COMMUN. STATIST.—THEORY METH., 25(11), 2557-2573 (1996)

MARGINALLY OVERSATURATED DESIGNS

Lih-Yuan Deng[1], Dennis K.J. Lin[2] and Jiannong Wang[3]

[1]The University of Memphis, Memphis, TN. 38138

[2]Pennsylvania State University, University Park, PA. 16802

[3]American Express, New York, NY. 10285

Key Words: Hadamard matrix; resolution rank; screening experiment; supersaturated design.

ABSTRACT

A special class of supersaturated design, called marginally over saturated design (MOSD), in which the number of variables under investigation (k) is only slightly larger than the number of experimental runs (n), is presented. Several optimality criteria for supersaturated designs are discussed. It is shown that the resolution rank criterion is most appropriate for screening situations. The construction method builds on two major theorems which provide an efficient way to evaluate resolution rank. Examples are given for the cases n=8, 12, 16, and 20. Potential extensions for future work are discussed.

1. INTRODUCTION

Screening experiments typically contain a large number of potential factors. It is not unusual, however, that among those factors only a small portion are, in fact, active. The usual advice given to practitioners is to run the so-called main-effect designs (Resolution III designs in the orthogonal case) which require at least $k+1$ runs for investigating k factors. However, estimating all main effects may be wasteful if the goal is only to detect the active factors. This is particularly true when only a small number of runs is desired. In such a situation, a supersaturated design can save considerable costs.

A supersaturated design is a fraction of a factorial design having n factor-level combinations while the number of factors, k, is more than $n-1$. When such a design is used, the abandonment of orthogonality is inevitable, because otherwise, the factor columns would form a set of more than n orthogonal vectors in n-dimensional space. Apart from some ad hoc procedures and computer-generated designs, the construction of supersaturated designs has been addressed only recently; see for example, Lin (1991, 1993a, 1995), Wu (1993), Tang and Wu (1993), and Deng, Lin and Wang (1994).

When running an experiment with a small run size, users should be aware of the increased risk of both false positive and false negative signals, corresponding to Type I and II errors in classical hypothesis testing settings. However, the difference between saturated main-effect designs and supersaturated designs is not of nature but only of degree. As shown in Lin (1995), this degree increase with $k-n$. That is, if k is slightly larger than n, then the difference is only marginal. A supersaturated design is equally as good as a saturated design in this case.

Of particular interest in this paper is a special class of supersaturated designs called the marginally over-saturated design (MOSD). In these designs,

the number of factors, k, is only slightly larger than the number of runs, n. Specifically, we consider supersaturated designs for the cases $k = n$ and $n + 1$. Optimality criteria are discussed in Section 2, where the resolution rank criterion is shown to be the most appropriate for evaluating the goodness of a supersaturated design, and thus will be used for design construction. In Section 3, we prove the main theorems for our construction method. Section 4 gives some results for $n = 8, 12, 16$, and 20. Finally, summary comments, conclusions and possible extensions for future work are presented in Section 5.

2. OPTIMALITY CRITERIA

As previously mentioned, abandonment of orthogonality in a supersaturated design is inevitable. Since a lack of orthogonality results in lower efficiency in estimation, it is desirable to make the design as nearly orthogonal as possible when perfect orthogonality is unattainable. To be able to measure the degree of orthogonality of a supersaturated design, several criteria come to mind. A naive measurement is the maximum correlation among any pair of design columns, denoted by s.

$$s = \max\left(|s_{ij}|\right), \text{ where } s_{ij} = \mathbf{x}_i'\mathbf{x}_j, \text{ for all } 1 \le i < j \le k. \tag{1}$$

Here the i-th column of the design matrix $\mathbf{X}$ is denoted by $\mathbf{x}_i$. The smaller s, the better. Note that $s = 0$ implies perfect orthogonality.

Booth and Cox (1962) first proposed the $E(s^2)$ criterion to evaluate the goodness of a supersaturated design, and it has been extensively used by others. The $E(s^2)$ criterion is

$$E(s^2) = \sum s_{ij}^2 / \binom{k}{2}, \text{ where } s_{ij} = \mathbf{x}_i'\mathbf{x}_j, \text{ for all } 1 \le i < j \le k. \tag{2}$$

In a sense, the $E(s^2)$ criterion is an average of the s_{ij}^2 values between all possible pairs of design columns $\mathbf{x}_i$ and $\mathbf{x}_j$.

Lin (1995) modified the $E(s^2)$ criterion, by taking the run size n into account and proposed an equivalent criterion, called the mean square correlation ρ as shown below,

$$\rho = \sum r_{ij}^2 / \binom{k}{2}, \text{ where } r_{ij} = Corr(\mathbf{x}_i, \mathbf{x}_j), \text{ for all } 1 \le i < j \le k. \tag{3}$$

It is apparent that this criterion is more appropriate than $E(s^2)$ for comparing supersaturated designs with different run sizes.

Now, let $\mathbf{X}$ be the design matrix with entries ± 1, and let c be the number of active factors, i.e., the number of design columns of the projected design matrix. For a given $\mathbf{s} = (l_1, \cdots, l_c)$, which is a set of size c from $(1, \cdots, k)$, we can construct a $n \times c$ sub-matrix $\mathbf{X}_s$ from $\mathbf{X}$. Following the idea of $E(s^2)$, which gives an efficiency measurement in an average sense, we can measure the goodness of $\mathbf{X}$ as follows:

$$V_c(\mathbf{X}) = \frac{1}{\binom{k}{c}} \cdot \sum_{|s|=c} v(\mathbf{X}_s), \tag{4}$$

where $v(\mathbf{X}_s)$ is a function that measures the "orthogonality" of $\mathbf{X}_s$ where the summation is taken over all possible choices of s.

As an extension of classical design optimality, some natural choices of $v(\mathbf{X}_s)$ are:

$$v(\mathbf{X}_s) = \det(\mathbf{X}_s'\mathbf{X}_s)^{-1}; \tag{5}$$

$$v(\mathbf{X}_s) = \text{trace } (\mathbf{X}_s'\mathbf{X}_s)^{-1}; \tag{6}$$

$$v(\mathbf{X}_s) = \lambda_{(c)}(\mathbf{X}_s'\mathbf{X}_s)^{-1}, \tag{7}$$

where $\lambda_{(c)}$ denotes the largest eigenvalue of the matrix $(\mathbf{X}_s'\mathbf{X}_s)^{-1}$.

Note that when $c = k \le n$, these criteria correspond to (i) D optimal, (ii) A optimal, and (iii) E optimal criteria, respectively. When $k > n$, the

value of c cannot be larger than k, and in fact, is normally much smaller than k. Note also that when $c = 2$, all criteria are reduced to the criterion similar to the one that proposed by Booth and Cox (1962) and Lin (1993a). One should also note that the first two criteria, (5) and (6), will optimize a design by minimizing $E(\frac{1}{n^2-s^2})$, whereas the criterion considered by Booth and Cox (1962) and Lin (1993a) is $E(s^2)$.

Clearly, if a vector $\mathbf{y}$ is orthogonal to a group of vectors $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_p)$, then the regression sum squares must be null, when we regress $\mathbf{y}$ on $(\mathbf{z}_1, \ldots, \mathbf{z}_p)$, that is, $\mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = 0$. Thus the value of $\mathbf{y}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$, or equivalently $\mathbf{b}'(\mathbf{Z}'\mathbf{Z})\mathbf{b}$, where $\mathbf{b} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$ is the vector of regression coefficients, provides a good measurement on how orthogonal the vector $\mathbf{y}$ is to $\mathbf{Z}$. Motivated by this, a series of criteria, called "B optimality," measures the dependency of a column to all other $c-1$ columns by computing the regression coefficients of one column in $\mathbf{X}_s$, $\mathbf{x}_i$, over the remaining columns $\mathbf{X}_{s-i}$. For any specific projection design $\mathbf{X}_s$ with size $n \times c$, we can average $\mathbf{x}_i'\mathbf{X}_{s-i}(\mathbf{X}_{s-i}'\mathbf{X}_{s-i})^{-1}\mathbf{X}_{s-i}'\mathbf{x}_i$ over all possible i $(i = 1, 2, \ldots, c)$ as a measure of the design orthogonality. Of course, the value of c is typically small (See, for example, Lin, 1993b).

In general, consider a class of new functions $v_g(\mathbf{X}_s)$ to measure the "orthogonality" of $\mathbf{X}_s$ for $V_c(\mathbf{X})$ in (4):

$$v_g(\mathbf{X}_s) = \sum_{i \in s} \beta_{s-i}'(\mathbf{X}_{s-i}'\mathbf{X}_{s-i})^g \beta_{s-i},$$

namely,

$$V_c(\mathbf{X}) = \frac{1}{\binom{k}{c}} \cdot \sum_{|s|=c} \sum_{i \in s} \beta_{s-i}'(\mathbf{X}_{s-i}'\mathbf{X}_{s-i})^g \beta_{s-i}, \tag{8}$$

where

1. $\beta_{s-i} = (\mathbf{X}_{s-i}'\mathbf{X}_{s-i})^{-1}\mathbf{X}_{s-i}'\mathbf{x}_i$,

2. $\mathbf{x}_i$ is the $n \times 1$ column corresponding to the i-th unit in s,

3. $\mathbf{X}_{s-i}$ is the $n \times (c-1)$ matrix corresponding to units in $s - \{i\}$,

and g can be any scalar value to present the degree of penalty to the near-singularity of the $\mathbf{X}'_{s-i}\mathbf{X}_{s-i}$ matrix. In principle, the B-criteria can be applied to any design when the projection property is of concern, regardless of the number of levels, the number of factors and the number of runs. For details, see, Deng, Lin and Wang (1996).

Note that once the few dominant active factors are identified, the initial design is then projected into a much smaller dimension. The implicit assumption with $E(s^2)$ is that there are, at most, two active factors. If the number of active factors c, is larger than 2, there is no guarantee that the projective (reduced) design will be of full rank, i.e., a main effects model consisting only of those active factors may not be estimable.

A criterion based upon such an important projection property, called resolution rank, is defined as follows.

Definition.: Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k)$ be a $n \times k$ matrix. Then the *resolution-rank* of $\mathbf{X}$ (r-rank, for short) is defined as $r = \max\{c$: for any $(\mathbf{x}_{i1}, \cdots, \mathbf{x}_{ic})$ of $\mathbf{X}$, $\mathbf{x}_{i1}, \cdots, \mathbf{x}_{ic}$ are linearly independent$\}$.

Clearly, if a supersaturated design, $\mathbf{X}$, has an r-rank of f, then when $\mathbf{X}$ is projected on to any submatrix of f (or less) factors, the main effects of the projected design are all estimable. Moreover, in many situations where the r-ranks are very different for two supersaturated designs, most of the other criteria values considered in this section are nearly the same.

3. MAIN THEOREMS

For a MOSD, the number of factors k is only slightly larger than $n - 1$. The basic idea here is to begin with an orthogonally based saturated design

and search for the best column(s) to optimize the r-rank criterion. For a two-level design, a Hadamard matrix is generally adopted as the orthogonal base. Since the evaluation for r-rank is very time-consuming, the brute-force search is not possible, even for small values of n. The following theorem has proven useful for constructing one additional column.

Theorem 1. Let $\mathbf{X} = [\mathbf{H}, \mathbf{v}]$, $\mathbf{w} = \mathbf{H}'\mathbf{v}$, where $\mathbf{H}$ is a Hadamard matrix of order n. Let R_1 be the number of non-zero entries in $\mathbf{w}$. Then

$$r = R_1,$$

where r is the r-rank of $\mathbf{X}$.

Proof. Pre-multiplying $\mathbf{H}$ on both sides of equation

$$\mathbf{w} = \mathbf{H}'\mathbf{v},$$

and noticing that $\mathbf{H}\mathbf{H}' = n\mathbf{I}$, we have

$$\begin{aligned} n\mathbf{v} &= \mathbf{H}\mathbf{w} \\ &= (\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n)\,(w_1, w_2, ..., w_n)' \\ &= \sum_{i=1}^{n} w_i \mathbf{h}_i. \end{aligned}$$

Therefore, $\mathbf{v}$ can be written as a linear combination of the columns of $\mathbf{H}$ and the coefficient associated with $\mathbf{h}_i$ is w_i. Since there are R_1 non-zero entries in $\mathbf{w}$, we have

$$r \leq R_1,$$

where r is the r-rank of $\mathbf{X}$.

On the other hand, since r is the r-rank of $\mathbf{X}$, there exists $r+1$ columns in $\mathbf{X}$ that are linearly dependent. Clearly, these $r+1$ can not all be columns of $\mathbf{H}$ because $\mathbf{H}$ is an orthogonal matrix. Therefore, we know that $\mathbf{v}$ is a linear combination of the remaining r columns. ¿From the unique representation of $\mathbf{v}$, we know that

$$r \geq R_1.$$

Combining these two inequalities, we have $r = R_1$. □

Remarks:

Usually, $\mathbf{v}$ is chosen as being column-balanced, that is, $\mathbf{v}$ has equal numbers of high- and low-level values. Clearly, there are a total of $\binom{n}{n/2}$ different cases. For large values of n, a complete search of the MOSD may not be possible. As a alternative, $\mathbf{v}$ can be selected via

(a) Product method: $\mathbf{D}(\mathbf{h}_i)\mathbf{h}_j, 1 \leq i < j \leq n$, where $\mathbf{D}(\mathbf{h}_i)$ is the diagonal matrix with diagonal vector $\mathbf{h}_i$, the i-th column vector of $\mathbf{H}$, $i = 1, 2, ...n$. In this case, there is a total of $\frac{n(n-1)}{2}$ cases.

(b) Random Permutation method: $\mathbf{P}\mathbf{h}_i$, where $\mathbf{P}$ is a matrix representing a row permutation. Since the number of permutations is very large, we restrict it to a fixed (say, 1000) number of searches.

Our empirical study shows that the product method in general fails to find a "good" design (in terms of r-rank) as compared with the random permutation method. We will only report the outcomes of the random permutation method in this paper. To add two additional columns to an orthogonal base, the following theorem provides an efficient approach.

Theorem 2. Let $\mathbf{X} = [\mathbf{H}, \mathbf{v}_1, \mathbf{v}_2]$ and $\mathbf{w}_1 = \mathbf{H}'\mathbf{v}_1$, $\mathbf{w}_2 = \mathbf{H}'\mathbf{v}_2$, where $\mathbf{H}$ is a Hadamard matrix of dimension n. Let

$$R_1 = \min[S(\mathbf{w}_1), S(\mathbf{w}_2)]$$

and

$$R_2 = \min[S(b_1\mathbf{w}_1 + b_2\mathbf{w}_2)] + 1,$$

where $S(\mathbf{u})$ represents the number of non-zero elements in the vector $\mathbf{u}$ and b_1, b_2 can take on all possible values. Then

$$r = \min[R_1, R_2],$$

where r is the r-rank of $\mathbf{X}$.

Proof. Using a similar technique as in Theorem 1, we see that

$$\begin{aligned} n\mathbf{v}_1 &= \sum_{i=1}^{n} w_{1i}\mathbf{h}_i \\ n\mathbf{v}_2 &= \sum_{i=1}^{n} w_{2i}\mathbf{h}_i, \end{aligned}$$

where $\mathbf{w}_1 = (w_{11}, w_{12}, \cdots, w_{1n})'$ and $\mathbf{w}_2 = (w_{21}, w_{22}, \cdots, w_{2n})'$. This implies that

$$r \leq R_1.$$

Furthermore, for any scalars b_1 and b_2, we have

$$n(b_1\mathbf{v}_1 + b_2\mathbf{v}_2) = \sum_{i=1}^{n}(b_1 w_{1i} + b_2 w_{2i})\mathbf{h}_i.$$

In other words, any linear combination of the columns of $\mathbf{v}_1$ and $\mathbf{v}_2$ can be represented by a combination of the columns in $\mathbf{H}$, with coefficients $(b_1 w_{1i} + b_2 w_{2i})$. Therefore, from the definition of r-rank, we have

$$r \leq R_2.$$

Consequently,

$$r \leq \min[R_1, R_2].$$

On the other hand, since r is the r-rank of $\mathbf{X}$, there exist $r+1$ columns in $\mathbf{X}$ that are linearly dependent. Clearly, these $r+1$ can not all be in columns of $\mathbf{H}$ because $\mathbf{H}$ is an orthogonal matrix. Therefore, we know that there are three possibilities:

1. Only $\mathbf{v}_1$ is contained in these $r+1$ column vectors.

2. Only $\mathbf{v}_2$ is contained in these $r+1$ column vectors.

3. Both $\mathbf{v}_1$ and $\mathbf{v}_2$ are contained in these $r+1$ column vectors.

For the case (1) or the case (2), we can see that

$$r \geq R_1.$$

For the case (3), we can see that

$$r \geq R_2.$$

Combining all three cases, we have

$$r \geq \min[R_1, R_2].$$

Hence,

$$r = \min[R_1, R_2].$$

□

Remarks:

- It is not necessary to consider all choices of b_1 and b_2 as in R_2. Rather, we need to consider only the solution set of

$$b_1 w_{1i} + b_2 w_{2i} = 0,$$

with either b_1 or b_2 set to 1. That is, we need consider $b_1 = 1, b_2 = -w_{1i}/w_{2i}$ or $b_2 = 1, b_1 = -w_{2i}/w_{1i}$ in R_2, for $1 \leq i \leq n$ with $w_{1i} \neq 0$ and $w_{2i} \neq 0$.

- If it is not practically possible to test for all the solutions of b_1 and b_2, (due to a large dimension n and computing time consideration) we can consider only the most likely cases (to produce many zeros in coefficients) when $b_1 = 1$ and $b_2 = \pm 1$. In this case, the R_2 computed is

only an estimate. However, according to our extensive empirical study, it is an excellent estimate of r-rank for MOSD. Deng, Lin and Wang (1994) proposed a similar estimate of the true r-rank for a constructed supersaturated design.

4. SOME RESULTS

In this section, we provide some useful MOSD's based on Theorems 1 and 2, for n=8, 12, 16, and 20 and $k = n$ or $n + 1$. In what follows, MOSD's with high r-rank are constructed using the following procedures:

- $\mathbf{v}_1$ is selected first (using permutation methods) with the highest r-rank for $[\mathbf{H}, \mathbf{v}_1]$.
- $\mathbf{v}_2$ is then selected (again using permutation methods) with the highest r-rank for $[\mathbf{H}, \mathbf{v}_1, \mathbf{v}_2]$.

Table 1 summarizes the optimal r-rank possible for these cases under consideration. Note that in all cases, the r-rank is no less than $n/2$, even when two columns are added. In fact, Srivastava (1975) shows, in "search design" context, that the number of active factors must be less than $n/2$ to insure the identifiability. All designs in Table 1 fulfill such a condition. We next discuss each design in detail.

4.1 Case $n = 8$.

Columns ($\mathbf{1}$, $\mathbf{x}_1$, $\mathbf{x}_2$, ..., $\mathbf{x}_7$) form a Hadamard matrix. By Theorems 1 and 2, we found the optimal added columns $\mathbf{v}_1$ and $\mathbf{v}_2$ to form a MOSD as given below. Table 2 summarizes most of its optimalities as discussed in Section 2. In addition, this design has $(s, \rho, E(s^2))$=(4,0.0625,4).

Table 1. Resolution Rank for Some Marginally Over-Saturated Designs

n	Number of Columns Added		
	0	1	2
8	7	4	4
12	11	9	7
16	15	13	10
20	19	17	15

Table 2. Various Criteria Values, $n = 8$.

f	A	D	B_2	B_1	B_0
2	0.267	4.792	6.400	0.800	0.100
3	0.435	3.756	19.200	2.511	0.368
4	0.653	2.979	38.400	5.321	0.958

1	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{v}_1$	$\mathbf{v}_2$
+	+	+	+	−	+	−	−	+	+
+	+	+	−	+	−	−	+	+	+
+	+	−	+	−	−	+	+	+	−
+	−	+	−	−	+	+	+	+	+
+	+	−	−	+	+	+	−	−	+
+	−	−	+	+	+	−	+	−	−
+	−	+	+	+	−	+	−	−	−
+	−	−	−	−	−	−	−	−	−

4.2 Case $n = 12$.

Columns $(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{11})$ form a Hadamard matrix. By Theorems 1 and 2, we found the optimal added columns $\mathbf{v}_1$ and $\mathbf{v}_2$ to form a MOSD as given below. Table 3 summarizes most of its optimalities as discussed in Section 2. In addition, this design has $(s, \rho, E(s^2))$=(4,0.02707,3.89744).

Table 3. Various Criteria Values, $n = 12$.

f	A	D	B_2	B_1	B_0
2	0.171	7.947	6.681	0.557	0.046
3	0.264	6.489	20.044	1.734	0.159
4	0.366	5.314	40.088	3.621	0.369
5	0.479	4.365	66.813	6.341	0.736
6	0.609	3.600	100.220	10.073	1.363
7	0.767	2.985	140.308	15.090	2.455

1	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$	$\mathbf{x}_9$	$\mathbf{x}_{10}$	$\mathbf{x}_{11}$	$\mathbf{v}_1$	$\mathbf{v}_2$
+	+	+	−	+	+	+	−	−	−	+	−	+	−
+	+	−	+	+	+	−	−	−	+	−	+	+	+
+	−	+	+	+	−	−	−	+	−	+	+	+	+
+	+	+	+	−	−	−	+	−	+	+	−	+	+
+	+	+	−	−	−	+	−	+	+	−	+	+	+
+	+	−	−	−	+	−	+	+	−	+	+	−	+
+	−	−	−	+	−	+	+	−	+	+	+	+	−
+	−	−	+	−	+	+	−	+	+	+	−	−	+
+	−	+	−	+	+	−	+	+	+	−	−	−	−
+	+	−	+	+	−	+	+	+	−	−	−	−	−
+	−	+	+	−	+	+	+	−	−	−	+	−	−
+	−	−	−	−	−	−	−	−	−	−	−	−	−

4.3 Case $n = 16$.

Columns $(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{15})$ form a Hadamard matrix. By Theorems 1 and 2, we found the optimal added columns $\mathbf{v}_1$ and $\mathbf{v}_2$ to form a MOSD as given below. Table 4 summarizes most of its optimalities as discussed in Section 2. In addition, this design has $(s, \rho, E(s^2))$=(8,0.01654,4.23529).

4.4 Case $n = 20$.

Columns $(\mathbf{1}, \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{19})$ form a Hadamard matrix. By Theorems 1 and 2, we found the optimal added columns $\mathbf{v}_1$ and $\mathbf{v}_2$ to form a MOSD as given below. Table 5 summarizes most of its optimalities as discussed in Section 2. In addition, this design has $(s, \rho, E(s^2))$=(12,0.00971,3.88571).

Table 4. Various Criteria Values, $n = 16$.

f	A	D	B_2	B_1	B_0
2	0.127	11.325	7.529	0.471	0.029
3	0.195	9.543	22.588	1.443	0.097
4	0.265	8.052	45.176	2.960	0.216
5	0.340	6.802	75.294	5.076	0.407
6	0.421	5.756	112.941	7.869	0.703
7	0.509	4.879	158.118	11.440	1.155
8	0.608	4.144	210.824	15.937	1.850
9	0.721	3.529	271.059	21.560	2.933
10	0.858	3.014	338.824	28.612	4.667

1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	v_1	v_2
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	+
+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	+
+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+
+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+
+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+
+	+	−	−	+	+	−	−	−	−	+	+	−	−	+	+	+	−
+	+	−	−	−	−	+	+	+	−	+	−	+	−	+	−	+	+
+	+	−	−	−	−	+	+	−	+	−	+	−	+	−	+	−	−
+	−	+	−	+	−	+	−	+	+	−	−	−	−	+	+	+	+
+	−	+	−	+	−	+	−	−	−	+	+	+	+	−	−	−	+
+	−	+	−	−	+	−	+	+	−	−	+	−	+	+	−	+	−
+	−	+	−	−	+	−	+	−	+	+	−	+	−	−	+	−	−
+	−	−	+	+	−	−	+	+	−	+	−	−	+	−	+	−	−
+	−	−	+	−	+	+	−	−	+	−	+	+	−	+	−	−	−
+	−	−	+	−	+	+	−	−	+	+	−	−	+	+	−	−	−

5. CONCLUSION

In this paper, we present a special class of supersaturated design called marginally over saturated design (MOSD), for which $k = n$ or $n + 1$. The fundamental theorems for constructing such designs are provided. Although examples are given here only for $n = 8, 12, 16$ and 20, in principle, the theorems

Table 5. Various Criteria Values, $n = 20$.

f	A	D	B_2	B_1	B_0
2	0.101	14.830	7.065	0.353	0.018
3	0.153	12.779	21.195	1.086	0.058
4	0.207	11.019	42.390	2.229	0.126
5	0.262	9.506	70.649	3.822	0.231
6	0.320	8.209	105.974	5.913	0.386
7	0.380	7.088	148.364	8.559	0.608
8	0.443	6.131	197.818	11.839	0.923
9	0.512	5.316	254.338	15.851	1.369
10	0.585	4.603	317.922	20.726	2.002
11	0.664	3.986	388.571	26.630	2.913
12	0.753	3.469	466.286	33.760	4.244
13	0.860	3.020	551.065	42.479	6.231
14	0.985	2.632	642.909	53.246	9.301
15	1.148	2.302	741.818	66.767	14.233

1	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	v_1	v_2
+	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	−
+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	−	+
+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	+
+	−	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+
+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+
+	+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	−	+
+	+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	+
+	−	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	+	+
+	+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−
+	−	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	+	−
+	+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	+
+	−	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	+
+	−	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−
+	−	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	+	−
+	−	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	−
+	+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	−	−
+	+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	−	−
+	−	+	+	−	−	+	+	+	+	−	+	−	+	−	−	−	−	+	+	−	−
+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

can be used to construct MOSD for any $n = 4t$, as long as the computing time permits.

Several optimality criteria has been discussed here. Because MOSD's are used mainly for screening purposes, the projection property plays an important role. Specifically, a main-effects only model consisting of only those active factors *must* be estimable. Since we do not know which factors are active in advance, the criterion of resolution rank seems most appropriate for evaluating the goodness of a MOSD.

Some possible extensions from this work are:

(1) a generalization of Theorems 1 and 2 to $k \geq n + 2$;

(2) a generalization of Theorems 1 and 2 to factors having more than two levels (a generalized Hadamard Matrix can be used as a starting point);

(3) a generalization of Theorems 1 and 2 to an arbitrary design matrix $\mathbf{D}$, rather than one with an orthogonal base defined as $\mathbf{H}$.

In conclusion, we live in a supersaturated world in which there are always more factors than we can handle. For screening a large numbers of factors, a supersaturated design have proven their worth. The false signal rates, present in most screening situations, will also increase with large k. In these cases where the experimental costs are moderate, and both Types I and II errors are of concern, we view the MOSD as a compromise between the classical saturated and supersaturated designs.

BIBLIOGRAPHY

Booth, K. H. V. and Cox, D. R. (1962), "Some Systematic Supersaturated Designs," *Technometrics*, **4**, 489-495.

Deng, L. Y, Lin, D. K. J. and Wang, J. (1996), "A Measurement of Multi-factor Orthogonality," *Statistics and Probability Letters* (to appear).

Deng, L. Y, Lin, D. K. J. and Wang, J. (1994), "Supersaturated design using Hadamard matrix,"IBM Watson Research Center Report, RC 19470 (84601).

Lin, D. K. J. (1991), "Systematic Supersaturated Designs,"Working Paper 264, University of Tennessee, College of Business Administration..

Lin, D. K. J. (1993a), "A New Class of Supersaturated Designs," *Technometrics,* **35**, 28-31.

Lin, D. K. J. (1993b), "Another Look at First-Order Saturated Designs: The p-efficient Designs," *Technometrics,* **35**, 284-292.

Lin, D. K. J. (1995), "Generating Systematic Supersaturated Designs," *Technometrics,* **37**, 213-225.

Srivastava, J. N. (1975), "Designs for Searching Non-negligible Effects", in: Survey of Statistical Design and Linear Models, North-Holland, Amsterdam, 507-519.

Tang, B. and Wu, C. F. J. (1993), "Construction of Supersaturated Designs ,"preprint.

Wu, C. F. J. (1993), "Construction of Supersaturated Designs Through Partially Aliased Interactions," *Biometrika,* **80**, 661-669.