combination of observations for a main effect, the row and column effects add out and the effect is estimable. Considering AB , the first and second columns of $\mathcal{G}_{1c}$ are the same resulting in all treatment combinations in the principal column having A, B both high or both low. All have positive sign in the AB effect. As the remaining columns are cosets of the principal column, each column has all treatment combinations with an even number of A, B high or all with an odd number of A, B high. Thus, all treatment combinations in a column have the same sign in the AB effect and AB is completely confounded with columns. The third and fourth columns of $\mathcal{G}_{1c}$ are also the same and CD is not estimable. Likewise, AD and BE are inestimable due to repeated columns in $\mathcal{G}_{1r}$. Conversely, consider AC . The corresponding columns of $\mathcal{G}_{1c}$ are different as are those of $\mathcal{G}_{1r}$. Thus in both the principal column and principal row, exactly half the treatment combinations have positive sign in the AC contrast. This property extends through all columns and rows, leading to estimability of the effect.

In $\mathcal{D}2$, all main effects are estimable since each column of $\mathcal{G}_{2c}$ and $\mathcal{G}_{2r}$ contains at least one element of unity. Sub-matrix $\mathcal{G}_{2c}$ is the same as $\mathcal{G}_{1c}$ and, by the same reasoning as above, AB and CD are not estimable. The columns of $\mathcal{G}_{2r}$ are all different and so all other interactions are estimable.

The notable feature of $\mathcal{G}_3$ is that the third column of $\mathcal{G}_{3c}$ contains only zeros, with the consequence that C is low in all treatment combinations in the principal column of $\mathcal{D}3$. Thus, C is confounded with columns and is not estimable. All other main effects are estimable as are all interactions except DE .

Properties highlighted in Example 1 prompt general conditions on AGMs relating to estimability of main effects and interactions. As seen in Example 1, an effect is either estimable with full information or is completely confounded with blocks and is inestimable. Some notation is introduced to facilitate statement of the conditions. Let X_U be the set of non-zero vectors of length U over $GF(2)$. Then $|X_U| = 2^U - 1$. For example:

$$X_3 = \left\{ \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Condition 1 can be supplemented by a second condition:

Condition 2A: Each column of $\mathcal{G}_c$ ($\mathcal{G}_r$) is a vector of X_p (X_q).

Together, Conditions 1 and 2A are necessary and sufficient for $\mathcal{G}$ to be the AGM for a $2^p \times 2^q$ array containing 2^{p+q-n} replicates of a 2^n factorial, from which all main effects are estimable. Condition 2A is given in Cheng and Tsai (2013).

Designs $\mathcal{D}1$ and $\mathcal{D}2$ of Example 1 have AGMs which satisfy Conditions 1 and 2A. These designs have different properties with respect to estimation of interactions and we seek an alternative to Condition 2A which, with Condition 1, relates to estimation of all main effects and to maximising the number of estimable two factor interactions. From discussion in Example 1, for an AGM satisfying Condition 1, an interaction is estimable *iff* the corresponding columns in $\mathcal{G}_c$ and $\mathcal{G}_r$ are distinct. Thus, a pair of common columns in $\mathcal{G}_c$ corresponds to an interaction that is not estimable. The set X_p has cardinality $2^p - 1$. The number of columns of $\mathcal{G}_c$ can be expressed uniquely as $n = \alpha(2^p - 1) + \beta$ where $\alpha = \lfloor n/(2^p - 1) \rfloor$, with $\lfloor \cdot \rfloor$ denoting the integer part of, and $\beta \in \{0, 1, \dots, 2^p - 2\}$. By manipulation of Binomial coefficients, an upper bound for the number of pairs of distinct vectors in a set of n vectors from X_p is:

$$\omega = \binom{n}{2} - \alpha\beta - (2^p - 1)\binom{\alpha}{2}. \quad (2.2)$$

This is also an upper bound for the number of estimable interactions from an array and reduces to $\binom{n}{2}$ when $n \leq 2^p - 1$. The bound of (2.2) prompts an alternative to Condition 2A, labelled Condition 2B, which concerns estimability of both main effects and interactions:

Condition 2B: $\mathcal{G}_c$ is formed from $\alpha + 1$ copies of each of β vectors of X_p and of α copies of each of the remaining $2^p - \beta - 1$ vectors, and the columns of $\mathcal{G}_r$ are distinct.

Together, Conditions 1 and 2B are necessary and sufficient for an AGM to yield an array from which all main effects and ω interactions are estimable. Both conditions are satisfied by $\mathcal{D}2$ of Example 1 but not by $\mathcal{D}1$. No arrangement of a 2^5 factorial in a $2^2 \times 2^3$ array can provide estimates of all main effects and of more than $\omega = 8$ interactions. In general, for a single replicate array with $q > 2^p - p - 1$, then $n = p + q > 2^p - 1$ and $\mathcal{G}_c$ will contain repeated columns. For such parameter combinations it is not possible to estimate all main effects and interactions from one array. Note that for $q \geq 3$, we have $n \leq 2q < 2^q - 1$, indicating that $|X_q| > n$, that is, the number of distinct vectors available for use in $\mathcal{G}_r$ exceeds n .

3. Single replicate arrays

We now investigate the arrangement of a single replicate of a 2^n factorial in a $2^p \times 2^q$ array, with $p + q = n$. The task is to give AGMs for different p, q combinations which generate arrays yielding estimates of all main effects and of ω interactions. The arrays can be used as single replicate designs, after randomisation of rows and columns, or combined to form multi-replicate designs, which are covered in §4 and §5.

3.1. Estimability of main effects

We first focus on main effects and give constructions for AGMs which satisfy Conditions 1 and 2A and thus yield single replicate arrays from which all main effects are estimable. There are two cases.

THEOREM 1. *For $p + q \geq 4$, an AGM of the following form yields a replicate of a 2^{p+q} factorial in a $2^p \times 2^q$ array enabling estimation of all main effects:*

Case 1: pq even

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\frac{I_p \mid J_{p,q}}{J_{q,p} \mid I_q} \right) \quad \text{and}$$

Case 2: pq odd or q even

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\frac{I_p \mid J_{p,q}}{J_{q,p} \mid I_q + J_{q,q}} \right), \quad (3.1)$$

where I_v and $J_{v,w}$ denote the $v \times v$ identity matrix and the $v \times w$ matrix with each term unity.

Proof Consider the $(p + q) \times (p + q)$ matrices

$$M_1 = \left(\frac{I_p \mid J_{p,q}}{0_{q,p} \mid I_q} \right)$$

and

$$M_2 = \left(\frac{I_p \mid J_{p,q}}{0_{q,p} \mid I_q + J_{q,q}} \right),$$

where $0_{v,w}$ denotes the $v \times w$ matrix, with all entries zero. All entries are in $GF(2)$. Matrix M_1 is full rank. For q even, M_2 has row reduced echelon form equal to M_1 , and is therefore also full rank.

Case 1 On adding the sum of rows $1, \dots, p$ to row i , for $i = p + 1, \dots, p + q$, working modulo 2, the matrix $\mathcal{G}$ is transformed to M_1 if p is even, and to M_2 if p is odd and q even. Thus, $\mathcal{G}$ has full rank.

Case 2 Likewise, on adding the sum of rows $1, \dots, p$ to row i , for $i = p + 1, \dots, p + q$, the matrix $\mathcal{G}$ is transformed to M_1 if p is odd, and to M_2 if p and q are both even. Thus, $\mathcal{G}$ has full rank.

In both cases $\mathcal{G}$ has full rank, satisfying Condition 1. Further, in both cases it is evident that Condition 2A is satisfied, and the result follows.

Case 1 covers all p, q pairs except those with p, q both odd, and Case 2 covers all pairs except those with p even and q odd. No constructions are given for $p + q \leq 3$ since no degrees of freedom would be available for estimation of σ^2 . The constructions of Theorem 1 have similarities to those given by Wang (2016). Case 2 always applies for $p = 1$. The AGM for $p = 1$ has relevance later and is covered now as a Corollary.

COROLLARY 1. *For $p = 1$ and $q \geq 3$, an AGM which arranges a single replicate of a 2^{q+1} factorial in a 2×2^q array from all main effects are estimable is:*

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\frac{1}{1_q} \middle| \frac{1_q^T}{I_q + J_{q,q}} \right). \quad (3.2)$$

EXAMPLE 2. *Examples of AGMs for a single replicate of a 2^5 factorial from which all main effects are estimable with (i) $p = 2, q = 3$ and (ii) $p = 1, q = 4$ are:*

$$(i) \left(\begin{array}{cc|ccc} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{array} \right) \quad \text{and} \quad (ii) \left(\begin{array}{c|cccc} 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{array} \right).$$

In many cases the arrays of Theorem 1 can be improved on if interactions are also of interest. To demonstrate this, denote the design corresponding to the array of Example 2 (i) by $\mathcal{D}4$. In common with $\mathcal{D}1$ and $\mathcal{D}2$ of Example 1, $\mathcal{D}4$ has $p = 2, q = 3$ and satisfies Conditions 1 and 2A. Of the three designs, $\mathcal{D}2$ is the only one which satisfies Condition 2B and so achieves the bound of (2.2) with eight estimable interactions. Designs $\mathcal{D}1$ and $\mathcal{D}4$ only provide estimates of six interactions.

We now develop AGM constructions which satisfy Conditions 1 and 2B and therefore yield arrays enabling estimation of all main effects and of ω interactions. AGMs are given for all p, q with $p + q = n$, where $p + q \geq 4$, except for the case $p = q = 2$ which is considered separately.

3.2. Constructions for $3 \leq p \leq q$

LEMMA 1. *For $p \geq 3$, let $K_p = (k_{i,j})$ be the $p \times p$ matrix with $k_{i,j} = 1$ if $i \leq j$ and $k_{i,j} = 0$ if $i > j$, and let $L_p = (l_{i,j}) = K_p + K_p J_{p,p} + I_p$. Then the columns of the $p \times 2p$ matrix $(K_p L_p)$ are $2p$ distinct columns from X_p .*

Proof The j th column of K_p has first j elements unity and last $p - j$ elements zero. Thus the columns of K_p are all distinct and all in X_p . Now consider L_p . By construction, we have $l_{i,j} = p + 1 - i$ if $i \geq j$ and $l_{i,j} = p + 2 - i$ if $i < j$. Therefore the columns of L_p are distinct and, since $l_{p,1} = l_{p,2} = \dots = l_{p,p} = 1$, each is a column of X_p . For p odd, $l_{1,j} = 0$, for $j \geq 2$, and $l_{2,1} = 0$. For p even, $l_{2,j} = 0$, for $j \geq 3$, and $l_{3,1} = l_{3,2} = l_{3,3} = 0$. Therefore, each column of L_p contains at least one element zero which comes above an element of unity and so none is a column of K_p . Thus, the columns of $(K_p L_p)$ are distinct columns of X_p as required.

The condition $p \geq 3$ is needed in Lemma 1 since the columns of $(K_p L_p)$ are not distinct for $p \leq 2$.

THEOREM 2. For $3 \leq p \leq q$, an AGM on the following form arranges a replicate of a 2^{p+q} factorial in a $2^p \times 2^q$ array, from which all main effects and ω interactions are estimable:

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\begin{array}{c|c|c} I_p & I_p + J_{p,p} & X_* \\ \hline K_p & L_p & 0_{p,q-p} \\ \hline 0_{q-p,p} & 0_{q-p,p} & I_{q-p} \end{array} \right),$$

where X_* is a $p \times (q-p)$ matrix such that each column of X_p appears in the $p \times n$ matrix $(I_p \ I_p + J_{p,p} \ X_*)$, α or $\alpha + 1$ times, where $\alpha = [(q+2)/(2^p - 1)]$.

Proof Reducing $\mathcal{G}$ to row echelon form, working modulo 2, gives

$$\left(\begin{array}{c|c|c} I_p & I_p + J_{p,p} & X_* \\ \hline 0_{p,p} & I_p & K_p X_* \\ \hline 0_{q-p,p} & 0_{q-p,p} & I_{q-p} \end{array} \right),$$

which confirms that $\mathcal{G}$ is full rank, satisfying Condition 1. By the construction of X_* , every column of $\mathcal{G}_c$ is in X_p and the columns have multiplicities consistent with Condition 2B. By Lemma 1, the first $2p$ columns of $\mathcal{G}_r$ are distinct columns of X_q . Further, the final $q-p$ columns are also distinct columns of X_q and are different to any of the first $2p$ columns. Thus Condition 2B is satisfied which establishes that all main effects and ω interactions are estimable from the array generated by $\mathcal{G}$.

EXAMPLE 3. Examples of AGMs for single replicate arrays from which all main effects and ω interactions are estimable are given for (i) $p = 3, q = 3$ and (ii) $p = 3, q = 6$:

$$(i) \left(\begin{array}{c|c|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ \hline 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{array} \middle| \begin{array}{c|c|c} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \hline 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \end{array} \right) \text{ and } (ii) \left(\begin{array}{c|c|c|c|c|c} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ \hline 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right).$$

All effects of interest are estimable from (i). By comparison, for $p = q = 3$, the construction of Wang (2017) gives an array with six inestimable interactions, and requires three single replicate arrays to achieve estimability of all main effects and interactions. From (ii), all effects of interest bar AH and BI are estimable.

Due to the reliance of Theorem 2 on Lemma 1, the construction does not apply to arrays with $p < 3$. Separate constructions are required for $p = 2$ and $p = 1$.

3.3. Constructions for $p = 2, q \geq 3$

No array comprising a single replicate of the 2^{q+2} factorial will give estimates of all main effects and interactions because there are $q+2 \geq 5$ columns in $\mathcal{G}$, but only three

vectors in X_2 . The following result gives constructions which enable estimation of all main effects and maximise the number of estimable interactions.

THEOREM 3. *For $q \geq 3$, an AGM of the following form yields a replicate of a 2^{q+2} factorial in a $2^2 \times 2^q$ array from which all main effects and ω interactions are estimable:*

$$\mathcal{G} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\begin{array}{c|c|c} I_2 & F & X_* \\ \hline E & I_3 + EF & 0_{3,q-3} \\ \hline 0_{q-3,2} & 0_{q-3,3} & I_{q-3} \end{array} \right), \quad (3.3)$$

where E and F are the matrices

$$E = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \quad F = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix},$$

and X_* is a $2 \times (q-3)$ matrix selected so that each vector of X_2 appears in the $2 \times n$ matrix $(I_2 \ F \ X_*)$, α or $\alpha + 1$ times, where $\alpha = [(q+2)/3]$.

Proof Reducing $\mathcal{G}$ to row echelon form, gives

$$\left(\begin{array}{c|c|c} I_2 & F & X_* \\ \hline 0_{3,2} & I_3 & EX_* \\ \hline 0_{q-3,2} & 0_{q-3,3} & I_{q-3} \end{array} \right),$$

which confirms that $\mathcal{G}$ is full rank, satisfying Condition 1. By the selection of columns of X_* , every column of $\mathcal{G}_c$ is a vector of X_2 and the multiplicities are consistent with Condition 2B. Now consider the 3×5 top left hand submatrix of $\mathcal{G}_r$:

$$\left(\begin{array}{c|c} E & I_3 + EF \end{array} \right) = \left(\begin{array}{cc|cc} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \end{array} \right).$$

This submatrix comprises distinct vectors of X_3 . Thus, the columns of $\mathcal{G}_r$ are distinct vectors of X_q . It follows that Condition 2B is satisfied which establishes the result.

EXAMPLE 4. *An AGM for a replicate of the 2^6 factorial arranged in a $2^2 \times 2^4$ array from which all main effects and ω interactions are estimable is:*

$$\mathcal{G} = \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right). \quad (3.4)$$

Of the fifteen interactions, all are estimable except AE , BD and CF .

Arrays from Theorem 3 improve on arrays and constructions elsewhere in the literature, with regards to estimation of interactions. For $q = 3$, the array of Theorem 3 has two confounded interactions. With the same parameters, the construction of Wang (2017) and arrays in Choi and Gupta (2008) each have three or more confounded interactions. However, it is noted that Choi and Gupta (2008) do not restrict attention to interactions of only two factors.

3.4. Construction for $p = 2, q = 2$

Theorem 3 does not apply to $p = q = 2$. In this case Condition 2B cannot be satisfied since the number of columns of $\mathcal{G}_r$ exceeds the cardinality of X_2 . Therefore there is no arrangement of a replicate of the 2^4 factorial in a 4×4 array which enables estimation of all main effects and of $\omega = 5$ interactions. The two arrays from Theorem 1 enable estimation of all main effects and of four of the six interactions and cannot be improved on. The AGM obtained from (3.1) is

$$\mathcal{G} = \left(\begin{array}{cc|cc} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{array} \right). \quad (3.5)$$

In practice, a design consisting solely of this array would not be recommended for experimentation since only one degree of freedom is available for estimating σ^2 . A two replicate design arranged in two 4×4 arrays is considered in §4.

Wang (2017) and Choi and Gupta (2008) also give single replicate designs with $p = q = 2$. Both designs enable estimation of all main effects. The design of Wang (2017) is equivalent to (3.5) and gives estimates of four interactions, but only three interactions are estimable from the design of Choi and Gupta (2008).

3.5. Constructions for $p = 1, q \geq 3$

Corollary 1 yields a single replicate 2×2^q array from which all main effects are estimable. Technically, the AGM of Corollary 1 satisfies Condition 2B, but since the upper bound for the number of estimable interactions given by (2.2) is $\omega = 0$, no interactions are estimable from the array. This is readily seen from (3.2), since every column of $\mathcal{G}_c$ is identical. Single replicate arrays can be constructed enabling estimation of some interactions but these arrays do not enable estimation of all main effects. This is investigated in §4.

Theorems 2, 3 and Corollary 1 give single replicate arrays for all parameter combinations with $1 \leq p \leq q$ where $q \geq 3$. For parameter combinations with $3 \leq p \leq q$ and $p + q \leq 2^p - 1$, the arrays of Theorem 2 enable estimation of all main effects and interactions. Arrays for all the other parameter combinations enable estimation of all main effects and of ω interactions. In general, the arrays of Theorems 2 and 3 have advantages over constructions of Wang (2017) which have $\{p(p-1) + q(q-1)\}/2$ inestimable interactions.

For cases where not all interactions are estimable or where, as for the design generated by (3.5), the number of degrees of freedom available for estimation of σ^2 is very small, multi-replicate designs are advised. These are investigated in §4 and §5.

3.6. Bespoke single replicate arrays

There are situations with $p \geq 2$ and $q \geq 3$ in which an array providing estimates of fewer than ω interactions may be preferable to one from Theorem 2 or Theorem 3. Retaining the AGM structures of Theorems 2 and 3 and the requirement that columns of X_* are vectors of X_p , but relaxing the requirement regarding selection of X_* to achieve as near

equal as possible multiplicities of columns in $\mathcal{G}_c$ can reduce ω , but may be beneficial for a given experimental situation.

EXAMPLE 5. *To investigate an industrial process, six factors are identified for experimentation. A design comprising a replicate of a 2^6 factorial arranged in a $2^2 \times 2^4$ array is required. Due to the nature of the factors and process it can be assumed that there are no interactions between a subset of three factors and no interaction between a further two factors.*

The design given by Theorem 3 has $X_* = (1, 1)^T$ and yields an array in which all main effects and 12 of the 15 interactions are estimable, with AE, BD and CF being inestimable. In allocating factors to labels, at least one interaction of interest will be inestimable. If $X_* = (1, 0)^T$ is used in (3.3), then only 11 interactions are estimable: the inestimable interactions being AE, AF, EF and BD . Allocating the three factors that do not interact to A, E, F and the further two factors that do not interact to B, D , gives an array from which all effects of interest are obtained.

4. Multiple single replicate arrays

We now consider designs comprising a number of arrays each containing a single replicate. For an m replicate 2^n factorial design, each replicate is arranged in a separate $2^p \times 2^q$ array where $p + q = n$. The model given by (2.1) is adjusted to incorporate replicates as an additional blocking factor:

$$y_{ijkl} = \mu + \tau_i + \rho_{jl} + \gamma_{kl} + \alpha_l + \epsilon_{ijkl},$$

where, y_{ijkl} is the observation on application of the i th treatment combination to the experimental unit in the j th row and k th column of the l th replicate and α_l is the effect of the l th replicate. Other terms are obvious extensions of terms in (2.1), with γ_{kl} and ρ_{kl} being the effects of the j th row and the k th column in the l th replicate.

4.1. Multiple array designs for $p \geq 2, q \geq 3$

For p, q combinations with both $p \geq 2$ and $q \geq 3$, individual replicates are constructed using results from §3. The AGM for the i th replicate is denoted by $\mathcal{G}^i$, with submatrices $\mathcal{G}_c^i$ and $\mathcal{G}_r^i$. If $p \geq 3$ and $q \leq 2^p - p - 1$, the construction of Theorem 2 gives replicates which each provide estimates of all effects of interest. Otherwise, that is if $p \geq 3$ and $q > 2^p - p - 1$ or if $p = 2$ and $q \geq 3$, whilst every replicate yields estimates of all main effects, no single replicate enables estimation of all interactions. The columns of generator matrices from Theorems 2 and 3 are rearranged to give AGMs for a sufficient number of individual replicates, so that the resulting arrays together provide estimates of all interactions. Main effects are estimated with full efficiency and some interactions are partially confounded. This is now demonstrated.

EXAMPLE 6. *A two replicate design is required for a 2^6 factorial with each replicate arranged in a $2^2 \times 2^4$ array.*

The AGM of (3.4) is used as $\mathcal{G}^1$. This provides estimates of all main effects and all interactions except AE , BD and CF . The columns of $\mathcal{G}^1$ are rearranged to give $\mathcal{G}^2$:

$$\mathcal{G}^2 = \left(\frac{\mathcal{G}_c^2}{\mathcal{G}_r^2} \right) = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \quad (4.1)$$

No vertical construction lines are included in (4.1), since $\mathcal{G}^2$ is a rearrangement of an existing AGM. The second replicate provides estimates of all effects of interest except AF , BE and CD . Thus, the design comprising both arrays enables estimation of all main effects and nine interactions with full information. The remaining six interactions are partially confounded and are estimated with 1/2 relative information.

For an array of Theorem 2 or Theorem 3, any non-estimability of interactions occurs as a consequence of repetition of columns in $\mathcal{G}_c$. Stacking the sub-matrices $\mathcal{G}_c^1$ and $\mathcal{G}_c^2$ of Example 6 demonstrates the property that results in full estimability.

$$\left(\frac{\mathcal{G}_c^1}{\mathcal{G}_c^2} \right) = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

Although both sub-matrices necessarily have repeated columns, the 4×6 stacked matrix has no repeated columns, indicating that each interaction is estimable in at least one replicate. In general, for an m replicate design we define $\mathcal{S}_c$ and $\mathcal{S}_r$ to be the $mp \times n$ and $mq \times n$ matrices formed by stacking the AGM sub-matrices:

$$\mathcal{S}_c = \begin{pmatrix} \frac{\mathcal{G}_c^1}{\mathcal{G}_c^2} \\ \vdots \\ \frac{\mathcal{G}_c^m}{\mathcal{G}_c^m} \end{pmatrix}, \quad \mathcal{S}_r = \begin{pmatrix} \frac{\mathcal{G}_r^1}{\mathcal{G}_r^2} \\ \vdots \\ \frac{\mathcal{G}_r^m}{\mathcal{G}_r^m} \end{pmatrix}.$$

Let the AGMs for m replicates be obtained from Theorem 2 or 3, possibly with rearranged columns. Then, columns of $\mathcal{G}_r^i$ comprise n distinct vectors from X_q and columns of $\mathcal{G}_c^i$ are vectors from X_p . Every main effect is estimable from every replicate. An interaction is estimable from at least one replicate *iff* the corresponding columns of $\mathcal{S}_c$ are different. Further, denoting the i th column of $\mathcal{G}_c^k$ by g_{ci}^k , the interaction $A_i A_j$ is estimable from

$$m_{i,j} = m - \sum_{k=1}^m \left[\frac{p - (g_{ci}^k - g_{cj}^k)^T (g_{ci}^k - g_{cj}^k)}{p} \right] \quad (4.2)$$

replicates, where $[.]$ denotes *the integer part of*. Calculations in (4.2) are evaluated in $\mathbb{R}$ and $\mathbb{R}^p$ and not in $GF(2)$. The interaction $A_i A_j$ is estimated with $m_{i,j}/m$ relative information.

The set X_p has cardinality $2^p - 1$ and therefore there are $(2^p - 1)^m$ different possibilities for columns of $\mathcal{S}_c$, each obtained by stacking m , not necessarily distinct, vectors from X_p . The bound of (2.2) can be extended to give an upper bound for the number of interactions estimable from a set of m arrays. The number of columns of $\mathcal{S}_c$ can be expressed as $n = \alpha_m(2^p - 1)^m + \beta_m$ where $\alpha_m = \lfloor n/(2^p - 1)^m \rfloor$ and $\beta_m \in \{0, 1, \dots, (2^p - 1)^m - 1\}$. Then, an upper bound for the number of pairs of distinct columns from the columns of $\mathcal{S}_c$, and hence for the number of interactions estimable from at least one array is:

$$\omega_m = \binom{n}{2} - \alpha_m \beta_m - (2^p - 1)^m \binom{\alpha_m}{2}. \quad (4.3)$$

Note that (4.3) reduces to (2.2) for $m = 1$. The bound leads to:

PROPOSITION 1. *If $p \geq 2$, $q \geq 3$, a design comprising $\lceil \log_{2^p-1} n \rceil$ single replicate arrays exists such that all main effects are estimable from every array and each interaction is estimable from at least one array.*

By Proposition 1, for $p \geq 3$, designs can be constructed in two replicates which enable estimation of all effects of interest, for $n \leq 49$. Likewise, for $p = 2$, $q \geq 3$, two and three replicates are sufficient for $n \leq 9$ and $n \leq 27$, respectively. In practice, it is straightforward to construct designs in $\lceil \log_{2^p-1} n \rceil$ arrays with estimation properties consistent with Proposition 1. For small n , it may be desirable to use a design with more than $\lceil \log_{2^p-1} n \rceil$ single replicate arrays, in order to increase the degrees of freedom available for estimation of σ^2 .

EXAMPLE 7. *A multi-replicate design for a 2^8 factorial is required, with each replicate arranged as a $2^2 \times 2^6$ array.*

By Proposition 1, a design can be constructed in two single replicate arrays, such that all main effects are estimable from both arrays and every interaction is estimable from at least one array. Using Theorem 3, a possible $\mathcal{G}_c^1$ is:

$$\mathcal{G}_c^1 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

Rearranging the columns of $\mathcal{G}_c^1$ in the order 1, 2, 4, 3, 8, 5, 6, 7 gives a $\mathcal{G}_c^2$, with the submatrices stacking as:

$$\mathcal{S}_c = \begin{pmatrix} \mathcal{G}_c^1 \\ \mathcal{G}_c^2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

All columns of $\mathcal{S}_c$ are distinct. The second replicate is constructed with $\mathcal{G}^2$ formed by arranging the columns of $\mathcal{G}^1$ in the order given above. All effects of interest except AE , AG , EG , BD , BH , DH , CF are estimable from the first replicate. The non-estimable interactions from the second replicate are AF , AH , FH , BC , BE , CE , DG . Thus, the two replicate design provides estimates of all main effects and of 14 interactions with full information and of the remaining 14 interactions with 1/2 relative information.

Proposition 1 and the approach of Example 7 can be used to construct designs comprising multiple single replicate arrays for all p, q pairs except for $p = q = 2$ and $p = 1$. These cases are now considered.

4.2. Multiple array design for $p = q = 2$

For $p = q = 2$, the sub-matrix $\mathcal{G}_r$ for a single replicate array will not have all columns distinct. Therefore, attention needs to be given to both $\mathcal{S}_c$ and $\mathcal{S}_r$ when constructing a multi-replicate design. We use the generator matrix given in (3.5) for $\mathcal{G}^1$. This first replicate provides estimates of all required effects except AB and CD . For a second replicate, switch columns 2 and 3 of $\mathcal{G}^1$ to give

$$\mathcal{G}^2 = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}, \quad \mathcal{S}_c = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix}, \quad \mathcal{S}_r = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}.$$

The second replicate gives estimates of all effects of interest except AC and BD . Thus, a design comprising these replicates has all main effects and two interactions estimable with full efficiency and the remaining four interactions estimable with $1/2$ relative information. Note that $\mathcal{S}_c$ and $\mathcal{S}_r$ each contain four distinct vectors from X_4 . Of the 31 degrees of freedom, eight are available for estimating σ^2 . The properties of the two replicate design can be contrasted with those of the single replicate design with $p = q = 2$ in §3.4.

4.3. Multiple array design for $p = 1, q \geq 3$

For single replicate arrays with $p = 1$, the only $\mathcal{G}_c$ sub-matrix that satisfies Condition 2A is $\mathcal{G}_c = 1_n^T$. This is reflected in the construction of Corollary 1. An array produced by this construction gives estimates of all main effects but of no interactions. In this work, estimation of main effects is taken to have greater priority than estimation of interactions and up to this point all constructions have involved arrays from which all main effects are estimable. However, in the case of a multi-replicate design with $p = 1$, in order to estimate any interactions it is necessary to use some arrays which do not satisfy Condition 2A. With this adjustment, single replicate arrays are used which allow estimation of a subset of the main effects and of the interactions. With careful selection, a design comprising a set of arrays will give estimates of all effects of interest.

LEMMA 2. *For $q \geq 3$, an AGM of the following form yields a single replicate of a 2^{q+1} factorial in a 2×2^q array from which $q_0 + 1$ main effects and $(q_0 + 1)(q - q_0)$ interactions are estimable, where $q_0 \in \{0, 1, \dots, q - 1\}$:*

$$\mathcal{G} = \begin{pmatrix} \mathcal{G}_c \\ \mathcal{G}_r \end{pmatrix} = \begin{pmatrix} 1 & 1_{q_0}^T & 0_{q-q_0} \\ \hline 1_{q_0} & I_{q_0} + J_{q_0, q_0} & 0_{q_0, q-q_0} \\ \hline 1_{q-q_0} & J_{q-q_0, q_0} & I_{q-q_0} \end{pmatrix}. \quad (4.4)$$

The proof is along the same lines as those of Theorems 2 and 3 and is not included. The condition $q \geq 3$ ensures that the columns of $\mathcal{G}_r$ are distinct vectors of X_q and therefore that estimable effects of interest depend entirely on $\mathcal{G}_c$. Main effects $A_1, \dots, A_{q_0+1}$ are estimable as are all interactions involving one of these factors and one of $A_{q_0+2}, \dots, A_n$.

THEOREM 4. For $q \geq 3$, let $\mathbf{m} = \lceil \log_2(q+2) \rceil$ and let the $\mathbf{m} \times (q+1)$ matrix S_* be formed from $q+1$ distinct vectors of $X_{\mathbf{m}}$. Then S_* is the $\mathcal{S}_c$ matrix for a design in $\mathbf{m}$ replicates of a 2^{q+1} factorial, each arranged as a 2×2^q array, such that each main effect and interaction is estimable from at least one replicate.

Proof From the definition of $\mathbf{m}$, we have $2^{\mathbf{m}-1} - 1 < q+1 \leq 2^{\mathbf{m}} - 1$. Of the vectors in $X_{\mathbf{m}}$, $2^{\mathbf{m}-1}$ have the j th element unity and $2^{\mathbf{m}-1} - 1$ have j th element zero, for $j = 1, \dots, \mathbf{m}$. Therefore, in any set of $q+1$ distinct vectors of $X_{\mathbf{m}}$, at least one has entry unity in the j th position for $j = 1, \dots, \mathbf{m}$. Thus, each row of S_* contains at least one element of unity and is a $1 \times (q+1)$ vector corresponding to $\mathcal{G}_c$ for a single replicate array for a 2^{q+1} factorial formed from Corollary 1 or Lemma 2, possibly after rearrangement of columns. Let the i th row of S_* be $\mathcal{G}_c^i$. The $\mathcal{G}_r^i$ sub-matrix that completes $\mathcal{G}^i$ is obtained by rearranging the columns of the appropriate $\mathcal{G}_r$ sub-matrix of (3.2) or (4.4) so that $\mathcal{G}^i$ corresponds to (3.2) or (4.4) with columns rearranged.

Consider the design comprising $\mathbf{m}$ single replicate 2×2^q arrays with AGMs $\mathcal{G}^1, \dots, \mathcal{G}^{\mathbf{m}}$. The effects of interest estimable from the i th replicate are identified directly from $\mathcal{G}_c^i$. Since each column of S_* contains at least one element unity, each main effect is estimable from at least one replicate. Finally, since the columns of S_* are all different, every interaction is estimable from at least one replicate. The result follows.

Note that there is no requirement for S_* of Theorem 4 to be full rank.

EXAMPLE 8. A multi-replicate design for a 2^5 factorial is required, with each replicate arranged as a 2×2^4 array.

By Theorem 4, designs can be constructed in $\mathbf{m} = 3$ single replicate arrays. We start by selecting five distinct vectors of X_3 to form $\mathcal{S}_*$. For example,

$$\mathcal{S}_* = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix} \quad \text{giving} \quad \mathcal{S}_c = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}. \quad (4.5)$$

Using Lemma 2, and rearranging columns so that the first row of each AGM is consistent with the corresponding row of $\mathcal{S}_c$, gives:

$$\mathcal{G}^1 = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 1 \end{pmatrix} \quad \mathcal{G}^2 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \quad \mathcal{G}^3 = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \end{pmatrix}.$$

The main effect of E is estimable from each replicate, whereas A , B and C are each obtained from two replicates and D is only estimable from the second replicate. Each interaction is estimable from at least one replicate. For example, BD is only estimable from the first replicate whereas AD is estimable from all three replicates.

Unlike other constructions in this work, designs from Theorem 4 do not have all main effects estimable from every replicate. Indeed, at most one replicate will have $\mathcal{G}_c = 1_{q+1}^T$,

so at least $\mathbf{m} - 1$ replicates will not provide estimates for the full quota of $q + 1$ main effects. Properties of designs of Theorem 4 can be identified directly from S_* . Here, calculations are in $\mathbb{R}$ and $\mathbb{R}^m$. The i th element of $1_{\mathbf{m}}^T S_*$ is the number of replicates providing an estimate of A_i . Further, let the $(q + 1) \times (q + 1)$ matrix $T_* = (t_{ij})$ have $t_{ij} = (s_i - s_j)^T (s_i - s_j)$, where s_i is the i th column of S_* . Then t_{ij} is the number of replicates giving estimates of $A_i A_j$, for $i \neq j$. The numbers of main effect and interaction estimates from the $\mathbf{m}$ replicates are $1_{\mathbf{m}}^T S_* 1_{\mathbf{m}}$ and $1_{\mathbf{m}}^T T_* 1_{\mathbf{m}}/2$. The selection of vectors from $X_{\mathbf{m}}$ to form S_* does have implications on the estimability properties of the design. This is now considered briefly and illustrated with reference to Example 8.

EXAMPLE 8. continued For the S_* matrix of (4.5),

$$1_3^T S_* = \begin{pmatrix} 2 & 2 & 2 & 1 & 3 \end{pmatrix} \quad \text{and} \quad T_* = \begin{pmatrix} 0 & 2 & 2 & 3 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 2 & 0 & 1 & 1 \\ 3 & 1 & 1 & 0 & 2 \\ 1 & 1 & 1 & 2 & 0 \end{pmatrix}. \quad (4.6)$$

The total numbers of main effect estimates and interaction estimates from the three replicates are $1_3^T S_* 1_3 = 10$ and $1_3^T T_* 1_3/2 = 16$, respectively. It is interesting to consider the impact of changing the selection of vectors of X_3 used in S_* . First relabel S_* and T_* of (4.5) and (4.6) as S_{*1} and T_{*1} , and denote the design by $\mathcal{D}81$. An alternative design, $\mathcal{D}82$, is obtained by replacing the final column of S_* by $(1, 0, 0)^T$. We denote this alternative matrix by S_{*2} with corresponding matrix T_{*2} to give

$$S_{*2} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix}, \quad 1_3^T S_{*2} = (2 \ 2 \ 2 \ 1 \ 1), \quad T_{*2} = \begin{pmatrix} 0 & 2 & 2 & 3 & 1 \\ 2 & 0 & 2 & 1 & 1 \\ 2 & 2 & 0 & 1 & 3 \\ 3 & 1 & 1 & 0 & 2 \\ 1 & 1 & 3 & 2 & 0 \end{pmatrix}. \quad (4.7)$$

In $\mathcal{D}82$, the main effect E is only estimable from one replicate and the total number of main effect estimates is $1_3^T S_{*2} 1_3 = 8$ compared to the ten estimates from $\mathcal{D}81$. The number of interaction estimates from $\mathcal{D}82$ is $1_3^T T_{*2} 1_3/2 = 18$: an increase on the 16 estimates from $\mathcal{D}81$. Focussing on those effects estimated with smallest relative information, from the design summaries in (4.6) and (4.7) it is observed that in $\mathcal{D}81$ one main effect and five interactions are estimable from only one replicate but that in $\mathcal{D}82$ two main effects and four interactions have this property.

Dash *et al.* (2013) provide designs in $\mathbf{m}$ replicates of a 2^{q+1} factorial arranged in 2×2^q arrays for $2 \leq n = q + 1 \leq 9$ which enable estimation of all effects of interest. Theorem 4 gives designs with equivalent estimability properties to those of Table 3 of Dash *et al.* (2013), in the sense that the numbers and distributions of independent estimates of main effects and two factor interactions are the same.

EXAMPLE 8. continued The two designs given in Dash *et al.* (2013) for three 2×2^4 single replicate arrays of a 2^5 factorial both have estimability properties equivalent to

that of a design of Theorem 4, $\mathcal{D}83$ say, with:

$$\mathcal{S}_{*3} = \begin{pmatrix} 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}, \quad 1_3^T \mathcal{S}_{*3} 1_3 = (2 \ 2 \ 1 \ 1 \ 3), \quad T_{*3} = \begin{pmatrix} 0 & 2 & 1 & 3 & 1 \\ 2 & 0 & 3 & 1 & 1 \\ 1 & 3 & 0 & 2 & 2 \\ 3 & 1 & 2 & 0 & 2 \\ 1 & 1 & 2 & 2 & 0 \end{pmatrix}.$$

Design $\mathcal{D}83$ has $1_3^T \mathcal{S}_{*3} 1_3 = 9$ independent main effect estimates and $1_3^T T_{*3} 1_3 / 2 = 18$ independent interaction estimates. Two main effects and four interactions are estimable from only one replicate.

In general, selection of vectors from X_m with a greater number of elements unity gives greater emphasis to estimation of main effects. Selection of vectors so that the number of non-zero elements in a row of S_* is close to $(q+1)/2$ gives greater emphasis to estimation of interactions. This is demonstrated in Example 8. For some values of q there are a large number of selections of $q+1$ vectors from X_m . In such cases, judicious choice of vectors and arrangement of these to form S_* enables construction of bespoke designs that provide estimates of selected effects with high relative information. Conversely, for other values of q , the range of estimability properties of designs from Theorem 4 are very limited. This is expressed in the following corollary to Theorem 4:

COROLLARY 2. *If $q = 2^k - 2$ for some integer $k \geq 3$, then the designs of Theorem 4 all have equivalent estimability properties. If $q = 2^k - 3$, then the designs of Theorem 4 can be partitioned into k types with respect to estimability properties.*

Proof The estimability properties of an m array design obtained from Theorem 4 are determined by the columns of S_* . If $q = 2^k - 2$ or $2^k - 3$, for $k \geq 3$, then $m = k$ and S_* is formed from distinct vectors of X_k . For $q = 2^k - 2$, matrix S_* comprises all $q+1 = 2^k - 1$ vectors of X_k and thus all designs have identical S_* and T_* matrices, bar a reordering. For $q = 2^k - 3$, matrix S_* comprises all but one of the vectors of X_k . The number of zeros in the omitted column vector can be any one of $0, 1, \dots, k-1$, giving k estimability structures.

5. Multi-replicate arrays

We now consider single array designs containing two or more replicates. The model of (2.1) applies and the AGM has order $(p+q) \times n$ with $n < p+q$. Further, $q < n$, since only arrays with incomplete blocks are considered. The design consists of 2^{p+q-n} replicates in a $2^p \times 2^q$ array. The approach used is to adapt the constructions for single replicate arrays of §3 and the selection of sets of single replicate arrays of §4. Fewer degrees of freedom are required for nuisance parameters, i.e. blocks or replicates, than for a design comprising several arrays with the same number of runs. Note that $p \geq 2$ and that, in terms of p and q , possible values for n are $p+1, \dots, p+q-1$.

5.1. Multi-replicate array for $p \geq 3, q < n < p+q$

We give constructions satisfying Conditions 1 and 2B. Define

$$\mathcal{H}_{cp} = \left(\begin{array}{c} I_p \\ I_p + J_{p,p} \end{array} \right) \quad \text{and} \quad \mathcal{H}_{rp} = \left(\begin{array}{c} K_p \\ L_p \end{array} \right),$$

where K_p and L_p are as given in Lemma 1, and let $M(a)$ denote the first a columns of matrix M .

THEOREM 5. *An AGM of the form specified below generates a $2^p \times 2^q$ binary array containing 2^{p+q-n} replicates of a 2^n factorial, from which all main effects and ω interactions are estimable:*

Case 1: $n \leq 2p$

$$\mathcal{G} = \left(\begin{array}{c} \mathcal{G}_c \\ \mathcal{G}_r \end{array} \right) = \left(\begin{array}{c} \frac{\mathcal{H}_{cp}(n)}{\mathcal{H}_{rp}(n)} \\ \hline 0_{q-p, n-q+p} \quad I_{q-p} \end{array} \right),$$

Case 2: $n > 2p$

$$\mathcal{G} = \left(\begin{array}{c} \mathcal{G}_c \\ \mathcal{G}_r \end{array} \right) = \left(\begin{array}{c} \frac{\mathcal{H}_{cp}}{\mathcal{H}_{rp}} \quad \begin{array}{c} X_* \\ 0_{p, n-2p} \end{array} \\ \hline 0_{q-p, n-q+p} \quad I_{q-p} \end{array} \right),$$

where X_* is a $p \times (n-2p)$ matrix such that each vector of X_p appears in the $p \times n$ matrix $\mathcal{G}_c$, α or $\alpha + 1$ times, where $\alpha = \lfloor n/(2^p - 1) \rfloor$.

The proof is by an argument similar to that of Theorem 2 with the minor adjustment that it is necessary to demonstrate that $\mathcal{G}_c$ and $\mathcal{G}_r$ are each full rank in order for the design to be binary with respect to rows and columns. Since the columns of $\mathcal{G}_r$ are distinct, estimability properties are determined by the submatrix $\mathcal{G}_c$. If $n \leq 2^p - 1$, then the columns of $\mathcal{G}_c$ are also distinct and the array of Theorem 5 provides estimates of all effects of interest with full information. Otherwise, some interactions will be inestimable.

The arrays of Theorem 5 are resolvable in both rows and columns. The 2^q treatment combinations of the principle row partition the 2^n treatment combinations into 2^{n-q} cosets. The 2^q treatment combinations in each coset occur together in 2^{p+q-n} rows. A similar property is noted in columns. This property causes the design to be resolvable in a number of alternative ways with respect to rows and columns. For example, for a two replicate design, that is a design with $2^{p+q-n} = 2$, the rows can be resolved into two replicates in $2^{2^{n-q}-1}$ ways. Further, for 2^{n-q} of these ways of resolving the rows, the individual replicates can be expressed in terms of generator matrices. Similarly, there are $2^{2^{n-p}-1}$ ways of resolving the columns into two replicates, of which 2^{n-p} can be given in terms of generator matrices. The property of being resolvable according to generator matrix constructions can be beneficial in the event of observation loss, as discussed in the following example.

EXAMPLE 9. *A two replicate design for a 2^5 factorial is required in a $2^3 \times 2^3$ array.*

By Theorem 5, an AGM for $p = q = 3$ and $n = 5$ which provides a design, $\mathcal{DA}$ say, from which all main effects and interactions are estimable is:

$$\mathcal{G}_{\mathcal{DA}} = \left(\frac{\mathcal{G}_c}{\mathcal{G}_r} \right) = \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 1 & \\ 0 & 1 & 0 & 1 & 0 & \\ 0 & 0 & 1 & 1 & 1 & \\ \hline 1 & 1 & 1 & 1 & 0 & \\ 0 & 1 & 1 & 0 & 0 & \\ 0 & 0 & 1 & 1 & 1 & \end{array} \right). \quad (5.1)$$

Note that $\mathcal{G}_{\mathcal{DA}}$ comprises the first five columns of the AGM of Example 3 (i). The design is:

$$\mathcal{DA} = \begin{array}{cccccccc} (1) & abcd & bc & ad & cde & abe & bde & ace \\ ae & bcde & abce & de & acd & b & abd & c \\ bd & ac & cd & ab & bce & ade & e & abcde \\ abde & ce & acde & be & abc & d & a & bcd \\ cde & abe & bde & ace & (1) & abcd & bc & ad \\ acd & b & abd & c & ae & bcde & abce & de \\ bce & ade & e & abcde & bd & ac & cd & ab \\ abc & d & a & bcd & abde & ce & acde & be \end{array}$$

Design $\mathcal{DA}$ yields full information on all the effects of interest. Designs containing two replicates of a 2^5 factorial arranged in an 8×8 array, using a construction approach involving quasi-Latin squares are given in Rao (1946) and investigated by many researchers, including Shah and Sinha (1996) and Choi and Gupta (2008). Such a design is:

$$\mathcal{DSS} = \begin{array}{cc|cccc|cccc} (1) & bcde & bc & de & abd & ace & abe & acd \\ abde & ac & acde & ab & bcd & e & bce & d \\ ae & abcd & abce & ad & cde & b & c & bde \\ bd & ce & cd & be & ade & abc & a & abcde \\ \hline SQ1 & SQ2 & & & & & & \\ SQ3 & SQ4 & & & & & & \\ \hline abc & abe & bde & bcd & (1) & ad & acde & ce \\ acd & ade & e & c & abce & bcde & bd & ab \\ cde & d & a & ace & be & abde & abcd & bc \\ bce & b & abd & abcde & ac & cd & de & ae \end{array} \quad (5.2)$$

Design $\mathcal{DSS}$ consists of squares, $SQ1 \dots, SQ4$, each containing a half replicate, with a complete replicate in rows 1-4, rows 5-8, columns 1-4 and columns 5-8. The construction is such that: rows 1-4 are confounded with ABC, ADE ; rows 5-8 are confounded with ABD, BCE ; columns 1-4 are confounded with ACE, BCD ; columns 5-8 are confounded with ACD, BDE . In addition, treatment combinations in each square are selected to be confounded with $ABCDE$. There is a slip in presentation of the design in Shah and Sinha (1996) in that the entries in rows of $SQ3$ are incorrectly arranged with the consequence that estimates of C, AE, BD are obtained only using $SQ2, SQ3, SQ4$. Comments on the design here relate to $\mathcal{DSS}$ as given in (5.2). The design cannot be constructed by the generator matrix approach of this paper. As with $\mathcal{DA}$, all main effects and interactions are estimable from $\mathcal{DSS}$ with full information. Further, $\mathcal{DSS}$ has the advantage that it also provides estimates of all higher order interactions.

However, $\mathcal{DA}$ has beneficial robustness properties in the event of observation loss or premature termination of the experiment that are not shared by $\mathcal{DSS}$. There are four ways of resolving the rows of $\mathcal{DA}$ into separate replicates according to a generator matrix construction, and likewise four ways of resolving the columns. Details of the four row resolutions are given below, in each case with the $\mathcal{G}_c$ sub-matrix for the set of rows including row 1 and information on effects which are not estimable from the individual replicates. The $\mathcal{G}_r$ sub-matrix is that of (5.1).

rows in each replicate	$\mathcal{G}_c$	non-estimable effects
1,2,3,4	$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$	C, AE, BD
5,6,7,8		
1,2,7,8	$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$	D, BC
3,4,5,6		
1,3,6,8	$\begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{pmatrix}$	E, AC
2,4,5,7		
1,4,6,7	$\begin{pmatrix} 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}$	AD, BE
2,3,5,8		

Consider the loss of any two rows of $\mathcal{DA}$ not containing the same eight treatment combinations, that is, any set of two rows except $\{1, 5\}$, $\{2, 6\}$, $\{3, 7\}$ or $\{4, 8\}$. Of the eight replicates listed above, exactly two will involve neither row and are preserved in the sub-design comprising the remaining six rows. These two replicates together give estimates of all main effects and interactions. For example, if rows 5 and 7 are lost, then replicates comprising rows 1,2,3,4 and 1,3,6,8 are preserved. The first of these gives estimates of all effects of interest except C, AE and BD . Likewise, all effects of interest except E and AC are obtained from the second replicate. Thus, all main effects and interactions are estimable from the sub-design. To summarise, for 24 of the 28 sub-designs comprising six rows, all effects of interest can be obtained. The same property is observed if two columns are lost.

Now consider the loss of two rows from $\mathcal{DSS}$. If one of the first four rows and one of the last four rows are lost, then no complete replicate remains. Even if two rows are lost from the first four or last four rows, it does not follow that all effects of interest will be estimable from the sub-design of the remaining six rows. For example, if rows 5 and 7 are lost then C, D, BE and AE are all inestimable. In fact there are only four pairs of rows which can be lost to give a sub-design from which all main effects are estimable. To summarise: for 16 of the 28 ways in which two rows can be lost, no effects of interest will be estimable; the loss of any pair of rows compromises at least one main effect or interaction. A similar property is noted in the event that two columns are lost.

If $n > 2^p - 1$, a $2^p \times 2^q$ array generated by Theorem 5 will provide estimates of all main effects but not of all interactions. Alternative constructions involving sub-arrays enable

estimation of more effects by confounding different interactions in each sub-array. Let q_1 be such that $p \leq q_1 < q$ and $p + q_1 \geq n$, then a $2^p \times 2^q$ array can be formed by arranging 2^{q-q_1} sub-arrays of size $2^p \times 2^{q_1}$ side by side. This is illustrated by the following example:

EXAMPLE 10. *A four replicate design in a $2^3 \times 2^7$ array is required for a 2^8 factorial.*

The AGM of Theorem 5 has a repeated column in $\mathcal{G}_c$ and so one interaction is not estimable. Two alternative constructions are suggested:

$q_1 = 6$ Two arrays, each of order $2^3 \times 2^6$ and containing two replicates are constructed by Theorem 5 and arranged side by side to form a single array. Each sub-array provides estimates of all effects of interest except for one interaction. Careful arrangement of the columns of each generator matrix gives a design in which all main effects and 26 of the 28 interactions are estimated with full efficiency and the remaining two interactions are estimated with 1/2 relative information.

$q_1 = 5$ Four single replicate arrays of order $2^3 \times 2^5$ are constructed by Theorem 2, with each array providing estimates of all effects of interest except for one interaction, a different interaction being inestimable from each replicate. The arrays are arranged side by side to give a design from which all main effects and 24 interactions are estimated with full information and the remaining four interactions with 3/4 relative information.

5.2. Two-replicate design for $p=2$, $n=q+1$

For $q > 2$, the approach of §4.1 is used to produce two single replicate arrays such that each replicate enables estimation of all main effects and together the replicates cover as many interactions as possible. For $n \leq 9$, all interactions can be estimated. The multi-replicate array will comprise the two arrays side by side. As with the two array design, main effects are estimated with full information. For $n \leq 9$ interactions will be estimated with full or half information. For $n > 9$ some interactions will be inestimable.

Finally, we deal with the case $p = q = 2$ separately, since the $\mathcal{G}_r$ submatrix for $p = q = 2$ does not have distinct columns. The first three columns of (3.5) can be used as the AGM for a binary design. All main effects are estimable as are AC and BC . Four degrees of freedom are available for estimating σ^2 . The array is:

$$\begin{array}{cccc} (1) & ab & abc & c \\ ac & bc & b & a \\ bc & ac & a & b \\ ab & (1) & c & abc \end{array}$$

6. Fractional Designs

Up to this point all arrays have contained at least one full 2^n factorial replicate and so have had $p+q \geq n$. In this section, we consider the exercise of arranging a 2^{n-t} fractional factorial in a $2^p \times 2^q$ array, with $p+q = n-t$. The use of fractional factorials extends the level of complexity substantially and only the surface of the problem is touched on

here. Sun *et al.* (1997) approach the related problem of constructing fractional factorial designs with one blocking factor by starting with a 2^{n-t} factorial, with a high ranking according to aberration, and investigating allocation of the treatment combinations to blocks to maximise the number of estimable main effects and two factor interactions.

A different approach is used here. We start with two independent entities: an AGM, $\mathcal{G}^B$, for a 2^{p+q} factorial in a $p \times q$ array, obtained using results already established, and a 2^{n-t} factorial. Adopting terminology of Das (1964), the n factors of the fractional factorial can be separated into $p + q$ *basic* factors, $A_1, \dots, A_{p+q}$, and t *added* factors, $A_{p+q+1}, \dots, A_n$, such that each added factor is aliased with an effect involving only the basic factors. The $p + q$ basic factors are mapped onto the columns of $\mathcal{G}^B$. The added factors correspond to t additional columns in the AGM. Denoting the sub-matrix of additional columns by $\mathcal{G}^A$, the final AGM is a $(p + q) \times n$ matrix:

$$\mathcal{G} = \left(\begin{array}{cc} \mathcal{G}^B & \mathcal{G}^A \end{array} \right) = \left(\begin{array}{cc} \mathcal{G}_c^B & \mathcal{G}_c^A \\ \mathcal{G}_r^B & \mathcal{G}_r^A \end{array} \right). \quad (6.1)$$

The sub-matrix $\mathcal{G}^A$ and subsequent properties of the design are dependent on the mapping. Each added factor, A_{p+q+i} for $i = 1, \dots, t$, is represented by an effect in the basic factors, and equivalently by v_i , a column vector of length $p + q$. The column of $\mathcal{G}^A$ corresponding to A_{p+q+i} has $g_{ci}^A = \mathcal{G}_c^B v_i$ and $g_{ri}^A = \mathcal{G}_r^B v_i$.

Fractional factorials used should have resolution at least IV to avoid main effects being aliased with other effects of interest. For fractional factorials of resolution V or higher, any non-estimable effects of interest are identifiable from the AGM in the usual way. For a fractional factorial of resolution IV, non-estimable main effects are recognised by inspection of $\mathcal{G}$, but attention has to be given to both $\mathcal{G}$ and to the effects of length four in the defining relationship to identify non-estimable interactions. Two examples are given for illustration.

EXAMPLE 11. *A design is required which comprises a 2^{5-1} fractional replicate arranged in a $2^2 \times 2^2$ array.*

The AGM of (3.5) is used for $\mathcal{G}_B$. The only defining relationship to yield a 2^{5-1} factorial of resolution V is $I = A_1 A_2 A_3 A_4 A_5$. Without loss of generality, let $A_1, \dots, A_4$ be the basic factors and map these onto columns of $\mathcal{G}_B$, giving the relabelling $A_1 \equiv A$, $A_2 \equiv B$, $A_3 \equiv C$, $A_4 \equiv D$. The added factor $A_5 \equiv E$ and the vector representation of E in terms of the basic factors is $v_E = (1, 1, 1, 1)^T$. Thus, $g_{cE}^A = \mathcal{G}_c^B (1, 1, 1, 1)^T = (1, 1)^T$ and $g_{rE}^A = \mathcal{G}_r^B (1, 1, 1, 1)^T = (1, 1)^T$. The AGM for the half replicate and the subsequent array are displayed below:

$$\mathcal{G} = \left(\begin{array}{cccccc} 1 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ \hline 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 & 1 \end{array} \right) \quad \begin{array}{llll} (1) & abde & abce & cd \\ acde & bc & bd & ae \\ bcde & ac & ad & be \\ ab & de & ce & abcd \end{array}$$

The estimability properties between the basic factors are retained from $\mathcal{G}_B$: the main effects A, B, C, D are estimable, as are the interactions AC, AD, BC and BD . The 2×1

vectors g_{cE}^A and g_{rE}^A are in X_2 , and so E is estimable. Finally, from $\mathcal{G}$ it is seen that no interactions involving E are estimable. Thus, all main effects but only four of the ten interactions are estimable. By comparison, the $2^2 \times 2^3$ array containing a full replicate of the 2^5 factorial obtained using Theorem 3, yields estimates of all main effects and all interactions excluding AE and BD .

EXAMPLE 12. A 2^{7-2} fractional replicate is required in a $2^2 \times 2^3$ array.

The AGM of Theorem 3, with $p = 2$, $q = 3$, is adopted for $\mathcal{G}_B$. The fractional factorial selected is the minimum aberration 2^{7-2} factorial from the catalogue of Chen *et al.* (1993). This has defining relationship $I = A_1 A_2 A_3 A_6 = A_1 A_2 A_4 A_5 A_7 = A_3 A_4 A_5 A_6 A_7$. We use trial end error to identify a mapping from the basic factors $A_1, \dots, A_5$ to $A, \dots, E$ which gives a design with good estimability properties. Without loss of generality, the added factors A_6, A_7 are mapped to F, G in that order. Priority is given to estimation of main effects over interactions. Designs corresponding to three mappings are considered below.

Mapping 1: The basic factors are mapped onto the columns of $\mathcal{G}_B$ in order, giving the relabelling $A_1 \equiv A$, $A_2 \equiv B$, $A_3 \equiv C$, $A_4 \equiv D$, $A_5 \equiv E$. Thus, the defining relationship is $I = ABCF = ABDEG = CDEFG$ and the vector representations of F, G in terms of the basic factors are $v_F = (1, 1, 1, 0, 0)^T$ and $v_G = (1, 1, 0, 1, 1)^T$. This gives $g_{cF}^A = \mathcal{G}_c^B v_F = (0, 0)^T$ and $g_{cG}^A = \mathcal{G}_c^B v_G = (0, 0)^T$. Hence, the main effects F, G are confounded with columns and are not estimable.

Mapping 2: Mapping 1 is amended by switching A_3 and A_4 , so $A_3 \equiv D$ and $A_4 \equiv C$. The new defining relationship is $I = ABDF = ABCEG = CDEFG$ and the vector representations of F, G are $v_F = (1, 1, 0, 1, 0)^T$ and $v_G = (1, 1, 1, 0, 1)^T$. This gives $g_{cF}^A = \mathcal{G}_c^B v_F = (1, 0)^T$, $g_{cG}^A = \mathcal{G}_c^B v_G = (1, 0)^T$, $g_{rF}^A = \mathcal{G}_r^A v_F = (1, 1, 1)^T$ but $g_{rG}^A = \mathcal{G}_r^A v_G = (0, 0, 0)^T$. Hence, F is estimable but G is confounded with rows and is not estimable.

Mapping 3: The basic factors are mapped onto the columns of $\mathcal{G}_B$ via $A_1 \equiv B$, $A_2 \equiv D$, $A_3 \equiv A$, $A_4 \equiv C$, $A_5 \equiv E$. The defining relationship is $I = ABDF = BCDEG = ACEFG$. The vector representation of F and corresponding column in $\mathcal{G}^A$ are as for Mapping 2. The representation for G is $v_G = (0, 1, 1, 1, 1)^T$, giving $g_{cG}^A = \mathcal{G}_c^B v_G = (0, 1)^T$ and $g_{rG}^A = \mathcal{G}_r^A v_G = (1, 0, 0)^T$. Hence, F and G are both estimable.

Of the mappings considered, Mapping 3 is preferred since it is the only one providing estimates of both added factors. The AGM for Mapping 3 is:

$$\mathcal{G} = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ \hline 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}.$$

The array gives estimates of all main effects and of ten of the 21 interactions. The interactions AE , AF , EF , BD , BG and DG are confounded with columns, whilst EG is confounded with rows. Finally, due to the involvement of $ABDF$ in the defining relationship, AB , AD , BF and DF are also non-estimable.

7. Discussion

The method of constructing an array from a principal row and column is common to approaches found elsewhere including Hinkelmann and Kempthorne (2005), Choi and Gupta (2008), Dash *et al.* (2013) and Wang (2017). The advance of this work is the body of array templates provided by results of §3 to §5 for construction of designs in rectangular arrays. These cover the vast majority of p, q, n combinations likely to be of practical interest. The approaches of §4 and §5 can be combined to produce designs comprising multiple arrays, each involving two or more replicates. This is especially useful where: not all effects of interest are estimable from one multiple array; few degrees of freedom are available for estimation of σ^2 . For example, use of a second two-replicate array to accompany that for the 2^3 factorial given in §5.2, would enable the estimation of all effects of interest and provide 12 degrees of freedom for estimation of σ^2 .

This work focuses on estimation of main effects and two factor interactions. The approach can be extended to encompass higher order interactions. In general, for an AGM satisfying Condition 1, a factorial effect will be estimable *iff* the corresponding columns of $\mathcal{G}_c$ do not have sum 0_p and those of $\mathcal{G}_r$ do not have sum 0_q . For example, the first three columns of $\mathcal{G}_{\mathcal{DA}}$ of Example 9 have sum $(1, 1, 1, 1, 0, 1)^T$ indicating that the three factor interaction ABC is estimable in $\mathcal{DA}$, whereas the second, third and fourth columns sum to $(0, 0, 0, 1, 0, 0)^T$ indicating that BCD is confounded with columns. In total, of the three factor interactions, six would be estimable from $\mathcal{DA}$, two are confounded with columns and two are confounded with rows. Note that in order for a three factor interaction and marginal terms to all be estimable, it is necessary that $p \geq 3$.

References

- Bailey, R.A. (1985). Patterns of confounding in factorial designs. *Biometrika* **64**, 597–603.
- Bailey, R.A. (1985). Factorial design and Abelian groups. *Linear Algebra and its Applications* **70**, 349–368.
- 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. *International Statistical Review* **61**, 131–145.
- Cheng, C.S., Tsai, P.W. (2013). Templates for design key construction. *Statistica Sinica* **23**, 1419–1436.
- Choi, K.C., Gupta, S. (2008). Confounded row-column designs. *Journal of Statistical Planning and Inference* **138**, 196–202.
- Dash, S., Parsad, R. and Gupta, V.K. (2013). Row-column designs for 2^n factorial 2-colour microarray experiments for estimation of main effects and two-factor interactions with orthogonal parameterization. *Agricultural Research* **2**, 172–182.

- Datta, A., Jaggi, S., Varghese, E., and Varghese, C. (2017). Generalized confounded row-column designs. *Communications in Statistics - Theory and Methods* **46**, 6213–6221.
- Das, M.N. (1964). A somewhat alternative approach for construction of symmetrical factorial designs and obtaining maximum numbers of factors. *Calcutta Statistical Association Bulletin* **13**, 1–17.
- Dean, A.M., Lewis, S.M. (1980). A unified theory for generalized cyclic designs. *Journal of Statistical Planning and Inference* **4**, 13–23.
- Dean, A.M., Lewis, S.M. (1992). Multidimensional designs for two-factor experiments. *Journal of the American Statistical Association* **87**, 1158–1165.
- Gopinath, P.P., Parsad, R. and Mandal, B.N. (2017). Incomplete row-column designs with factorial structure for estimating main effects with full efficiency. *Communications in Statistics - Theory and Methods* DOI: 10.1080/03610926.2017.1376091.
- Hinkelmann, K., Kempthorne, O. (2005). Design and analysis of experiments, Vol 2. Wiley, New York.
- John, J.A., Lewis, S.M. (1983). Experiments in generalized cyclic row-column designs. *Journal of the Royal Statistical Society, Series B* **45**, 245–251.
- John, J.A., Whitaker, D. (1993). Construction of resolvable row-column designs using simulated annealing. *Australian Journal of Statistics* **35**, 237–245.
- Kobilinsky, A. (1985). Confounding in relation to duality of finite Abelian groups. *Linear Algebra and its Applications* **70**, 321–347.
- Patterson, H.D., Bailey, R.A. (1978). Design keys for factorial experiments. *Journal of the Royal Statistical Society, Series C* **27**, 335–343.
- Preece, D.A. (1966). Some row and column designs for two sets of treatments. *Biometrics* **22**, 1–25.
- Preece, D.A. (1971). Some new balanced row-and-column designs for two non-interacting sets of treatments. *Biometrics* **27**, 426–430.
- Rao, C.R. (1946). Confounded factorial designs in quasi-Latin squares. *Sankhya* **7**, 295–304.
- Shah, K.R., Sinha, B.K. (1996). Row-column designs. Included in: Ghosh, S., Rao, C.R. (Eds.), Handbook of Statistics, vol. 13, Elsevier, New York, 903–937.
- 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.
- Wang, P.C. (2017). Orthogonal main-effect plans in row-column designs for two-level factorial experiments. *Communications in Statistics - Theory and Methods* DOI: 10.1080/03610926.2016.1242743.

- Williams, E.R., John, J.A. (1996). Row-column factorial designs for use in agricultural field trials. *Journal of the Royal Statistical Society, Series C* **45**, 39–46.
- Yates, F. (1937a). The design and analysis of factorial experiments. *Technical Communication Bureau of Soil Science, Harpenden* **35**.
- Yates, F. (1937b). A further note on the arrangement of variety trials: quasi-Latin squares. *Annals of Eugenics* **7**, 319–332.