Optimal Incomplete Block Designs

J. P. Morgan

Department of Statistics, Virginia Tech

31st August 2005; revised 31st March 2006; revised 24th July 2006

Abstract: Optimal incomplete block designs are pursued through the E-criterion of minimizing maximal variance. Methodology is developed for design choice based on graphs and an extensive catalog of downloadable designs compiled. Along with the general methodology, a near complete solution for E-optimal block designs is provided for up to 15 treatments. E-optimality is revealed as a flexible criterion that, depending on the application, can offer many choices for good designs.

1 Introduction

Blocking to improve efficiency of comparative experiments, deemed one of the three fundamental principles of experimentation by Fisher (1937), is a tool in every statistician's kit, and a concept known by scientists across the experimental spectrum. Yet aside from relatively sparse, "standard" designs, there is still much that is not known about the implementation of blocking, and known designs are not easily accessed. This paper is a move towards correcting these shortfalls. In doing so, a rigorous structure for determination of E-optimal block designs is established.

That blocking is a powerful tool for removing nuisance variation is a firmly established tenet of statistical practice. Considerations leading to the decision to block have been examined in many contexts and are a staple of modern texts, e.g. Box, Hunter, and Hunter (2005), Dean and Voss (1999), Kuehl (2000), Oehlert (2000); why and how one blocks will not be rehashed here. Rather, the purpose is to address the design question that follows: having selected blocks of experimental units, one must still decide which units are to receive which treatments. That decision can be

driven by the variances it induces. The variance reduction achieved by blocking is not a priori measurable, nor is it affected by the treatment/unit assignment. But for any given blocking system, the variances of estimators (up to a scalar multiple) are a function *only* of that assignment, so that these variances *can* be a priori calculated. Design selection can be based upon summary measures of estimator variances.

Formally, a *block design* is just an assignment of v treatments to n experimental units that have been partitioned into b blocks of k units each. While the texts cited above illustrate that block designs appear in multitudinous applications, the unifying, underlying problem in choice of block design is succinctly stated: *given* (v, b, k)*, what is the best assignment of treatments to units?* Let $\mathcal{D}(v, b, k)$ be the class of all possible assignments for given (v, b, k), and use d to denote members of $\mathcal{D}$, that is, to denote available designs. The problem is to choose $d \in \mathcal{D}$.

Many papers have been written addressing this problem; see Shah and Sinha (1989) for numerous references and later in this paper. The partial answers found are typically theorems of the form "if conditions X are satisfied then design Y is optimal." Due to combinatorial difficulties in optimization over discrete spaces, there has been no comprehensive solution for a practical range of v for all practical b and k. The main theoretical advance here allows computational power to be harnessed in bypassing a large measure of the combinatorial complexity.

Evaluation of a block design begins with the linear model. For y_{ju} the yield from unit u in block j, the model is

$$y_{ju} = \mu + \beta_j + \tau_{d[ju]} + e_{ju} \tag{1}$$

where as usual μ is an overall mean, β_j the effect of block j, τ_i the effect of treatment i, and the e_{ju} are uncorrelated, equivariable (σ^2), random errors. The design function $d[ju]$ in (1) reports the treatment assigned to unit (ju) by design d.

Now let $N_d = (n_{dij})$ be the $v \times b$ treatment$\times$block incidence matrix determined by design d; n_{dij} is the number of units in block j assigned treatment i. Let r_{di} be the total number of units receiving treatment i. Then the information matrix for estimation of treatment effects is

$$C_d = Diag(r_{d1}, \ldots, r_{dv}) - \frac{1}{k} N_d N_d' \tag{2}$$

All treatment contrasts are estimable if and only if C_d has rank $v - 1$ (Chakrabarti, 1962). $\mathcal{D}$ is restricted to contain only these designs.

Any generalized inverse of C_d is a covariance matrix for estimation of treatment contrasts $l'\tau$, that is, $\text{var}(\widehat{l'\tau}) = \sigma^2 l' C_d^- l$. Let $z_{di} = \frac{1}{\mu_{di}}$ for the nonzero eigenvalues $\mu_{d1} \leq \mu_{d2} \leq \ldots \leq \mu_{d,v-1}$ of C_d. Aside from σ^2, the z_{di} are variances for a normalized set of orthogonal treatment contrasts. Accordingly, choice of design can be based on minimizing summary functions of the z_{di}. The average variance of any set of $v-1$ orthonormal contrasts is proportional to $\sum_i z_{di}$, and a design d minimizing this quantity is said to be A-optimal. D-optimal designs minimize $\Pi_{i=1}^{v-1} z_{di}$, this being proportional to the volume of the confidence ellipsoid for any $v-1$ orthonormal contrasts. The primary focus of this paper will be minimization of the E-criterion z_{d1}, which is proportional to the largest variance over all (normalized) treatment contrasts. A broad discussion of the E-criterion and its history can be found in Shah and Sinha (1989, chapter 3).

Optimality work based on criteria such as just discussed took hold with papers by Kiefer (1975) and John and Mitchell (1977). The former established rigorous optimality proofs in some very special cases. The latter generated designs based on an optimality conjecture. Both worked with settings for which v divides n, meaning that equal replication is possible and so covering the great majority of applications; this will also be done here.

John and Mitchell's (1977) conjecture is based on an equivalence between designs and graphs. Exploring that idea, and its relationship to one of Kiefer's (1975) key results, requires a few preliminary definitions. A block design d is *binary* if all n_{dij} are 0 or 1, that is, if no treatment is assigned to more than one unit in any block. For binary d, the diagonal elements of $N_d N_d'$ are the replication numbers r_{di}. Moreover, if d is also equireplicate (all r_{di} are equal), then C_d in (2) is a function only of the off-diagonal elements of $N_d N_d'$, denoted by $\lambda_{dii'}$ and termed treatment concurrence counts ($\lambda_{dii'}$ is the number of blocks containing both treatments i and i'). For given (v, b, k), define the *associated parameters* (r, λ, p, q) by

$$r = \text{int}\left(\tfrac{bk}{v}\right), \quad \text{the } \textit{replication target} \qquad \lambda = \text{int}\left(\tfrac{r(k-1)}{v-1}\right), \quad \text{the } \textit{concurrence target}$$

$$p = bk - vr, \quad \text{the } \textit{plot excess} \qquad q = r(k-1) - \lambda(v-1), \quad \text{the } \textit{concurrence excess}$$

The plot excess p is the number of experimental units available beyond that needed for equal replication of treatments. The possibility of equal replication (all r_{di} achieve the replication target r) thus necessitates $p = 0$, which as already stated, is taken throughout the remainder of this paper. For an equally replicated design, the concurrence excess q is the number of units in the blocks

containing any given treatment i beyond that required for the $\lambda_{dii'}$ to achieve the concurrence target λ for all i' ($\neq i$).

If both p and q are 0, then it is possible that a binary design can achieve both targets just defined: $r_{di} = r$ for all i, and $\lambda_{dii'} = \lambda$ for all $i \neq i'$. Widely known, such a design is a *balanced incomplete block design*, or BIBD. One of Kiefer's (1975) contributions was to establish A-, D-, and E-optimality, and much more, of BIBDs. Most of the theory for optimal block designs in the intervening 30 years, working with $(p, q) \neq (0, 0)$, can be viewed as trying to establish optimality for designs that come as "close" to these targets as possible. That approach makes intuitive sense, which is strengthened by the following decomposition of the information matrix.

For binary, equireplicate d and J an all-ones matrix, C_d can be written as

$$\begin{aligned} C_d &= rI - \frac{1}{k} N_d N_d' = rI - \frac{r}{k} I - \frac{1}{k} (\lambda_{dii'})_{i \neq i'} \\ &= \frac{r(k-1)}{k} I - \frac{1}{k} [(\lambda_{dii'})_{i \neq i'} - (\lambda)_{i \neq i'}] - \frac{1}{k} (\lambda)_{i \neq i'} \\ &= \left(\frac{v\lambda + q}{k} \right) I - \frac{\lambda}{k} J - \frac{1}{k} \Delta_d \end{aligned} \tag{3}$$

where the *discrepancy matrix* Δ_d has elements $\delta_{dii'}$, called *discrepancies*, defined by

$$\delta_{dii'} = \begin{cases} 0 & \text{if } i = i' \\ \lambda_{dii'} - \lambda & \text{if } i \neq i'. \end{cases} \tag{4}$$

Any discrepancy matrix is symmetric with constant row sums of q.

BIBDs have $q = 0$ and all $\delta_{dii'} = 0$, so that $C_d = \frac{\lambda v}{k}(I - \frac{1}{v} J)$. A matrix of the form $\alpha(I - \frac{1}{v} J)$ is said to be *completely symmetric*. Binarity and complete symmetry are the crucial ingredients of Kiefer's (1975) proof for the many optimalities of BIBDs. The intuitive argument above says to choose binary d so that Δ_d is as close to completely symmetric as possible. This brings us to John and Mitchell's (1977) work. A *regular graph design* (RGD) is a binary, equireplicate block design with all $\delta_{dii'}$ being 0 or 1. They conjectured "if $\mathcal{D}(v, b, k)$ contains a RGD, then a RGD is A-, D-, and E-optimal." There is no known counterexample to this conjecture for either A- or D-optimality. The situation will be seen to be different for the E-criterion.

The graph terminology is due to a simple equivalence. A graph is just a collection of v vertices with $\delta_{ii'}$ (say) edges connecting vertices $i \neq i'$. The matrix $\Delta = (\delta_{ii'})$ is called the *adjacency matrix*

of the graph. All graphs correspond to all adjacency matrices, that is, all symmetric matrices with nonnegative integer entries and zero diagonals. A graph is *regular* if its adjacency matrix has constant row sums, called the *valency*. The discrepancy matrix Δ_d is precisely the adjacency matrix for a regular graph of valency q provided, as in John and Mitchell's (1977) definition, it contains no negative values $\delta_{dii'}$. But there is also a graph equivalence if some $\delta_{dii'} < 0$.

Say a regular graph with adjacency matrix Δ is *1-regular* if all $\delta_{ii'} \in \{0, 1\}$, *2-regular* if not 1-regular and all $\delta_{ii'} \in \{0, 1, 2\}$, and so on. Define the $\overline{m}$*-complement* of Δ to be $\overline{m}(J - I) - \Delta$. Equivalence of discrepancy matrices and regular graphs is through this complement.

LEMMA 1 *For any nonzero discrepancy matrix Δ_d as defined by (3) and (4), let $\overline{m}_d = \max_{i \neq i'} \delta_{dii'}$ and $\underline{m}_d = \min_{i \neq i'} \delta_{dii'}$. Then*

$$\Delta = \overline{m}_d(J - I) - \Delta_d$$

is the adjacency matrix of an m-regular graph, $m = \overline{m}_d - \underline{m}_d$. Conversely, let Δ be the adjacency matrix for an m-regular graph with at least one $\delta_{ii'} = 0$, and write $\overline{m} = \lceil s/(v-1) \rceil$ where $s = 1'\Delta 1/v$ is the constant row sum of Δ. Then

$$\Delta_d = \overline{m}(J - I) - \Delta$$

is a discrepancy matrix with $\overline{m}_d = \overline{m}$ and $\underline{m}_d = \overline{m} - m$.

PROOF Given Δ_d, the difference $\overline{m}_d(J - I) - \Delta_d$ is symmetric and zero on the diagonal, contains only nonnegative integers with maximum $m = \overline{m}_d - \underline{m}_d$, and has constant row sums $(v-1)\overline{m}_d - q$. Conversely, given m-regular Δ, it is likewise easily checked that the difference $\overline{m}(J - I) - \Delta$ meets the conditions for a discrepancy matrix with constant row sums $q = \overline{m}(v - 1) - s$; the choice $\overline{m} = \lceil s/(v - 1) \rceil$ forces $0 \leq q \leq v - 2$. □

Starting with an m-regular graph, the corresponding discrepancy matrix will have negative entries if $\overline{m} < m$, which in turn depends on the valency $s = 1'\Delta 1/v$. The optimality work here will require only $\overline{m} = m-1$, with complement denoted $\Delta^c = (m-1)(J-I)-\Delta$ for an m-regular graph Δ. Then Δ^c contains at least one entry $\delta^c_{ii'} = -1$, and the valency of Δ is $s = (m - 1)(v - 1) - q$.

With this terminology, the RGDs of John and Mitchell (1977) are designs whose discrepancy matrices correspond to 1-regular graphs; these designs will now be denoted by RGD(1). Likewise a

2-regular graph design, or RGD(2), is a binary, equireplicate design for which Δ_d is the adjacency matrix of a 2-regular graph. A 2-regular complement graph design, or RGD(2^c), is a binary, equireplicate design for which the discrepancy matrix is $\Delta_d = J - I - \Delta$ for Δ the adjacency matrix of a 2-regular graph. RGD(m)s have all $\lambda_{dii'} \in \{\lambda, \lambda+1, \ldots, \lambda+m\}$ (equivalently, all $\delta_{dii'} \in \{0, 1, \ldots, m\}$) and RGD($m^c$)s have all $\lambda_{dii'} \in \{\lambda-1, \lambda, \lambda+1, \ldots, \lambda+m-1\}$ (all $\delta_{dii'} \in \{-1, 0, 1, \ldots, m-1\}$).

The idea is to look beyond the RGD(1)s for good designs, that is, to consider discrepancy matrices other than the 1-regular graphs. As will be seen, the RGD(m)s and RGD(m^c)s form a sufficiently rich class to solve the E-optimality problem. Presumably this would entail enumerating m-regular graphs for all possible m, but the number of graphs grows so rapidly with v that complete enumeration quickly becomes impossible. Section 2 develops theory for the E-criterion through which computational resources can be effectively brought to bear in eliminating many competing graphs prior to full enumeration. This theory allows determination of all E-best graphs for $v \leq 15$ in section 3. Section 4 provides techniques for building the graph-based designs comprising the online catalog described in section 5. Summary discussion, including choice of criterion and the practical advantages of the E-approach, is in section 6.

2 A method for solving the E-problem

Minimization of z_{d1} is, of course, equivalent to maximizing μ_{d1}. Workers in E-optimality have devised various upper bounds for μ_{d1}, then sought designs that achieve one of these bounds. The first goal here is to generalize these bounds so that the E-problem becomes more tractable both analytically and computationally. Lemma 2 applies to any symmetric, nonnegative definite matrix with zero row and column sums.

LEMMA 2 *Partition C_d as*

$$C_d = \begin{pmatrix} C_{d11} & C_{d12} \\ C_{d21} & C_{d22} \end{pmatrix}$$

where C_{d11} is $t \times t$ for some $t \geq 1$. Let w be any normalized t-vector and write $x = w'1$. Then

$$\mu_{d1} \leq \left(\frac{v}{v - x^2}\right) w' C_{d11} w$$

PROOF Let $y' = (w', 0')$ and decompose y by $y = (I - \frac{1}{v}J)y + \frac{1}{v}Jy = y_1 + y_2$. Starting with the facts that $\mu_{d1} = \min_{z'1=0} z'C_d z/z'z$ and $y_1'1 = 0$ gives

$$\mu_{d1} \leq \frac{y_1'C_d y_1}{y_1'y_1} = \frac{w'C_{d11}w}{w'w - \frac{1}{v}x^2} = \left(\frac{v}{v-x^2}\right) w'C_{d11}w. \qquad \square$$

Among previously employed E-bounds following from lemma 2 are:

COROLLARY 3 *The smallest nonzero eigenvalue μ_{d1} of any $C_d = (c_{dii'})$ in (2) satisfies (i) $\mu_{d1} \leq vc_{dii}/(v-1)$, (ii) $\mu_{d1} \leq (c_{dii} + c_{di'i'} - 2c_{dii'})/2$, and (iii) $\mu_{d1} \leq v(\sum_{i\in H} c_{dii} + 2\sum\sum_{i\neq i'\in H} c_{dii'})/h(v-h)$ where H is any h-subset of treatments.*

PROOF These follow by taking $t = 1, 2$, and h respectively, and in the latter two cases, $w = (1, -1)'$ and $w = 1_h$ and normalizing. $\square$

Corollary 3 and related bounds appear repeatedly throughout the literature, beginning with Jacroux (1980a,b) and Cheng (1980). They can be effective in winnowing out many possible information matrices, but are limited by choice of particular w. The advance here is not just in a simplified, unifying proof, but in showing the options for fully flexible choice of w. For a given order of submatrix, the sharpest bound afforded by lemma 2 is:

COROLLARY 4 *Let W_t be the set of normalized t-vectors. Then*

$$\mu_{d1} \leq \min_{w\in W_t} \left(\frac{v}{v-(w'1)^2}\right) w'C_{d11}w$$

To relate corollary 4 to choice of graph, observe that the positive eigenvalues of C_d as given in (3) are $\frac{v\lambda+q}{k} - \frac{1}{k}$(eigenvalues of Δ_d other than q). The smallest positive eigenvalue μ_{d1} of C_d corresponds to the largest eigenvalue (other than q) of Δ_d, called the *e-value of* Δ_d. It is now an easy step to:

COROLLARY 5 *Among binary, equireplicate designs, the E-optimal choice minimizes the e-value of Δ_d. Moreover, for any binary, equireplicate d,*

$$\mu_{d1} \leq \frac{v\lambda+q}{k} - \frac{1}{k} \max_{w\in W_t} \left(\frac{v}{v-(w'1)^2}\right) w'\widetilde{\Delta}_{d11}w$$

where $\widetilde{\Delta}_{d11}$ is any $t \times t$ principal submatrix of $\widetilde{\Delta}_d = \Delta_d - \frac{q}{v}J$.

For obvious reasons, a discrepancy matrix with minimal e-value for fixed v and q will be termed *E-best*. Corollary 5 is the result needed to find E-best graphs (and so E-optimal information matrices) without a full enumeration, for it can eliminate many graphs by virtue of poorly performing subgraphs. It will make an otherwise overwhelming combinatorial problem computationally feasible.

3 Determining E-best Graphs

The implementation of corollary 5 to select the best discrepancy matrix (that is, the best graph) will be described following an example to help fix the concepts.

EXAMPLE 1 Consider $v = 7$ treatments and any (b, k) such that the concurrence excess is $q = 2$. The E-best 1-regular graph (e-value=1.247) is

$$\Delta_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

while the E-best discrepancy matrix (e-value=1) is found by complementing a 2-regular graph:

$$\Delta_2^c = \begin{pmatrix} 0 & 2 & 2 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & 0 & 0 & 0 & 0 \\ 2 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 & 1 \\ 0 & 0 & 0 & 2 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 & 2 & 0 \end{pmatrix}^c = \begin{pmatrix} 0 & -1 & -1 & 1 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 & 1 & 1 & 1 \\ -1 & -1 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & -1 & 0 & 0 \\ 1 & 1 & 1 & -1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 \\ 1 & 1 & 1 & 0 & 0 & -1 & 0 \end{pmatrix}$$

One approach to finding these two matrices could be to enumerate all 1-regular and 2-regular graphs on 7 vertices, but that would not be sufficient for the optimality argument unless m-regular graphs could be shown inferior by other methods for every $m \geq 3$. The results of section 2 will be employed both to eliminate some m entirely, and to greatly reduce the enumeration for other, still feasible m.

Example 1 is just one example where designs can be built which are E-better than RGD(1)s. Actual designs and other examples will be presented later. Before us now is a graph enumeration

problem. The simple graphs (all $\delta_{ii'} \in \{0, 1\}$, regular or not), which include the 1-regular graphs, have been extensively studied by combinatorialists. Their numbers grow rapidly: while there are only four distinct possibilities for $v = 3$, there are $12,005,168$ for $v = 10$. Multiple edges adds another order of complexity so that a full enumeration even for the John-Mitchell range ($v \leq 12, r \leq 10$) would be a daunting, if not currently impossible, task. Fortunately, corollary 5 allows the E-best graphs to be determined with far less effort.

Corollary 5 provides information on the e-value of a graph via a subgraph. If graphs are built up one vertex at a time, then at each step all those failing to meet a specified bound can be discarded. This is the basic strategy fleshed out below. It is stated in terms of m-regular graphs; complements require only a slight modification. Because labelling of treatments is irrelevant, two graph matrices (or submatrices) Δ_1 and Δ_2 are equivalent, termed *isomorphic*, if one can be found from the other by a permutation of rows and columns ($\Delta_1 = P\Delta_2 P'$ for some permutation matrix P). Clearly only nonisomorphic graphs are sought. The algorithm begins with $m = 1$.

1. Guess the best e-value for given v and q, say e^*.
2. Enumerate possible order 3 submatrices of adjacency matrices for m-regular graphs.
3. Apply corollary 5 (omitting the constants $(v\lambda+q)/k$ and $-1/k$) to each submatrix, eliminating those with e-value greater than e^*.
4. For the remaining submatrices, generate nonisomorphic extensions with one additional row and column.
5. If current matrix order is less than v, return to step 3.
6. Keep the E-best (there may be several) of the remaining graphs.
7. Attempt to eliminate graphs with larger m, and complement graphs, by applications of lemma 2 (to a general Δ_d, see appendix A). If successful, the search is complete. If not, search again after incrementing m by 1.

If at any point in this process there are no remaining graphs, then return to step 1 and start over with larger e^*.

There are imbedded, but surmountable, computational problems here, such as how to efficiently generate extensions in step 4, which includes isomorphism checking and avoidance of previously

eliminated substructures. Calculation of the corollary 5 bound is easily accomplished numerically. By working sequentially in m, a search for (say) 2-regular graphs can exclude 1-regular graphs. Given the best graphs, elimination of nonbinary and unequally replicated designs is typically trivial (see appendix B), so that the graph search is really the main task. Table 1 summarizes results of this search for all (v, q) with $v \leq 15$. As there are far too many E-best discrepancy matrices to include here (over 1300) they have been made available online at *designtheory.org*, a website devoted to free storage and access to block designs and many of their properties, both statistical and combinatorial. The graphs are stored there as xml files, in *external representation format* (see Bailey *et al*, 2006), including notes on complementation. They are also available as R lists of discrepancy matrices at *www.stat.vt.edu/∼jpmorgan/discrepancy*.

4 Building E-optimal Designs

The work to this point solves only half the problem. Now that the E-best graphs are known, the corresponding designs must be found. Treatments can often, but not always, be assigned to blocks so that the desired discrepancies are actually achieved.

EXAMPLE 2 For $(v, b, k) = (7, 7, 5)$ the associated concurrence parameters are $(\lambda, q) = (3, 2)$. Example 1 displays the E-best discrepancy matrix Δ_2^c, but it can be shown (appendix B) that there is no corresponding design. In fact, the E-optimal design is this RGD(1), for which the discrepancy matrix is Δ_1 of example 1:

1	2	3	4	5	6	7
2	3	4	5	6	7	1
3	4	5	6	7	1	2
4	5	6	7	1	2	3
5	6	7	1	2	3	4

This design has all $\lambda_{dii'} \in \{\lambda, \lambda+1\} = \{3, 4\}$. For this v and k, the next b for which $q = 2$ is $b = 28$. Now a RGD(2^c) with discrepancy matrix Δ_2^c can be found:

1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2
2	2	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
3	3	3	3	4	4	4	4	4	4	5	5	4	4	4	4	4	4	5	5	4	4	4	4	4	4	5	5
4	4	5	5	5	5	5	5	6	6	6	6	5	5	5	5	6	6	6	6	5	5	5	5	6	6	6	6
6	7	6	7	6	6	7	7	7	7	7	7	6	6	7	7	7	7	7	7	6	6	7	7	7	7	7	7

and all $\lambda_{dii'} \in \{\lambda - 1, \lambda, \lambda + 1\} = \{6, 7, 8\}$. Given this design, the same discrepancy matrix can be achieved in $\mathcal{D}(7, 28 + 21s, 5)$ for any $s \geq 1$ by adding the blocks of a BIBD in $\mathcal{D}(7, 21s, 5)$. Thus E-optimal blocking of seven treatments into blocks of size five is solved for all b for which $q = 2$.

Example 2 illustrates this fact:

LEMMA 6 *If E-optimal d in $\mathcal{D}(v, b, k)$ has an E-best discrepancy matrix, $k \geq 3$, and $\mu_{d1} \geq \left(\frac{v}{v-1}\right) \frac{r(k-1)-2}{k}$, then $d \cup d^*$ is E-optimal in $\mathcal{D}(v, b + b^*, k)$ for any BIBD $d^* \in \mathcal{D}(v, b^*, k)$.*

PROOF Among binary, equireplicate designs, this is immediate from (3), the fact that information matrices for different sets of blocks are additive, and C_{d^*} is completely symmetric. The conditions on k and μ_{d1} are sufficient to extend the result to the full class; see appendix B. □

Designs d as in lemma 6 are said to be *BIBD-extendable*. Lemma 6 says that if the existence problem can be solved for small b, then it can otherwise be left to the wealth of knowledge on BIBDs (see Mathon and Rosa, 2006). The problem of building designs thus centers not on (v, b, k) but on (v, q, k). The number of treatments v and the concurrence excess q specify the relevant graph(s) and valency, after which given k specifies the relevant b, that is, the smallest b for which the (v, q, k) combination is feasible. Not every (v, q, k) is feasible; for instance, q can only be zero for $k = v - 1$.

With the E-best graphs for $v \leq 15$ in hand, there is an obviously large number of designs to be found, even with the extendability concept capping the possibilities for b. With the focus on online cataloging for easy access, only a sketch of the techniques for building the designs will be offered.

Soicher (2003) developed the GAP Design software for building block designs. It has been an invaluable tool in the current endeavor. GAP Design can accept a concurrence structure as input, that is, it can be told the desired graph (or complement), and *given sufficient time* will either find a design or conclude its nonexistence. Unfortunately, performance decreases rapidly with growing v, or b, or k. Performance is enhanced by the software's ability to accept a required automorphism group for the sought design. Statisticians have long used group-based methods for building designs, though not always in this generality nor employing this terminology, explained next.

An automorphism of a block design is a permutation of the treatments that reproduces the design. For instance, the first design in example 2 (call it d_1) is reproduced by the cyclic permutation

$(1 \to 2, 2 \to 3, \ldots, 7 \to 1)$. The set of all automorphisms of a design is its automorphism group; for d_1 this includes cycles at any lag t: $(1 \to 1+t, 2 \to 2+t, \ldots, 7 \to 7+t)$ (mod 7). This is a useful concept for building designs in that designs with "large" automorphism groups are determined by a correspondingly "small" subset of their blocks. For example, all blocks of d_1 can be found from any one block using the cyclic permutations. In searching for d_1, it is vastly easier to examine all possibilities for one block than for seven blocks. Use of the automorphism group is the underlying concept for the method of differences introduced by Bose (1939) for BIBD construction and extended many times by statistical researchers to a wide variety of contexts, including cyclic, generalized cyclic, and α-designs (see John and Williams, 1995).

EXAMPLE 3 For 11 treatments in 44 blocks of size 3, the concurrence excess is $q = 4$ and the E-best discrepancy matrix is the complement of a 2-regular graph:

$$\Delta^c = \begin{pmatrix} 2(J_4 - I_4) & 0 \\ 0 & J_7 - I_7 \end{pmatrix}^c = \begin{pmatrix} I_4 - J_4 & J_{4,7} \\ J_{7,4} & 0 \end{pmatrix}$$

The blocks of a design having this discrepancy matrix are found from the 14 blocks:

1	1	1	1	2	2	2	2	2	2	2	2	2	2
2	5	5	8	7	7	5	5	6	6	6	9	8	9
5	8	8	11	8	8	10	10	7	11	11	11	9	10

by simultaneously applying the cyclic permutations $(2, 3, 4)$, $(5, 6, 7)$, and $(8, 9, 10)$, holding 1 and 11 fixed. For instance, block $(1, 2, 5)$ also gives $(1, 3, 6)$ and $(1, 4, 7)$; each of the 14 "starter" blocks above gives 3 blocks of the design. The design is completed by two blocks, $(2, 3, 4)$ and $(5, 6, 7)$, invariant to the permutations.

The automorphism group of a block design must be a subgroup of the automorphism group of the associated graph. Many of the designs have been built by first determining the graph automorphism group, choosing a likely subgroup in light of the given b and k, then executing GAP Design with these inputs. Often several iterations, changing subgroups, was needed. In some cases b is too large (due to excessive run times) for GAP Design to be useful. These were solved by various trial-and-error constructions or by asking GAP Design to complete a likely subdesign.

In some cases no BIBD-extendable design exists. There are then two plausible approaches for finding E-optimal designs. One is to widen the graph search: allow larger e^* in the section 3 algorithm until the best graph for which a (necessarily non-extendable) design does exist is determined,

completing the optimality proof as shown in appendix B. The graph catalog at *designtheory.org* contains all additional graphs determined in this manner. Alternatively, one can bypass the graph approach and opt for a full enumeration of designs. This is feasible only with small (v, b), and usually only once the class of competitors has been reduced by theoretical arguments. Examples of both approaches are in appendix B.

Where nonexistence of a BIBD-extendable design has been determined, this was sometimes accomplished with GAP Design, but more often with combinatorial arguments relative to (v, b, k) and the desired graph (see appendix B).

5 The Design Catalog

The designs have been cataloged in downloadable form at *designtheory.org*. Each design in the catalog is notated as to corresponding graph, provided it was found using a cataloged graph. Each list of designs for given (v, q, k, b) is notated as to containing (or not) all possible E-optimal designs. By "all" is meant one design for every distinct E-optimal information matrix for which at least one design exists.

In the sense just discussed, *the E-optimality problem has been completely solved for* $v \leq 11$, $3 \leq k \leq v - 1$, save the case $(v, q, k) = (10, 6, 5)$, where design existence for some of the 89 E-best graphs remains unsettled. For larger v, as seen in Table 1, the number of E-best graphs is sometimes quite large! Due both to this and to the increasing difficulty in finding larger designs, the more modest goal of determining at least one BIBD-extendable E-optimal design for each (v, q, k) has been pursued. In this sense, *the E-optimality problem for* $v = 12, 13, 15$ *and* $3 \leq k \leq v - 1$ *has been solved except for one case for each*: $(v, q, k) = (12, 2, 5)$, for which the smallest b is 120; $(v, q, k) = (13, 4, 5)$, for which the smallest b is 65; and $(v, q, k) = (15, 12, 6)$ for which the smallest b is 20 (though see the caveat for $v = 15$ below). The work has been more problematic for $v = 14$, where only $q = 1, 4, 6, 12$ have been solved in their entirety. In addition $q = 2, 7$ lack only designs where BIBD-extendable designs have been proven not to exist. For other values of q, one or more of $k = 5, 6, 7$ are yet to be solved. See the catalog for details.

The approach of solving only the smallest b and thereafter relying on extendability presumes existence of the needed BIBDs. They are all known (Mathon and Rosa, 2006) except $(v, b, k) =$

$(15, 21, 5)$, which does not exist. Let b_q be the smallest b for $v = 15, k = 5$ and given q. Solution of all b ($= b_q + 21s$ for all $s \geq 0$) for $k = 5$ and q means that a BIBD-extendable design must be found for both b_q and $b_q + 21$. While b_q has been solved in every case, $b_q + 21$ has been solved only for $q = 12$. Thus as of this writing there is a single design missing for $v = 15, k = 5$ and each of $q = 2, 4, 6, 8, 10$.

The catalog contains all designs with $q \neq 0$ as described for $k = 3, \ldots, \lfloor \frac{v}{2} \rfloor$ and $k = v - 2$. Omission of $k = v-1$ is because $p = 0 \Rightarrow q = 0$ (and a BIBD exists). For other $k > \lfloor \frac{v}{2} \rfloor$ all E-optimal designs are found as the complements of all E-optimal designs with block size $v - k$; discrepancy matrices for a design and its complement are based on the same graph. Complementation fails only when the design with smaller k would be single-replicate and thus disconnected; for these few cases the larger k designs are cataloged.

Block size $k = 2$ stands apart from other $k < v$ in that the bounds of corollary 3 often fail to eliminate even unequally replicated competitors relative to the best equireplicate designs. Now consider, the adjacency matrix of any regular graph is equivalent to a block design with $k = 2$: g edges connecting vertices i and i' correspond to g blocks containing treatments i and i'. Given v, b, $k = 2$, q, λ, and an E-best discrepancy matrix Δ, the adjacency matrix $\Delta + \lambda$ corresponds to an E-optimal block design within the binary, equireplicate subclass $\mathcal{D}_0(v, b, k)$ of $\mathcal{D}(v, b, k)$. All discrepancy matrices found here have all entries ≥ -1, so this method fails to produce a design only if $\lambda = 0$ and at least one $\delta_{ii'} = -1$ for every E-best discrepancy matrix. Wherever this occurs, the list of graphs has been extended beyond the E-best as needed until a discrepancy matrix with all nonnegative entries results. Thus the graph catalog *completely solves the E-optimality problem for* $k = 2$ *and* $v \leq 15$ relative to $\mathcal{D}_0(v, b, k)$.

6 Discussion

One accomplishment of this paper is the near-complete solution to the E-optimal block design problem for $v \leq 15$. Having the corresponding catalog of designs online changes accessibility for experimentalists to a wide variety of useful designs from highly problematic to trivial. Existing catalogs are paper bound, dispersed in journals, and far from comprehensive in not covering all b and k possibilities for a range of v.

But mathematical optimality is not the entire story for good designs in experimental practice. The weakness of the E-criterion, and of A-, of D-, and others, is that they reduce a high-dimensional problem involving many variances and covariances to a single number. In doing so, they fail to capture the richness of the many forms of information available in the many competing designs from which a selection will be made.

The A-criterion (minimize average variance) is especially vulnerable on this point, in that it usually points to designs based on a single graph as being "best." This is akin to many routine statistical problems comparing distributions: sometimes the mean is an appropriate comparative measure, sometimes it is sharply misleading, and rarely (other than in theory) does it capture all the relevant aspects of the problem. For choice of block design, there is a (typically ill-specified) distribution of contrasts of interest with associated variances that are functions of the design. The problem is to "optimize" that distribution. The mean may do that, but it may not. In any case, small differences in the mean are usually less meaningful than are other aspects of the distribution.

To clarify the dimensionality problem, consider what it means for one design to dominate another. Design d_1 dominates d_2 if $C_{d_1} \neq C_{d_2}$ and d_1 estimates every contrast with variance no greater than that achieved by d_2: $l'C_{d_1}^{-}l \leq l'C_{d_2}^{-}l$ for every contrast vector $l \Leftrightarrow C_{d_1} - C_{d_2}$ is nonnegative definite. No nonzero matrix with zero trace can be nonnegative definite, so:

LEMMA 7 *If d_1 and d_2 are binary, then d_1 does not dominate d_2.*

Since there is no dominant design, an *optimal* design will be one which performs well when evaluated from several relevant perspectives, which can include any of (i) one or more formal optimality criteria, (ii) reflection of treatment structure in variances of contrasts, and even (iii) simplicity of interpretation. Many of the E-best discrepancy matrices (including examples 1 and 3) can be written in a block partitioned form with completely symmetric submatrices on the diagonal and constant submatrices off-diagonal, affording simplicity in the collection of variances. For instance, the design of example 3 will estimate elementary contrasts $\tau_i - \tau_{i'}$ with three distinct variances, corresponding to within and between the two groups of sizes 4 and 7 defined by Δ. The best RGD(1) produces five distinct variances for these contrasts in no discernible pattern.

An advantage of the E-approach as developed here is the revealing of many available information structures. Being E-optimal, they tend to behave *very* well with respect to the A-criterion (typically

A-efficiency $> 99\%$, and $< 98\%$ only for b and k both small; see Table 2 for definitions and specific values). Thus the E-criterion, easier to apply (by virtue of the results here), is also a highly effective proxy for A that offers more design choices. Users can browse the catalogs of matrices and designs, weighing the relative merits of competing designs for the experiment at hand. Alternatively, an experimenter could propose a discrepancy matrix (and tackle the corresponding design construction problem). As discrepancy matrices, with the constants (v, b, k), carry all of the information in binary, equireplicate designs, those found here provide standards against which any proposed alternative can be compared for performance in terms of (i)-(iii) above.

As stated in section 1, the work here is for n a multiple of v, covering a large majority of applications. When this does not hold, equireplication is not possible, nor is the decomposition (3) for Δ_d of form (4). It is also known that E-optimal block designs in this case need not be binary (Morgan and Uddin, 1995). Work is in progress for approaching unequally replicated designs through lemma 2, but the route is not through regular graphs. Small cases can be solved enumeratively if restricted to binary designs; a catalog for $v + b \leq 16$ can be found at *designtheory.org*.

Finally, the John-Mitchell conjecture is no longer generally viable for E-optimality. Constantine (1986) provided the first counterexample, for $(v, b, k) = (54, 135, 2)$. Bagchi (1994) then found designs for $v \geq 26$, $v \equiv 2 \pmod 4$ which are E-better then RGD(1)s. The combinations found here for which no E-best graph is 1-regular are $(v, q) = (7, 2), (11, 4), (12, 3), (13, 4), (13, 6), (14, 3), (14, 5)$, and $(15, 4)$, comprising many values of k and infinitely many b for which RGD(1)s must be E-inferior. The relative improvement is reflected in the ratios of e-values for these graphs relative to the best of the 1-regular graphs: respectively $1.25, 1.40, 1.53, 1.38, 1.30, 1.41, 1.53, 1.62$. For 27 other pairs (v, q) with $v \leq 15$, 1-regular graphs are not uniquely E-best (see Table 1). Many authors (e.g. Jacroux, 1985; Cheng and Constantine, 1986; Cheng, 1992) have worked to show optimality and high efficiency for RGD(1)s, and this paper does not contradict those results. What this paper says is that there are often many good designs other than RGD(1)s, that a view of optimality that extends beyond minimization of a single criterion will in any case lead to consideration of designs other than RGD(1)s, and that such a view is necessary for skillful practice in the art of designing an experiment. Moreover, routine statistical application requires that designs be readily available, a requirement the online catalog addresses.

A Bounds for the e-value of Δ_d

As noted earlier, lemma 2 can be applied to any symmetric, nonnegative definite matrix with zero row sums. The nonnegative condition can be set aside to bound the smallest eigenvalue other than one zero. Write $C_d = \frac{v\lambda+q}{k}(I-\frac{1}{v}J)-\frac{1}{k}\widetilde{\Delta}_d$ where $\widetilde{\Delta}_d = \Delta_d - \frac{q}{v}J = ((\tilde{\delta}_{dii'}))$. Corollary 4 says work can proceed using $\widetilde{\Delta}_d$ rather than C_d, freeing the problem from dependence on b, k, and λ. Minimizing the e-value of Δ_d is equivalent to maximizing $\tilde{e}_{d1}$, the smallest eigenvalue (after setting one zero aside) of $-\widetilde{\Delta}_d$. Analogous to part (iii) of corollary 3, $\tilde{e}_{d1} \leq -v(\sum_{i\in H}\tilde{\delta}_{dii} + 2\sum\sum_{i<i'\in H}\tilde{\delta}_{dii'})/h(v-h)$, which provides a bound for examining feasibility of individual discrepancy values. For instance, taking $h=2$, if some $\delta_{dii'}=2$ then this inequality reduces to $\tilde{e}_{d1} \leq 2(q-v)/(v-2)$. Since e-value $= -\tilde{e}_{d1}$, existence of a 1-regular graph with e-value less than $2(v-q)/(v-2)$ eliminates $\delta_{dii'}=2$ from consideration. This (and similar arguments for negative discrepancies) is the technical reason the search strategy in section 3 begins with $m=1$.

B Optimality in the full class

Equation 3 says that a design with E-best discrepancy matrix is E-optimal over all equireplicate, binary designs. Remaining competitors are eliminated using corollary 3(i):

$$\mu_{d1} \leq \left(\frac{v}{v-1}\right)\frac{(r-1)(k-1)}{k} \quad \text{for any unequally replicated design} \tag{5}$$

$$\mu_{d1} \leq \left(\frac{v}{v-1}\right)\frac{r(k-1)-2}{k} \quad \text{for any nonbinary design.} \tag{6}$$

For $k>2$, the bound (6), used in lemma 6, is at least as large as (5). In *every tabled case* for $v \leq 15$ and $k>2$, the BIBD-extendable designs exceed bound (6). Let $d^* \in \mathcal{D}(v,b^*,k)$ be one of the tabled, BIBD-extendable designs, and let $\tilde{d} \in \mathrm{D}(v,\tilde{b},k)$ be a BIBD. Writing $\bar{b} = b+\tilde{b}$, let $\bar{d}^* \in \mathcal{D}(v,\bar{b},k)$ denote the design whose blocks are the $\bar{b}$ blocks of d^* and $\tilde{d}$ together. Suppose there is a nonbinary (or unequally replicated) design $\bar{d} \in \mathcal{D}(v,\bar{b},k)$ for which $\mu_{\bar{d}1} > \mu_{\bar{d}^*1}$. Write its information matrix as $C_{\bar{d}} = [C_{\bar{d}} - \frac{\tilde{\lambda}v}{k}(I-\frac{1}{v}J)] + \frac{\tilde{\lambda}v}{k}(I-\frac{1}{v}J) = C_{\hat{d}} + \frac{\tilde{\lambda}v}{k}(I-\frac{1}{v}J)$. Then $C_{\hat{d}}$ is a symmetric matrix of rank $v-1$, and whether or not a design exists with this information matrix, the diagonal elements of $C_{\hat{d}}$ are those of some nonbinary (or unequally replicated) design in $\mathcal{D}(v,b^*,k)$. Thus (6) holds for $\hat{d}$ and consequently $\mu_{\hat{d}1} < \mu_{d^*1}$. But $\mu_{\hat{d}1} = \mu_{\bar{d}1} - \frac{\tilde{\lambda}v}{k} > \mu_{\bar{d}^*1} - \frac{\tilde{\lambda}v}{k} = \mu_{d^*1}$, a contradiction. This

completes the proof of lemma 6.

Non-extendable designs often beat (5) and (6) as well. The 7-block design in example 2, with discrepancy matrix Δ_1 of example 1, has $\mu_{d1} = \frac{v\lambda+q-e}{k} = \frac{7(3)+2-1.247}{5} = 4.35 > 4.2$, the larger of the two bounds. Since no design exists for any discrepancy matrix that is E-better than Δ_1, this non-extendable design is E-optimal over all of $\mathcal{D}(7,7,5)$. In other settings where no extendable design exists, more work is needed. A typical argument is shown next.

For $(v,b,k) = (15,10,3)$ the associated parameters are $(r,\lambda,q) = (2,0,4)$. Here it was possible to generate all RGD(1)s. The uniquely best of these has $\mu_{d1} = 2/3$, surpassing neither (5) nor (6). Its dual design with $(v,b,k) = (10,15,2)$ has smallest eigenvalue 1 (E-ordering is preserved in the class of equireplicate duals - see Shah and Sinha, 1989, pp. 27-8). If some pair of treatments concurs twice in an equireplicate design $\bar{d}$ in the dual setting, then by the third part of corollary 3 with $h = 2$, $\mu_{\bar{d}1} \leq 0.625 < 1$. Thus among equireplicate duals, the best must be a RGD(1), which implies the same for equireplicate designs in $\mathcal{D}(15,10,3)$.

It remains to rule out unequally replicated designs, binary or not. Take treatment 1 to have replication 1 and notice that the largest possible value of C_{dii} is $20/k$. If some treatment occurs twice in the block with treatment 1, then there is a 2×2 submatrix of C_d of the form

$$\frac{1}{k}\begin{pmatrix} 2 & -2 \\ -2 & x \end{pmatrix} \quad \text{for integer } 2 \leq x \leq 20.$$

Evaluating this submatrix for each x using corollary 4 rules out this possibility. For two distinct treatments in the block with treatment 1, there is a 3×3 submatrix of the form

$$\frac{1}{k}\begin{pmatrix} 2 & -1 & -1 \\ -1 & x & -z \\ -1 & -z & y \end{pmatrix} \quad \text{for postive integers } 1 \leq x \leq y \leq 20 \text{ and } z \geq \max\{1, x+y-10\}.$$

Another round of applications of corollary 4 for the x, y, z combinations completes the result.

Ruling out existence of a design with a specified discrepancy matrix is sometimes possible through computer search. When not, a frequently successful method starts with a subset of v_1 treatments, $k \leq v_1 < v$. Let x_i be the number of blocks containing i treatments from that subset. Then $\sum_{i=1}^{k} x_i \leq b$, $\sum_{i=1}^{k} i x_i = r v_1$, and $\sum_{i=1}^{k} i(i-1)x_i = \lambda v_1(v_1-1) + s$, where s is the sum of the elements of the corresponding discrepancy submatrix. If there are no nonnegative integer solutions to this system, then there is no design. One example is $(v,b,k) = (7,7,5)$ of example 2, which admits no solution for treatments $1,\ldots,5$ in Δ_2^c of example 1.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation. The author, also supported for six months by the U. K. Engineering and Physical Sciences Research Council, warmly acknowledges facilities and collegiality provided by Queen Mary University of London.

REFERENCES

Bagchi, S. (1994). Optimality and construction of some rectangular designs. *Metrika* **41**, 29–41.

Bailey, R. A., Cameron, P. J., Dobcsanyi, P., Morgan, J. P., and Soicher, L. H. (2006). Designs on the web. *Discrete Mathematics*, to appear.

Box, G. E. P., Hunter, J. S. and Hunter, W. G. (2005). *Statistics for Experimenters, 2nd edition.* Springer-Verlag, New York.

Bose, R. C. (1939). On the construction of balanced incomplete block designs. *Ann. Eugenics* **9**, 353–399.

Chakrabarti, M. C. (1962). *Mathematics of Design and Analysis of Experiments.* Asia Publishing House, Bombay.

Cheng, C.-S. (1980). On the E-optimality of some block designs. *J. Roy. Statist. Soc. Ser. B* **42**, 199–204.

Cheng, C.-S. (1992). On the optimality of (M.S)-optimal designs in large systems. *Sankhyā Ser. A* **54**, 117–125.

Cheng, C.-S. and Constantine, G. M. (1986). On the efficiency of regular generalized line graph designs. *J. Statist. Plann. Inference* **15**, 1–10.

Constantine, G. M. (1986). On the optimality of block designs. *Ann. Inst. Statist. Math.* **38**, 161–174.

Dean, A. and Voss, D. (1999). *Design and Analysis of Experiments.* Springer-Verlag, New York.

Fisher, R. A. (1937). *The Design of Experiments*. Oliver and Boyd, Edinburgh.

Jacroux, M. (1980a). On the E-optimality of regular graph designs. *J. Roy. Statist. Soc. Ser. B* **42**, 205–209.

Jacroux, M. (1980b). On the determination and construction of E-optimal block designs with unequal number of replicates. *Biometrika* **67**, 661–667.

Jacroux, M. (1985). Some sufficient conditions for the type I optimality of block designs. *J. Statist. Plann. Inf.* **11**, 385–396.

John, J. A. and Mitchell, T. J. (1977). Optimal incomplete block designs *J. Roy. Statist. Soc. Ser. B* **39**, 39–43.

John, J. A. and Williams, E. R. (1995). *Cyclic and Computer Generated Designs, 2nd edition. Monographs on Statistics and Applied Probability, 38.* Chapman & Hall, London.

Kuehl, R. O. (2000). *Design of Experiments: Statistical Principles of Research Design and Analysis* Brooks/Cole, Pacific Grove.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In *A Survey of Statistical Design and Linear Models*, Ed. J.N. Srivastava, pp. 333-353. North-Holland, Amsterdam.

Mathon, R. and Rosa, A. (2006). 2-(v, k, λ) designs of small order. In *The CRC Handbook of Combinatorial Designs, 2nd edition*, Ed. C. J. Colbourn and J. H. Dinitz, pp. 25–58. CRC Press, Boca Raton.

Morgan, J. P. and Uddin, N. (1995). Optimal, non-binary, variance balanced designs. *Statist. Sinica* **5**, 535-546.

Oehlert, G. W. (2000). *A First Course in Design and Analysis of Experiments.* Freeman, New York.

Shah, K. R. and Sinha, B. K. (1989). *Theory of optimal designs. Lecture Notes in Statistics, 54.* Springer-Verlag, New York.

Soicher, L. H. (2003). The DESIGN package for GAP. *designtheory.org/software/gap_design*

Table 1: Number of E-best graphs by type

v	q	# E-best Graphs	$\#(1, 2, 2^c, 3^c)$
4	1	2	$(1, 0, 1, 0)$
	2	1	$(1, 0, 0, 0)$
5	2	1	$(1, 0, 0, 0)$
6	1	3	$(1, 0, 2, 0)$
6	2	2	$(1, 0, 1, 0)$
	3	1	$(1, 0, 0, 0)$
	4	1	$(1, 0, 0, 0)$
7	2	1	$(0, 0, 1, 0)$
	4	1	$(1, 0, 0, 0)$
8	1	5	$(1, 0, 4, 0)$
	2	2	$(1, 0, 1, 0)$
	3	6	$(2, 0, 4, 0)$
	4	1	$(1, 0, 0, 0)$
	5	1	$(1, 0, 0, 0)$
	6	1	$(1, 0, 0, 0)$
9	2	1	$(1, 0, 0, 0)$
	4	13	$(4, 0, 9, 0)$
	6	1	$(1, 0, 0, 0)$
10	1	7	$(1, 0, 6, 0)$
	2	2	$(1, 0, 1, 0)$
	3	7	$(1, 0, 6, 0)$
	4	3	$(1, 0, 2, 0)$
	5	1	$(1, 0, 0, 0)$
	6	89	$(6, 26, 5, 52)$
	7	1	$(1, 0, 0, 0)$
	8	1	$(1, 0, 0, 0)$
11	2	1	$(1, 0, 0, 0)$
	4	1	$(0, 0, 1, 0)$
	6	8	$(1, 0, 7, 0)$
	8	1	$(1, 0, 0, 0)$
12	1	11	$(1, 0, 10, 0)$
	2	2	$(1, 0, 1, 0)$
	3	4	$(0, 0, 4, 0)$
	4	17	$(1, 0, 16, 0)$
	5	56	$(7, 0, 49, 0)$
	6	1	$(1, 0, 0, 0)$
	7	480	$(15, 101, 55, 309)$
	8	1	$(1, 0, 0, 0)$
	9	1	$(1, 0, 0, 0)$
	10	1	$(1, 0, 0, 0)$
13	2	1	$(1, 0, 0, 0)$
	4	1	$(0, 0, 1, 0)$
	6	1	$(0, 0, 1, 0)$
	8	18	$(2, 0, 16, 0)$
	10	1	$(1, 0, 0, 0)$
14	1	15	$(1, 0, 14, 0)$
	2	2	$(1, 0, 1, 0)$
	3	1	$(0, 0, 1, 0)$
	4	1	$(1, 0, 0, 0)$
	5	1	$(0, 0, 1, 0)$
	6	4	$(1, 0, 3, 0)$
	7	1	$(1, 0, 0, 0)$
	8	2213	$(21, 158, 198, 1836)$
	9	24	$(2, 0, 22, 0)$
	10	11	$(6, 0, 5, 0)$
	11	1	$(1, 0, 0, 0)$
	12	1	$(1, 0, 0, 0)$
15	2	1	$(1, 0, 0, 0)$
	4	16	$(0, 0, 16, 0)$
	6	234	$(7, 0, 227, 0)$
	8	19	$(2, 0, 17, 0)$
	10	1	$(1, 0, 0, 0)$
	12	1	$(1, 0, 0, 0)$

Table 2: Worst A-efficiencies for E-best non-RGD(1)s

v	q	$(b, k,$ A-efficiency)
6	1	$(6,3,.944),(3,4,.944)$
6	2	$(12,3,.993),(6,4,.993)$
7	2	$(7,5,.994)$
8	1	$(48,3,.996),(10,4,.983),(48,5,.999),(4,6,.983)$
8	2	$(40,3,.999),(6,4,.990),(40,5,.999),(8,6,.999)$
8	3	$(32,3,.997),(16,4,.997),(32,5,.999),(12,6,.999)$
9	4	$(6,3,.880),(9,4,.980),(9,5,.992),(6,6,.992),(18,7,.999)$
10	1	$(14,5,.992),(5,8,.992)$
10	2	$(10,5,.998),(10,8,.999)$
10	3	$(20,3,.986),(10,4,.986),(6,5,.986),(10,6,.997),(20,7,.999),(15,8,.999)$
10	4	$(20,5,.999),(20,8,.999)$
10	6	$(10,3,.854),(5,4,.854),(12,5,.981),(5,6,.960),(10,7,.992),(30,8,.999)$
11	4	$(44,3,.997),(22,4,.997),(22,7,.999),(44,8,.999)$
11	6	$(11,3,.951),(33,4,.998),(33,7,.999),(11,8,.999),(33,9,.999)$
12	1	$(24,3,.965),(12,4,.965),(60,5,.999),(18,6,.996),(60,7,.999),(12,8,.997),(24,.999),(6,10,.996)$
12	2	$(48,3,.999),(24,4,.999),(120,5,.999),(14,6,.999),(120,7,.999),(24,8,.999),(4,9,.999),(12,10,.999)$
12	3	$(28,3,.986),(36,4,.997),(48,5,.999),(10,6,.994),(48,7,.999),(3,8,.986),(28,9,.999),(18,10,.999)$
12	4	$(8,3,.930),(15,4,.989),(108,5,.999),(6,6,.989),(108,7,.999),(15,8,.999),(8,9,.998),(24,10,.999)$
12	5	$(32,3,.984),(27,4,.993),(36,5,.998),(24,6,.998),(36,7,.999),(27,8,.999),(32,9,.999),(30,10,.999)$
12	7	$(36,3,.966),(18,4,.966),(24,5,.990),(16,6,.990),(24,7,.997),(18,8,.997),(36,9,.999),(42,10,.999)$

Shown are (v, q) for which 1-regular graphs are either suboptimal or not uniquely best. A-efficiencies compare A-worst of E-best non-RGD(1)s versus A-best of E-best RGD(1)s. Evaluation for each (v, q, k) is for smallest possible b. Efficiencies increase quickly with b. Of the 121 cases with smallest b for $v = 13, 14, 15$, one A-efficiency is .924; two are .967; one is .977; five are in $[.980, .990]$; all others are $> .990$. Notice that even when there are many competing matrices, the range of A-behavior is *very* small (e.g. 480 E-best matrices for $(v, q) = (12, 7)$, but almost all A-efficiencies $\geq .990$). All values truncated to three places. Only rarely do A-best and E-best among RGD(1)s fail to coincide (see John and Mitchell, 1977).