A HIDDEN PROJECTION PROPERTY OF PLACKETT-BURMAN AND RELATED DESIGNS

J.C. Wang and C.F.J. Wu

IIQP Research Report
RR-93-01

January 1993

A HIDDEN PROJECTION PROPERTY OF PLACKETT-BURMAN AND RELATED DESIGNS

J. C. Wang and C.F.J. Wu

Department of Statistics and Actuarial Science
University of Waterloo
January 1993

ABSTRACT

Box and Hunter (1961) made an important observation that any fractional factorial design of resolution R has the property that when projected onto any R - 1 factors it becomes a full factorial design. This has a significant implication for statistical analysis. We observe that the Plackett-Burman and related designs have a hidden projection property with an analogous implication for the analysis. Because of complex aliasing, these designs have traditionally been used for screening main effects only. The hidden projection property suggests that complex aliasing actually allows some interactions to be entertained and estimated without making additional runs and provides a theoretical explanation for the success of an analysis strategy due to Hamada and Wu (1992). We give a detailed study of the hidden projection property for 12-run and 20-run designs with two levels and an 18-run design with three levels.

Key words and phrases: fractional factorial design, Plackett-Burman design, Hadamard matrix, orthogonal array, resolution, projection rationale, hidden projection, interaction, D efficiency, D_s efficiency.

1 Introduction

In the classic work of Box and Hunter (1961), they defined the resolution of a fractional factorial design to be the shortest wordlength R of the defining contrasts for the design. They further noted that the notion of resolution has an interesting statistical interpretation, that is, when the design is projected onto any subset of $R-1$ factors, it is a full factorial design. Therefore all the main effects and interactions among $R-1$ factors are orthogonal and estimable. Although this geometric projection property is obvious from the notion of strength in Rao's (1947) earlier work on orthogonal arrays, its full statistical implications were only recognized and exploited by Box and Hunter. Lin and Draper (1992), and independently Box and Bisgaard (1992), extended this projection property to Plackett-Burman (PB) designs, which cannot be defined by a group of defining contrasts. (Henceforth we will refer to Lin and Draper (1992) as LD.) Take the 12-run PB design as an example. LD showed that when projected onto any four factors, the design needs one additional run to make it into a 2^{4-1} design with resolution IV and five additional runs to make it into a full factorial design.

The projective rationale has two aspects. The first and the obvious one is the geometric projection, which was adopted by LD. The second and the more important aspect is the ability to entertain the estimation of interactions. For the usual 2^{n-k} designs (i.e., n factors with 2^{n-k} runs) with defining contrasts, these two aspects are equivalent. For the PB designs, however, the second aspect can be achieved even when geometric projection does not lead to a full factorial design or a fractional factorial design with high resolution . For the same example, we show in Section 2 that all the six two-factor interactions (2fi's) among the four factors can be estimated *without* adding any additional runs. By contrast the 2^{4-1} design obtained by adding one run has the defining relation $\mathbf{I} = \mathbf{1234}$ and only allows three out of the six 2fi's to be estimated because the six 2fi's are aliased in three pairs. A design is said to possess a *hidden projection* property if it allows some (or all) interactions to be estimated even when the projected design does not have the right resolution or other combinatorial design property for the same interactions to be estimated. For the PB designs their hidden projection property is a result of the complex aliasing patterns between the interactions and the main effects. For the 12-run design with four factors **1**, **2**, **3**, and **4**, any 2fi, say the interaction between **1** and **2**, is orthogonal to the main effects **1** and **2**, and is partially aliased with the main effects **3** and **4** with correlation $1/3$ or $-1/3$. Any other 2fi enjoys a similar property. No 2fi is fully aliased with any main effect, thus making it possible for all six 2fi's to be estimated along with the four main effects.

The hidden projection property provides a theoretical explanation for the success of an analysis strategy due to Hamada and Wu (1992) for entertaining and estimating interactions from PB-type experiments. Because of the complex aliasing, the PB designs have been used traditionally as a screening design, i.e., for estimating main effects only. The hidden projection property suggests that some interactions can be entertained and estimated without making additional runs at other settings.

In Section 2 we give a detailed analysis of the hidden projection property of the 12-run PB design for projections onto 3 to 7 factors. The results show that the number of estimable 2fi's is equal or close to the maximum degrees of freedom remaining for interactions. That is, no additional runs are necessary if only a moderate number of 2fi's are to be entertained. In Section 3 we extend the study to three 20-run designs including the PB design. The hidden projection property can also be applied to designs with more than two levels. An important example is the 18-run orthogonal array with seven 3-level factors. In Section 4 we study the hidden projection property of this array. By following the same approach we can study the hidden projection property for many other designs with complex aliasing. Finally we show in Section 5 how the hidden projection rationale can be exploited in data analysis and for run size savings. We demonstrate its advantages by reanalyzing the data in LD. In particular, we can identify the same 2fi as in the analysis of LD by adding only one run instead of the six runs required by the geometric projection approach.

2 Hidden projections of the 12-run Plackett-Burman design

Consider the 12-run PB design in Table 1. When collapsed onto any three factors, it consists

Table 1: 12-run Plackett-Burman Design.

run	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1	+	+	−	+	+	+	−	−	−	+	−
2	−	+	+	−	+	+	+	−	−	−	+
3	+	−	+	+	−	+	+	+	−	−	−
4	−	+	−	+	+	−	+	+	+	−	−
5	−	−	+	−	+	+	−	+	+	+	−
6	−	−	−	+	−	+	+	−	+	+	+
7	+	−	−	−	+	−	+	+	−	+	+
8	+	+	−	−	−	+	−	+	+	−	+
9	+	+	+	−	−	−	+	−	+	+	−
10	−	+	+	+	−	−	−	+	−	+	+
11	+	−	+	+	+	−	−	−	+	−	+
12	−	−	−	−	−	−	−	−	−	−	−

of two parts: a 2^3 design with eight points and a 2^{3-1} design with four points (see LD). For four factors, in addition to the four main effects, there are still seven degrees of freedom left. Can they be used to estimate the six 2fi's without adding more runs? It turns out that any four columns of the matrix in Table 1 can be chosen for the four factors because it is known (Draper, 1985; Wang, 1989) that, except for $n = 5$ and 6, any $12 \times n$ submatrices are isomorphic. (Any two matrices of 1 and -1's are isomorphic if one can be obtained from the other by permutations of rows, columns and sign changes.)

For the main effects and a given set of 2fi's, we can use the following D criterion for measuring the overall efficiency for estimating the collection of effects:

$$|X^t X|^{1/k}, \tag{1}$$

where $X = [\boldsymbol{x}_1/||\boldsymbol{x}_1||, \ldots, \boldsymbol{x}_k/||\boldsymbol{x}_k||]$, and $\boldsymbol{x}_i$ is the coefficient vector of the i-th effect. Because the columns of X are standardized, (1) achieves its maximum 1 if and only if the $\boldsymbol{x}_i$'s are orthogonal to each other, i.e., when the array is orthogonal. The vector **1** is not included in X since it is orthogonal to the $\boldsymbol{x}_i$'s. For the estimation of each individual effect, we use the following D_s criterion for measuring its efficiency:

$$\{\boldsymbol{x}_i^t \boldsymbol{x}_i - \boldsymbol{x}_i^t X_{(i)} (X_{(i)}^t X_{(i)})^{-1} X_{(i)}^t \boldsymbol{x}_i\} / \boldsymbol{x}_i^t \boldsymbol{x}_i, \tag{2}$$

which attains its upper bound 1 if and only if $\boldsymbol{x}_i$ is orthogonal to the other columns in X.

For $n = 4$, we give in Table 2 the values of D and D_s for 10 cases. The last one consisting of four main effects and six 2fi's is the most comprehensive. The other nine are its submodels. Therefore we can estimate all the 2fi's and the main effects without adding runs.

Table 2: Estimation efficiency for four factors and h interactions, $h = 1, \ldots, 6$. The D_s efficiency is given for each effect.

	D	effect									
case	value	**1**	**2**	**3**	**4**	**12**	**13**	**14**	**23**	**24**	**34**
1	.95	1	1	.88	.88	.78					
2	.92	1	.87	.87	.75	.76	.76				
3	.89	.85	.85	.85	.85	.63					.63
4	.89	1	.74	.74	.74	.74	.74	.74			
5	.89	.87	.87	.87	.62	.74	.74		.74		
6	.87	.85	.85	.74	.74	.76	.63			.63	
7	.85	.85	.73	.73	.62	.73	.73	.62	.62		
8	.83	.73	.73	.73	.73	.63	.63			.63	.63
9	.82	.72	.72	.62	.62	.72	.62	.62	.62	.62	
10	.80	.62	.62	.62	.62	.62	.62	.62	.62	.62	.62

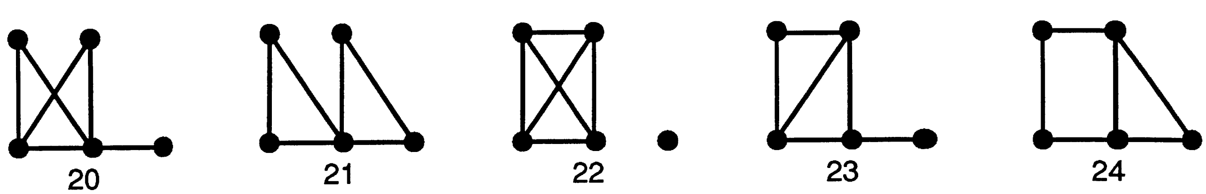
For $n = 5$, there are two non-isomorphic 12×5 submatrices: design 5.1 and design 5.2 in the notation of LD. Design 5.1 has two repeated runs, i.e., two runs with the same level combination. For example, in the design consisting of columns **1**, **2**, **3**, **4**, and **10**, runs 3 and 11 are identical. On the other hand, design 5.2 has two mirror image runs. For example, runs 7 and 11 are two mirror runs in the design consisting of columns **1** to **5**. As we show later, design 5.2 can entertain more models than design 5.1. Therefore we present the results primarily for design 5.2. In Table 3, there are altogether 24 cases. The last five contain six 2fi's in addition to the five main effects. The 2fi's

Table 3: Estimation efficiency of design 5.2 (columns **1**, **2**, **3**, **4**, and **5**) for five factors and h interactions, $h = 1, \ldots, 6$. The D_s efficiency is given for each effect.

	D	effect														
case	value	**1**	**2**	**3**	**4**	**5**	**12**	**13**	**14**	**15**	**23**	**24**	**25**	**34**	**35**	**45**
1	.93	1	1	.86	.86	.86	.67									
2	.89	1	.86	.86	.75	.75	.67	.67								
3	.88	.84	.84	.84	.8	.84	.59								.59	
4	.85	1	.74	.67	.74	.6	.59	.67	.59							
5	.85	.84	.86	.84	.6	.6	.59	.67			.59					
6	.85	.84	.84	.74	.74	.71	.67	.59				.59				
7	.83	.83	.77	.67	.77	.67	.56		.56						.56	
8	.81	1	.59	.59	.59	.59	.59	.59	.59	.59						
9	.81	.84	.73	.59	.73	.57	.59	.59	.59			.59				
10	.81	.83	.68	.67	.68	.6	.56	.67	.56						.56	
11	.74	.7	.42	.7	.54	.54	.52	.31			.52					.26
12	.81	.73	.73	.73	.67	.73	.59	.59					.59		.59	
13	.79	.67	.75	.67	.59	.59	.53		.53		.53				.53	
14	.78	.83	.59	.56	.59	.56	.59	.56	.59	.56		.56				
15	.78	.72	.72	.59	.59	.51	.56	.59	.56		.56	.59				
16	.77	.67	.67	.67	.53	.53	.53	.67	.53		.53				.53	
17	.71	.67	.33	.51	.53	.51	.51	.3	.51		.51					.24
18	.71	.53	.29	.51	.67	.53	.51	.3	.51						.51	.24
19	.76	.58	.58	.58	.58	.58	.52		.52		.52				.52	.52
20	.69	.5	.5	.5	.2	.5	.5	.5	.5	.22	.2				.5	
21	.69	.33	.5	.5	.5	.5	.5	.5	.5	.5		.2			.2	
22	.69	.5	.5	.5	.5	.17	.2	.5	.5		.5	.5		.2		
23	.69	.29	.5	.5	.5	.5	.5	.5	.5		.5	.22			.2	
24	.69	.5	.29	.5	.5	.5	.5	.29	.5		.5				.5	.22

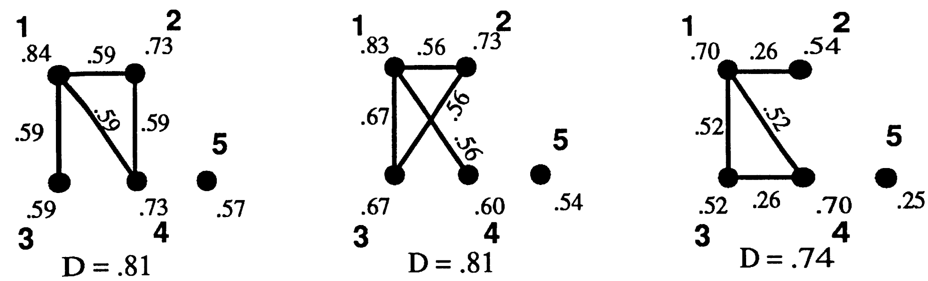
for these five cases can be graphically represented as in Figure 1. The remaining 19 cases can be

Figure 1: Graph representation of cases 20 to 24 of Table 3.



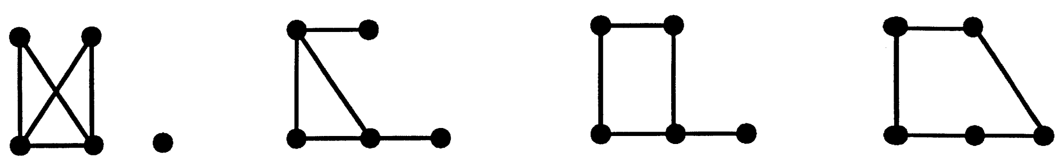
represented as subgraphs of these five graphs. In Figure 1 we do not give the column numbers of the factors because there can be several choices for the same graph. Take, for example, case 9 in Table 3. There are two other models for the same graph, but with different D and D_s values. They are shown in Figure 2 with the leftmost graph being case 9 in Table 3. To save space, we give only one model with good values of D and D_s for each unlabelled graph. The same is done for the rest of the paper.

Figure 2: Three isomorphic graphs with different models. (A main effect is represented by a node and a 2fi by a line connecting two nodes. The D_s value for an effect is given next to a node or line.)



Note that design 5.1 is inferior to design 5.2. Because design 5.1 has two repeated runs, there are only 10 degrees of freedom for estimating effects. Consequently, it can entertain at most five 2fi's. It can entertain models represented by 15 non-isomorphic graphs, all of which are, however, subgraphs of the 24 graphs for design 5.2. Furthermore, for the same model, design 5.2 has higher efficiencies than design 5.1. Among the 15 graphs, four have five 2fi's (in addition to the five main effects). They are given in Figure 3 and correspond to cases 15, 16, 18, and 19 of Table 3 respectively.

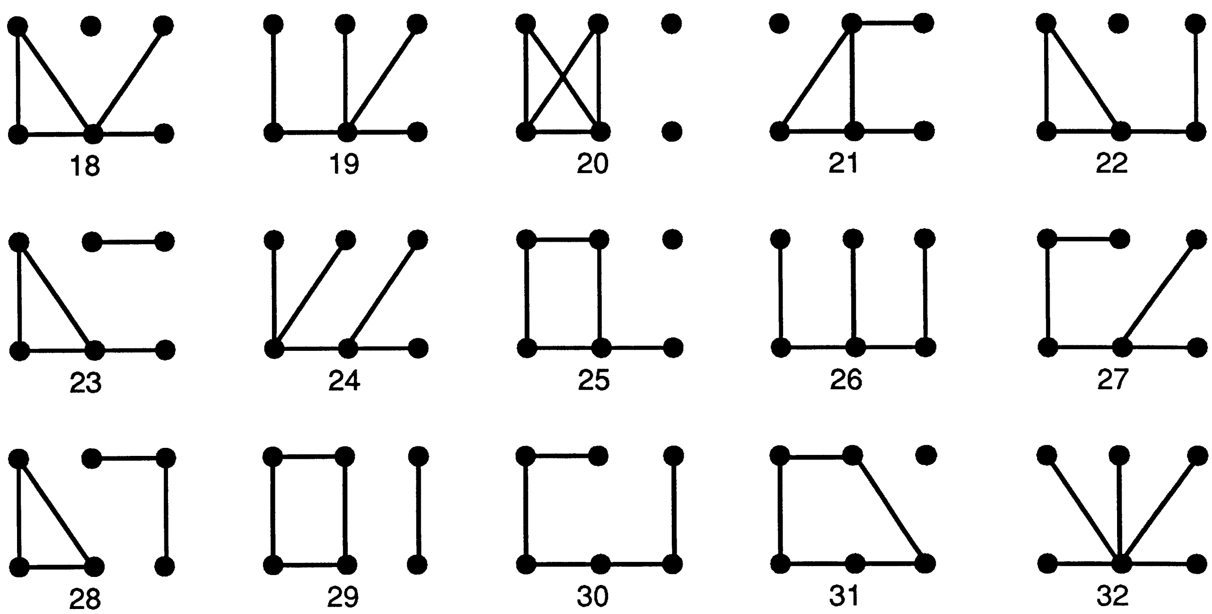
Figure 3: Graphs for design 5.1.



By contrast the geometric projection approach would require adding six and respectively ten runs for designs 5.1 and 5.2 so that the augmented designs have resolution V (see LD). If only six or fewer 2fi's are to be entertained, the hidden projection approach will be preferred.

For $n = 6$, there are two non-isomorphic 12×6 submatrices (Lin and Draper, 1992). Design 6.1 is characterized by having no mirror runs (e.g. columns **1** to **6**) while design 6.2 has two mirror image runs (e.g. runs 7 and 11 are mirror image runs in columns **1** to **5** and **7**). We only consider design 6.1 for reasons given below. There are altogether 32 cases given in Table 4. The last 15 cases has five 2fi's in addition to six main effects. Their graphs are given in Figure 4. The remaining cases can be viewed as their subgraphs.

Figure 4: Graph representation of cases 18 to 32 of Table 4.



Note that design 6.1 is superior to design 6.2 for the following reasons.

1. The two mirror image runs in design 6.2 provide no information about any 2fi because any 2fi column in the two runs is either $(1,1)^t$ or $(-1,-1)^t$, and therefore is confounded with the grand mean.

2. Design 6.2 can entertain models represented by 20 non-isomorphic graphs all of which are subgraphs of the 32 graphs for design 6.1.

Table 4: Estimation efficiency of design 6.1 (columns **1** to **6**) for six factors and h 2-factor interactions, $h = 1, \ldots, 5$. The D_s efficiency is given for each effect. (The interaction is given in the parenthesis after its D_s value.)

	D	main effect						2-factor interaction
case	value	**1**	**2**	**3**	**4**	**5**	**6**	
1	.92	1	1	.83	.83	.83	.83	.56(**12**)
2	.86	1	.83	.83	.67	.75	.67	.53(**12**), .53(**13**)
3	.86	.83	.83	.83	.75	.83	.75	.53(**12**), .53(**35**)
4	.81	1	.74	.67	.67	.58	.58	.52(**12**), .52(**13**), .52(**14**)
5	.81	.82	.82	.82	.58	.58	.58	.52(**12**), .52(**13**), .52(**23**)
6	.81	.82	.74	.82	.58	.74	.58	.52(**12**), .52(**13**), .52(**35**)
7	.81	.82	.74	.74	.67	.7	.67	.52(**12**), .52(**13**), .52(**46**)
8	.76	.8	.44	.73	.73	.73	.73	.44(**12**), .3(**35**), .3(**46**)
9	.77	.58	1	.58	.58	.58	.5	.51(**12**), .51(**23**), .51(**24**), .51(**25**)
10	.77	.82	.73	.67	.58	.5	.53	.51(**12**), .51(**13**), .51(**14**), .51(**23**)
11	.77	.82	.68	.67	.58	.58	.53	.51(**12**), .51(**13**), .51(**14**), .51(**35**)
12	.72	.8	.44	.7	.31	.7	.31	.44(**12**), .25(**13**), .25(**15**), .44(**46**)
13	.72	.4	.7	.7	.52	.57	.52	.44(**12**), .55(**13**), .25(**23**), .25(**46**)
14	.77	.67	.67	.67	.67	.5	.67	.51(**13**), .51(**14**), .51(**36**), .51(**46**)
15	.77	.67	.73	.58	.58	.67	.67	.51(**12**), .51(**13**), .51(**26**), .51(**46**)
16	.72	.8	.44	.7	.52	.7	.52	.44(**12**), .44(**13**), .25(**35**), .25(**46**)
17	.67	.42	.53	.67	.53	.67	.3	.3(**12**), .22(**23**), .22(**45**), .3(**46**)
18	.69	.5	.67	.5	.5	.5	.24	.44(**12**), .22(**13**), .44(**23**), .22(**24**), .44(**26**)
19	.69	.8	.44	.5	.29	.5	.24	.44(**12**), .22(**13**), .44(**14**), .22(**15**), .44(**46**)
20	.69	.5	.67	.5	.5	.5	.2	.44(**12**), .22(**13**), .44(**14**), .44(**23**), .22(**24**)
21	.74	.67	.67	.67	.5	.5	.5	.5(**12**), .5(**13**), .5(**14**), .5(**23**), .5(**35**)
22	.69	.36	.67	.5	.5	.5	.5	.44(**12**), .44(**13**), .44(**14**), .22(**23**), .22(**46**)
23	.61	.5	.33	.29	.5	.5	.2	.22(**12**), .22(**13**), .22(**14**), .22(**24**), .11(**56**)
24	.61	.5	.22	.5	.29	.29	.29	.22(**12**), .22(**13**), .22(**15**), .22(**34**), .22(**36**)
25	.69	.5	.67	.5	.29	.5	.5	.22(**12**), .44(**13**), .44(**14**), .22(**36**), .44(**46**)
26	.69	.8	.44	.5	.5	.5	.5	.44(**12**), .44(**13**), .44(**14**), .22(**35**), .22(**46**)
27	.61	.5	.33	.5	.29	.29	.29	.22(**12**), .22(**13**), .22(**25**), .22(**34**), .22(**36**)
28	.61	.29	.33	.29	.5	.5	.29	.22(**12**), .22(**23**), .22(**45**), .22(**46**), .22(**56**)
29	.61	.5	.33	.29	.5	.29	.29	.22(**12**), .22(**13**), .22(**24**), .22(**34**), .22(**56**)
30	.61	.5	.33	.29	.5	.5	.29	.22(**12**), .22(**13**), .22(**24**), .22(**45**), .22(**56**)
31	.61	.29	.17	.29	.5	.5	.29	.22(**13**), .22(**14**), .22(**35**), .11(**46**), .11(**56**)
32	.74	.5	1	.5	.5	.5	.5	.5(**12**), .5(**23**), .5(**24**), .5(**25**), .5(**26**)

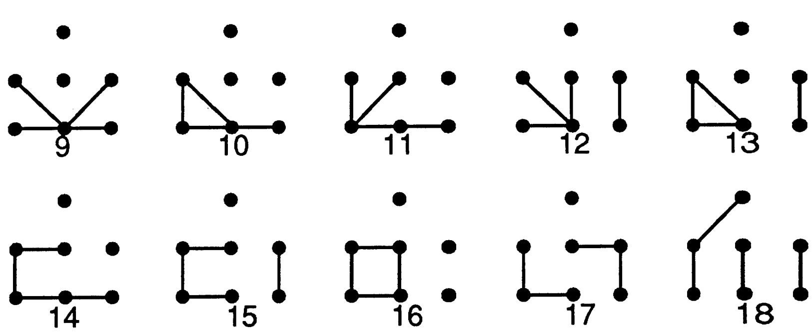
3. For the same model, design 6.1 has higher efficiencies than design 6.2.

For $n = 7$, all 12×7 submatrices are isomorphic. So we choose the first seven columns for the factors. The results are given in Table 5. Each of the last 10 cases has four 2fi's in addition to seven main effects. Their graphs are given in Figure 5. The remaining cases can be viewed as their subgraphs. With the exception of graph no. 18, they are subgraphs of those in Figure 4.

Table 5: Estimation efficiency for seven factors and h interactions, $h = 1, \ldots, 4$. The D_s efficiency is given for each effect. (Except for **12**, the interaction is given in the parenthesis after its D_s value.)

	D	effect								
case	value	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**12**	
1	.9	1	1	.8	.8	.8	.8	.8	.44	
2	.84	1	.8	.67	.8	.67	.67	.67	.44	.44(**14**)
3	.84	.8	.8	.8	.67	.8	.67	.67	.44	.44(**35**)
4	.78	1	.67	.57	.67	.67	.57	.57	.44	.44(**14**), .44(**15**)
5	.78	.8	.8	.57	.8	.57	.57	.57	.44	.44(**14**), .44(**24**)
6	.78	.8	.8	.67	.67	.57	.57	.57	.44	.44(**14**), .44(**23**)
7	.78	.8	.67	.67	.67	.67	.57	.57	.44	.44(**14**), .44(**35**)
8	.78	.67	.67	.67	.67	.67	.57	.67	.44	.44(**35**), .44(**47**)
9	.74	1	.57	.5	.57	.57	.5	.57	.44	.44(**14**), .44(**15**), .44(**17**)
10	.74	.8	.67	.5	.67	.57	.5	.5	.44	.44(**14**), .44(**15**), .44(**24**)
11	.74	.8	.57	.57	.57	.67	.5	.5	.44	.44(**14**), .44(**15**), .44(**35**)
12	.66	.67	.67	.67	.4	.25	.29	.29	.22	.22(**13**), .22(**14**), .22(**67**)
13	.66	.67	.67	.5	.29	.5	.5	.25	.22	.22(**14**), .22(**24**), .11(**36**)
14	.74	.67	.67	.67	.57	.57	.5	.5	.44	.44(**14**), .44(**23**), .44(**35**)
15	.74	.67	.57	.57	.67	.57	.5	.57	.44	.44(**14**), .44(**35**), .44(**47**)
16	.74	.67	.67	.67	.5	.5	.5	.67	.44	.44(**17**), .44(**23**), .44(**37**)
17	.66	.5	.4	.4	.4	.67	.5	.5	.22	.22(**14**), .22(**36**), .22(**56**)
18	.66	.67	.67	.67	.4	.18	.5	.5	.22	.22(**14**), .22(**35**), .11(**67**)

Figure 5: Graph representation of cases 9 to 18 of Table 5.



3 Hidden projections of three 20-run designs

According to Hall (1965), there are three nonisomorphic Hadamard matrices of order 20, which he called class N, P, and Q. Our computer search shows that class Q is equivalent to the cyclic design studied by Plackett and Burman (1946), a fact not pointed out by Hall. Note that in Table 6,

Table 6: Class Q Hadamard matrix of order 20.

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

the column (labelled **0**) consisting of all +'s cannot be used to study a factor effect. Similarly, for Classes N and P, only 19 columns can be used for studying factor effects.

For each of the three designs, projection onto any three columns consists of at least one 2^3, thus allowing all the factorial effects to be estimated with high efficiency. Details on the geometric projection property for $n = 3$ can be found in LD.

For $n = 4$ and 5, the hidden projection approach reveals some interesting aspects that are missed by the geometric projection approach. First we review the geometric projection property.

According to LD, for each of N, P, and Q, there are three nonisomorphic 20×4 submatrices:

1. Design 20-4.1. It has five runs with two repeats. As a result it has 15 out of the 16 ($= 2^4$) level combinations.
2. Design 20-4.2. It has one run with three repeats and six runs with two repeats and therefore has only 12 out of the 16 level combinations. One of the four possible projections onto three columns forms a design consisting of a 2^3 and three replicates of a 2^{3-1} with resolution III. Each of the remaining projections forms a design (denoted 20-3.1) which consists of a 2^{3-1} with resolution III and two replicates of a 2^3.
3. Design 20-4.3. It has one run with three repeats and six runs with two repeats and therefore has only 12 out of the 16 level combinations. Each of its projections onto three columns forms a design 20-3.1.

Designs 20-4.1 to 20-4.3 require the addition of 1, 4, and 4 runs respectively to complete a full factorial 2^4. If we are primarily interested in estimating the four main effects and six 2fi's, there is *no* need to add runs for estimating the ten effects. Using the hidden projection property, we have the estimation efficiencies for design 20-4.1: $D = .93$, $D_s = .86$ for each of the ten effects. For 20-4.2 (using columns **1**, **2**, **3**, **4**), $D = .8$, $D_s = .81$ for **4**, **14**, **24**, **34**, and $D_s = .53$ for **1**, **2**, **3**, **12**, **13**, **23**. For design 20-4.3, $D = .8$, $D_s = .81$ for each main effect, $D_s = .53$ for each 2fi. So if the 3-factor and 4-factor interactions are of little interest, which is usually the case, the hidden projection property would allow us to estimate the 2fi's without adding more runs.

Among the three projections, 20-4.1 is the best in terms of the geometric projection as well as the hidden projection property. For each of N, P, and Q, 20-4.1 appears 2736 times, 20-4.2 912 times, and 20-4.3 228 times. So there is no difference among N, P, and Q when $n = 4$.

For $n = 5$, there are respectively 10, 10, 9 nonisomorphic 20×5 submatrices for N, P, and

Q corresponding to the same matrices, which Lin and Draper (1991) called design 20-5.1 to 20-5.10 with design 20-5.10 not included in the collection for Q. (For brevity sake, we drop 20 in 20-5.i in the remainder of the section. To save space we refer to their paper for these designs.) The number of additional runs required to make a full factorial 2^5 ranges from 12 to 19 (Lin and Draper, 1991). If, however, the 2fi's are of primary interest, we need to add fewer runs to complete a 2^{5-1} design defined by $\mathbf{5} = \pm\mathbf{1234}$ because in any resolution V design, all the 2fi's are estimable. Our calculations show that, among the ten 20×5 submatrices, one requires adding three runs, two require adding four runs, and the rest require adding six to nine runs. By exploiting the hidden projection property, we can estimate most or all of the 2fi's without adding any run.

Straight but tedious calculations show that for designs 5.1, 5.4, 5.3 and 5.5, all the 10 2fi's are estimable with 5.1 and 5.4 having higher overall estimation efficiencies than 5.3 and 5.5. For designs 5.2, 5.6, 5.7 and 5.8, nine 2fi's are estimable while for designs 5.9 and 5.10, only seven are estimable. The results can be explained by the structures of the designs. Design 5.1 has no run with repeats, 5.4 has one run with two repeats, and the rest have at least two runs with repeats. The worst are designs 5.9 and 5.10 with six and seven runs respectively with repeats.

We can compare N, P, and Q in term of the frequencies of the best designs 5.1 and 5.4 among the projections. The best is Q, which has 1881 projections of design 5.1 and 1368 projections of design 5.4, while N has 1680 of design 5.1 and 1488 of design 5.2 and P has 1296 of design 5.1 and 1728 of design 5.2. Recall that Q does not have the "worst" design 5.10 among its projections, which may partially explain its superiority.

We conclude the section with a summary of results for $n = 6$. There are 59, 56 and 50 nonisomorphic 20×6 submatrices for N, P and Q respectively. The complexity may explain why Lin and Draper (1991) did not study the geometric projection for $n = 6$. Among the 59 submatrices, 20 have no repeated runs. At most 13 2fi's can be estimated. In Table 7 we give the percentages and cumulative percentages of projections that allow h 2fi's to be estimated, $h = 13, 12, 11, 10, 7$. Since at least 99.7% of the projections will allow 10 or more 2fi's to be estimated (with average efficiencies approximately $D = .65$, $D_s = .37$ for main effects, and $D_s = .32$ for 2fi's), the hidden projection property suggests that usually no additional runs are needed for studying important 2fi's. In practice, the number of important 2fi's seldom exceeds six. Design **Q** is again the best among N, P and **Q** because all of its 20×6 submatrices can entertain at least 10 2fi's!

Table 7: Percentages and cumulative percentages (in parentheses) of 20×6 submatrices of N, P and Q that can entertain h 2fi's, $h = 13, 12, 11, 10, 7$.

	h				
design	13	12	11	10	7
N	33.0(33.0)	47.9(80.9)	12.4(93.3)	6.6(99.9)	0.1(100)
P	39.0(39.0)	42.1(81.1)	11.2(92.3)	7.4(99.7)	0.3(100)
Q	29.4(29.4)	51.5(80.9)	13.2(94.1)	5.9(100)	

4 Hidden projections of $L_{18}(3^7)$

The hidden projection property also holds for three-level designs with complex aliasing. To save space we only consider the orthogonal array $L_{18}(3^7)$ (Masuyama, 1957) given in Table 8. This array plays an ubiquitous role in practical experimentation and in theoretical research because it is the smallest orthogonal array with three levels and complex aliasing.

For three factors (i.e. $n = 3$), there are three nonisomorphic 18×3 submatrices given as follows.

1. Design 18-3.1. It is a $\frac{2}{3}$ fraction of 3^3, consisting of one $\frac{1}{3}$ fraction defined by $\mathbf{C} = \mathbf{A} + \mathbf{B}$ (mod 3), and another $\frac{1}{3}$ fraction defined by $\mathbf{C} = \mathbf{A} + \mathbf{B} + 1$ (mod 3), where **A** and **B** are any two of the three columns. Any three columns in Table 8 not containing column **1** form a design of this type.

2. Design 18-3.2. It consists of two $\frac{1}{3}$ fractions of 3^3 in which the two fractions share three points in common. For instance, in columns **1**, **2**, and **7**, run 7 through run 15 form a $\frac{1}{3}$ fraction

Table 8: 18-run Orthogonal Array.

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

defined by col **7** = col **2** + col **1** (mod 3), and the remaining runs form another $\frac{1}{3}$ fraction defined by col **7** = col **2** + 2 · col **1** (mod 3), where col is an abbreviation for column. These two fractions share (0,0,0), (1,0,1), and (2,0,2).

3. Design 18-3.3. It contains two identical replicates of a $\frac{1}{3}$ fraction of 3^3. Only one choice of three columns (columns **1**, **3**, and **4**) is of this type.

Obviously design 18-3.1 is the best because it has 17 df's for estimating effects while designs 18-3.2 and 18-3.3 have, respectively, 14 and 8 df's. For efficiency comparison we consider design 18-3.1 with quantitative factors only. For each quantitative factor, we use $\ell = (-1, 0, 1)$ for its linear effect and $q = (1, -2, 1)$ for its quadratic effect. The four df's for each 2fi are represented by the linear-by-linear $(\ell \times \ell)$, linear-by-quadratic $(\ell \times q)$, quadratic-by-linear $(q \times \ell)$, and quadratic-by-quadratic $(q \times q)$ effects. The model consists of the grand mean, the three main effects **A**, **B**, **C** (with 6 df's) and the $\ell \times \ell$, $\ell \times q$, and $q \times \ell$ components of **A**×**B**, **B**×**C**, **A**×**C** (with 9 df's). For this model, the overall D efficiency is 0.83. The individual D_s efficiencies are: 0.78 for ℓ, 0.76 for q, 0.48 for $\ell \times \ell$, and 0.6 for $\ell \times q$ and $q \times q$. Because the model has 16 df's, there are two remaining df's, which allow for estimating two of the three $q \times q$ effects.

For four factors (i.e. $n = 4$), there are four nonisomorphic 18×4 submatrices given as follows:

1. Design 18-4.1. Any four columns not containing column **1** form a design of this type. Note that any three columns of this design form a design 18-3.1.

2. Design 18-4.2. One set of its three columns is a design 18-3.2 and the remaining three sets are design 18-3.1. Columns **1**, **2**, **3**, and **6** form a design of this type.

3. Design 18-4.3. One set of its three columns is a design 18-3.3 and the remaining three sets are design 18-3.1. Columns **1**, **2**, **3**, and **7** form a design of this type.

4. Design 18-4.4. One set of its three columns is a design 18-3.1 and the remaining three sets are design 18-3.2. Columns **1**, **2**, **4**, **7** form a design of this type.

Any of the four designs allows the four main effects (8 df's altogether) and the $\ell \times \ell$ components of the six interactions to be estimated. Their efficiencies are given in Table 9. Overall designs 18-4.1 and 18-4.2 are better than the other two. Design 18-4.1 is slightly better than design 18-4.2 because its D_s-efficiencies are less variated.

For $n = 5$ the same method can be used to find out how many $\ell \times \ell$ effects (in addition to the main effects) can be estimated. The details are omitted.

5 An Illustrative Example

In this section we illustrate the hidden projection approach to analysis by reanalyzing an example reported in LD. To save space we refer to LD for the data and design matrix. Initially a 12-run

Table 9: D and D_s efficiencies for 18 runs in 4 factors.

design		effect													
		linear				quadratic				linear×linear					
	D	A	B	C	D	A	B	C	D	AB	AC	AD	BC	BD	CD
18-4.1	.82	.62	.62	.62	.62	.87	.87	.87	.87	.53	.54	.54	.54	.54	.53
18-4.2	.84	.85	.78	.70	.85	.61	.91	.72	.74	.55	.49	.49	.51	.58	.57
18-4.3	.73	.44	.27	.54	.27	.65	.82	.78	.82	.33	.30	.33	.58	.35	.58
18-4.4	.72	.42	.51	.75	.51	.41	.57	.47	.57	.33	.25	.33	.49	.27	.49

Note. We use columns **2**, **3**, **4**, **5** for 18-4.1, columns **1**, **2**, **3**, **6** for 18-4.2, columns **1**, **2**, **3**, **4** for 18-4.3, and columns **1**, **2**, **4**, **7** for 18-4.4.

PB design was used to study 10 factors. Using standard analysis they identified five significant main effects **1**, **3**, **7**, **8** and **10**. Then they added six more runs, labelled 13 to 18, to make it a resolution V design. Based on the augmented data, they found the **7**×**8** interaction to be significant.

Using the hidden projection property, we can entertain and analyze some 2fi's based on the original 12 runs. Since columns **1**, **3**, **7**, **8** and **10** form design 5.1, the collapsed design can entertain at most five 2fi's whose graphs are given in Figure 3. It is easy to see from the graphs that any set of three 2fi's among the five factors are estimable. So we can study any three 2fi's without adding runs. Using forward selection in regression analysis with the five main effects and the 10 2fi's as candidate variables, we did not find any significant 2fi since their partial t values are small. To better estimate the 2fi's, we need to add runs. Since design 5.1 has 11 distinct runs, there are 21 remaining factor level combinations in the 2^5 design. Suppose the objective is to be able to estimate one more 2fi, say, xy. For choosing the additional run, we use the D criterion for the overall model (five main effects plus xy) and the D_s criterion for estimating xy. It turns out that the combination in run no. 14 of LD is the best, with $D = .96$ and $D_s = .83$ for any xy. (Note that run no. 14 is the mirror image of the two repeated runs, no. 5 and no. 10 in the design matrix of LD.) By adding this run to the original data and repeating the forward selection for model search, the variables are entered in the order: **10**, **7**, **8**, **1**, **3**, **7**×**8** and the partial t value for **7**×**8** is 2.55.

The fitted model is

$$\hat{y} = 72.3 + 22.1x_{10} + 16.4x_7 + 11.5x_8 - 8.9x_1 - 6.4x_3 + 2.4x_7x_8$$

and the adjusted R^2 increases from 98.3% to 99.1% by adding x_7x_8. So we can reproduce results and conclusions very close to those in LD by adding only one run.

Acknowledgements

This research is supported by the Natural Sciences and Engineering Research Council of Canada, General Motors of Canada and the Manufacturing Research Corporation of Ontario.

References

Box, G. E. P. and Hunter, J. S. (1961). The 2^{k-p} Fractional Factorial Designs. *Technometrics*, **3**, 311–351, 449–458.

Box, G. E. P. and Bisgaard, S. (1993). George's Column. *Quality Engineering*, to appear.

Draper, N. R. (1985). Small Composite Designs. *Technometrics*, **27**, 173–180.

Hall, M. Jr. (1965). Hadamard Matrix of Order 20. Jet Propulsion Laboratory Technical Report No. 32-761.

Hamada, M. and Wu, C. F. J. (1992). Analysis of Designed Experiments with Complex Aliasing. *Journal of Quality Technology*, **24**, 130–137.

Lin, D. K. J. and Draper, N. R. (1991). Projection Properties of Plackett and Burman Designs. Technical Report 885, University of Wisconsin, Dept. of Statistics.

Lin, D. K. J. and Draper, N. R. (1992). Projection Properties of Plackett and Burman Designs. *Technometrics*, **34**, 423–428.

Masuyama, M. (1957). On Difference Sets for Constructing Orthogonal Arrays of Index Two and Strength Two. *Reports of Statistical Applications Research, JUSE*, **5**, 27–34.

Plackett, R. L. and Burman, J. P. (1946). The Design of Optimum Multifactorial Experiments. *Biometrika*, **33**, 305–325.

Rao, C. R. (1947), "Factorial Experiments Derivable from Combinational Arrangements of Arrays," *Journal of the Royal Statistical Society*, Supplementary, 9, 128–139.

Wang, J. C. (1989), "Orthogonal Arrays and Nearly Orthogonal Arrays with Mixed Levels: Construction and Applications," Ph. D. thesis, University of Wisconsin, Madison, Department of Statistics.