

Center for Quality and Applied Statistics
Kate Gleason College of Engineering
Rochester Institute of Technology

The Efficiencies of Fractional Factorial Designs

Joseph G. Voelkel
Center for Quality and Applied Statistics
Rochester Institute of Technology
jgvcqa@rit.edu

Technical Report 2004–1

February 2004

The Efficiencies of Fractional Factorial Designs

JOSEPH G. VOELKEL, CQAS, R.I.T., ROCHESTER, NY

ABSTRACT. A frequently-stated advantage of fractional factorial (FF) designs over one-factor-at-a-time (1FAT) designs is their high relative efficiency. We study k -factor, $2k$ -run designs, where k is a power of 2, where the usually-stated relative efficiency is k in favor of the resolution IV FF design over an orthogonal 1FAT design. In this paper, we examine other ways to measure efficiency under main-effects-only modeling. If both designs are restricted to lie in smallest-sized hypercubes of equal volume, or in hyperspheres of equal volume, then the relative efficiency of the FF design is 1 under the full model (k effects active). We also show that the two designs can be considered identical, up to scale and rotation—that is, the 1FAT design can be rescaled and rotated to become a resolution IV FF design. The real advantage of the FF design lies in its projection properties. For example, if at most k' factors are expected to be active, then, under equal-volume designs, the efficiency of the FF is at least k/k' . Taking a size-of-effects-based approach, we rescale the FF in a different way. We rescale the design to have the same probability as a 1FAT design of producing too-large swings in the response (swings that produce unusable runs). Here, while the efficiency of the FF usually remains greater than 1, it can reach $1/k$ as a limiting case. We show how these measures of efficiency may be used to help scale back originally-planned settings in a fractional factorial design to reduce the chance of unusable runs while still maintaining the other advantages of FF designs.

Keywords: one factor at a time, projection properties, experimental design

1. INTRODUCTION

In an expository article, Czitrom (1999) illustrates many of the arguments for favoring fractional-factorial (FF) over one-factor-at-a-time (1FAT) designs. There is no question that the FF design is superior to the 1FAT design in a number of important ways. For example, a Resolution IV FF design can detect (groups of) two-factor interactions, and in a certain sense is less sensitive to outliers than 1FAT designs. Our interest lies in measures of the efficiency of the FF design. In this article, “efficiency” will always refer to the efficiency of a FF design relative to a 1FAT design.

Czitrom’s is one of few articles in the statistical literature that mention 1FAT designs. Daniel (1973) shows that a scientist’s natural inclination to do 1FAT runs can be arranged sequentially to lead to reduction of two-factor interaction bias. Morris (1991) considers *computational* (that is, deterministic, to be contrasted with *physical*) experiments, and shows how a series of one-factor-at-a-time designs may be used to gain certain knowledge about the system under study.

To measure efficiency, we will assume that main effects may be active, but (1) *there are no active interactions*, and that (2) *all active main effects are linear*. The first assumption is not unreasonable—the number of active interactions in any experiment is typically fairly small—0, 1, or 2 is common—and even this small number may be reducible by careful consideration in both the design and analysis stages. (A nice example of reduction in both stages is provided by Buckner, et al. (1997).) The second assumption is also reasonable—nonlinearities may be also be reduced or eliminated in the design and analysis stages. Arguments along such lines in the design stage are made in the robust-design arena, e.g. Phadke (1989, Chapter 6), while arguments in the analysis stage usually employ transformations, such as in Box and Cox (1964). This second assumption is needed to compare designs under the rescalings we consider.

We will consider two-level designs in which k factors will be studied in $n = 2k$ runs, where k is a power of 2. We assume that standard features of experimentation, such as randomization, will be performed. Say that, on a coded scale, $\mathbf{0}$ refers to the vector of current conditions. We use the standard way of creating a FF design in this case: use a resolution IV 2^{k-p} design, with coded settings of ± 1 .

By comparison, there is no standard way in which a set of 1FAT runs is defined. One definition would be to make a run at $\mathbf{0}$, and then make k additional runs where, at run i , only the level of factor i is changed, say to $+1$. Another definition would be to start at $\mathbf{0}$; then make several runs at which only the first factor's levels are varied; decide upon a “best” setting for that factor and use this to update the vector of current conditions; and continue to do so with each of the remaining factors.

We will instead examine the following way to implement a 1FAT strategy. For $i = 1, 2, \dots, k$,

- At run i , set factor i to setting -1 , and set all other factors to setting 0 .
- At run $i + k$, set factor i to setting $+1$, and set all other factors to setting 0 .

We have chosen to look at this class of designs because *both sets of designs are orthogonal main-effects designs of k factors in $2k$ runs*, and because the larger number of factors and large fractionation, say for $k \geq 8$, is precisely the kind of situation for which the efficiency of FF designs is considered to excel over 1FAT designs. Another reasonable comparison could be made between running a saturated, main-effects only, FF design in k factors in $n = k + 1$ runs (where n is a power of 2) and comparing this to a 1FAT design based on the first definition given above. However, the difference in the actual range of factor levels, and the lack of orthogonality of the 1FAT design makes this comparison less appealing for the purposes we wish to explore in this paper.

Denote the standard deviation of the experimental error by σ . Then the variances of each of the resulting estimated main half-effects b , the natural estimates to use under regression-based analysis, are

$$\begin{aligned}\text{Var}(b_{1FAT}) &= \sigma^2/2 \\ \text{Var}(b_{FF}) &= \sigma^2/2k.\end{aligned}$$

So, using this measure, the efficiency of the FF design is k .

2. EFFICIENCIES WITH EQUALLY-SIZED DESIGN REGIONS: FULL MODEL

Comparisons between designs are customarily made in regions that are, in some geometrical sense, the same size. For example, Lucas (1976) compares designs for quadratic models in either a fixed-size hypercube or a fixed-size hypersphere.

Consider the hypercube region. All points for the FF design and the 1FAT design in k factors lie on the 2-unit k -dimensional hypercube that is centered at the origin and whose sides are parallel to the coordinate (factor-level) axes. Note that if the model was later reduced, say to k' factors, then not all of the points in the 1FAT design would lie on the corresponding k' -dimensional hypercube—those from the remaining $k - k'$ factors would be at the center of the hypercube. However, all points of the FF design would still be on this hypercube. So it is reasonable to say that the efficiency of the FF is k in the *parallel hypercube region under the full model*, where the full model consists of the k linear, main-effect, terms.

Now suppose we adopt a more general geometric approach by not requiring the hypercube that envelopes a design to be restricted to lie parallel to the coordinate axes. Because the designs we consider are centered at the origin, it will be sufficient to consider rotations of each hypercube from its initial position. This seems reasonable—there is no reason to restrict the hypercube from rotating when considering its volume. Rotations need only be considered for the 1FAT design, because all points of the FF already lie on the extreme points of the hypercube.

Consider the $k = 2$, $n = 4$ case. (Here, the FF design is actually a full factorial, but we will still refer to it as FF.) See Figure 1. If we restrict the hypercubes to lie parallel to the axes, then the solid-lined square is needed for both designs. In this case the efficiency of the FF design is 2. However, if we allow the hypercube to rotate for the 1FAT design, and then rescale it to provide a hull for this design, this generates the dotted-line hypercube, whose sides are of length $\sqrt{2}$. When we rescale the 1FAT design to have a hypercube whose volume is the same as that of the FF's, then the efficiency of the FF becomes 1. Also, note that the 1FAT design, in the rotated space, has become a FF design.

It turns out that these results hold in the general case.

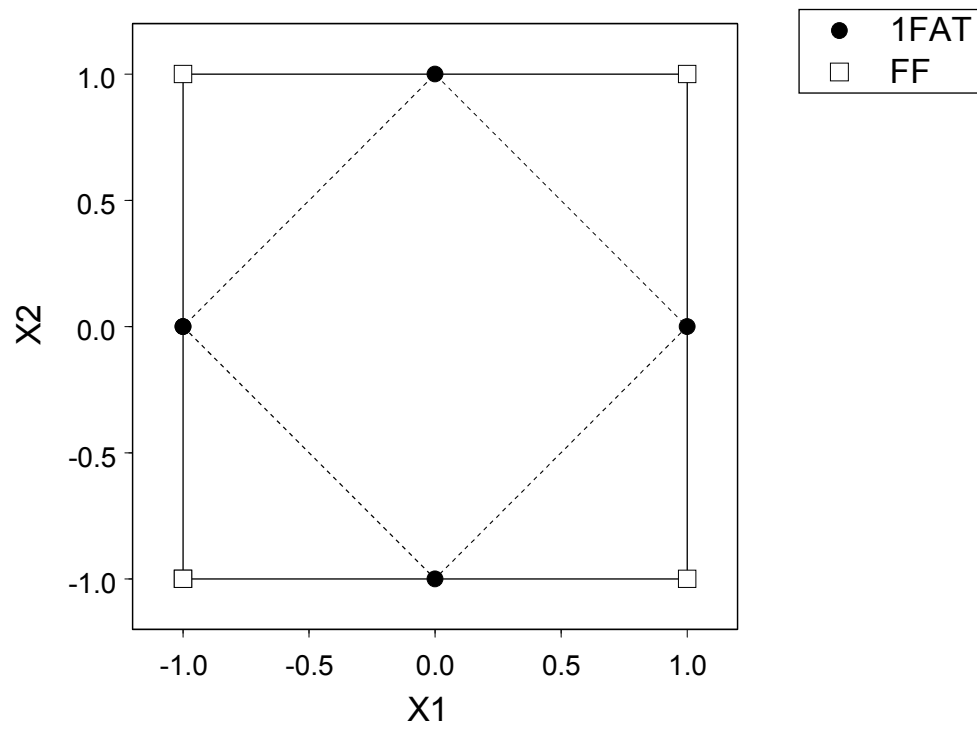


Figure 1: Hypercube Volumes Needed to Envelope 1FAT and FF Designs—Two-Factor Case.

Theorem 1. *Consider the FF and 1FAT designs in the k -factor, $2k$ -run, case under the full model when k is a power of 2. Then*

1. *If efficiency is measured based on same-volume hypercube hulls whose sides must be parallel to the coordinate axes, then the efficiency of the FF design is k .*
2. *If efficiency is measured based on same-volume hypercube hulls that may be rotated independently for each design, then the efficiency of the FF design is 1.*
3. *A 1FAT hypercube rotation matrix \mathbf{R} that can be used in (2) consists of one column of 1's and $(k-1)$ columns from an orthogonal, saturated, $(k-1)$ -factor FF design, all divided by \sqrt{k} .*
4. *In these rotated coordinates, the 1FAT design is a Resolution IV FF design.*
5. *(Equivalency of Designs) The FF design, aside from a scale factor, may be constructed to correspond to this Resolution IV FF design. Thus, the FF design and the 1FAT design are identical up to a scale and rotation transformation.*

The proof is contained in the Appendix. Result (5), along with the rotatability of the FF design under this full model, also means that the prediction-variance contours of the FF and 1FAT are identical.

The case $k = 2$ that we have looked at is trivial, so consider $k = 4$, as shown in Figure 2. Note that \mathbf{R} consists of a constant column and a 2^{3-1} design, scaled back by $\sqrt{4} = 2$, and that the rotated design is a scaled 2^{4-1} design.

By the nature of the proof, the result extends to Plackett-Burman-based designs, that is, Plackett-Burman designs that are folded over on all factors. However, the result does not need to hold in more general cases. For example, consider two 3-factor orthogonal designs: a 6-run 1FAT design using the construction method we have presented, and an 8-run 2^3 design. The optimal rotation for the first design allows it to be enclosed in a hypercube whose sides are $\pm 2/3$, as shown in the Appendix. Then, adjusting the designs for both this scale factor and the number of runs leads to an efficiency of 4 (original efficiency) $\times (2/3)^2 \times 6/8 = 4/3$, so the FF is still superior by this measure.

Next, consider comparisons on the hypersphere (of dimension k). All points for the 1FAT design for k factors are on the unit hypersphere, while all points for the FF are on a hypersphere of radius \sqrt{k} . By shrinking the FF design by a \sqrt{k} factor, both designs would be on the same-sized hypersphere. But when this is done, the efficiency of the FF design under the full model again becomes 1.

$$\mathbf{X} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{R} = \begin{pmatrix} 0.5 & -0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

$$\mathbf{X} \cdot \mathbf{R} = \begin{pmatrix} -0.5 & 0.5 & 0.5 & -0.5 \\ -0.5 & -0.5 & 0.5 & 0.5 \\ -0.5 & 0.5 & -0.5 & 0.5 \\ -0.5 & -0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & -0.5 & 0.5 \\ 0.5 & 0.5 & -0.5 & -0.5 \\ 0.5 & -0.5 & 0.5 & -0.5 \\ 0.5 & 0.5 & 0.5 & 0.5 \end{pmatrix}$$

Figure 2: 1FAT Design \mathbf{X} , Rotation Matrix \mathbf{R} , and Rotated Design.

3. EFFICIENCIES WITH EQUALLY-SIZED DESIGN REGIONS: REDUCED MODEL

How the response is affected by the factors also plays an important role. We can make design comparisons along these lines in two ways. First, we can consider various reduced models, but without any additional regard to the possible sizes of the effects. That is, we can consider effects to be either active or inactive. This is our approach in this section. Second, we can consider the possible values of the non-zero effects—this will be the basis of our approach in the next section.

For an example of the reduced-model approach in the hypersphere-volumen case, suppose we believe that at most one-half of the factors will be active, but we are not sure which ones. Then a unit $k/2$ -dimensional hypersphere would become the hull for the 1FAT design. However, for the FF design, the radius $1/\sqrt{k}$ hypersphere in k dimensions would become a radius $1/\sqrt{2k}$ hypersphere in $k/2$ dimensions, on an equal-volume basis, and so the FF is now twice as efficient as the 1FAT. In the general case, with at most k' factors active, the efficiency of the FF design on an equal-volume basis is k/k' . In the extreme case, with only one factor active, the efficiency becomes k .

For the hypercube case, if at most k' factors are expected to be active, and k' is a factor of 2, then, on an equal-volume basis, the rotated hypercube hull for the 1FAT design has length $1/\sqrt{k'}$, and the hypercube hull for the FF design still has length 1. So, the efficiency of the FF design is again k/k' . In both cases, these greater efficiencies are realized because of the projection properties of FF designs (e.g. Box and Hunter (1961), Cheng (1995, 1998), Tang (2001)). So, it would appear that *the true efficiencies of fractional factorial designs lie in their projection properties*. Because these projective properties are factor-coordinate-axis dependent, they work for the FF design but not the 1FAT design.

If we accept this equal-volume basis, and agree that the 1FAT-design factor levels of ± 1 are reasonable for defining the volume of design space, this means we should scale back the FF design by a $\sqrt{k'}$ factor. Using an example with $k = 16$ factors, say we believe all factors might be active. Then, to create a FF design of equal volume, we would need to set their levels to $\pm 1/4$, with an efficiency of 1. However, if we believed that at most 4 factors might be active, then we could set the FF factor levels to $\pm 1/2$, with an efficiency of 4.

For this reason, we should consider model-reduction-based information to be considered when comparing the two designs.

4. ZERO-INFORMATION RUNS AND RESTRICTED-INFORMATION REGIONS

Another way to measure efficiency is to incorporate information based on the size of the effects, which in turn affects the swings in the response. Because this is not a standard approach in comparing designs, we begin by explaining the approach in this section.

The FF design explores much larger regions of design space, but in doing so can greatly increase the chance of a large change in values of the response, which in turn may lead to a failed run or, as we call it here, a *zero-information run* (OIR). In fact, this work was initiated because of difficulties involved in running a designed experiment involving 27 factors. Although each factor's levels were carefully set *individually* to avoid failed runs, the *combination* of a reasonable number of active effects along with the large region of design space generated by a 2^{27-20} design resulted in a large fraction of failed runs, and the experiment was aborted.

Although such failed runs provide information in one sense, most experimenters try to create designs to avoid such possibilities. We will assume this to be the case. Here are some examples of OIR's:

- Let Y be a property of single-phase glass. In modeling durability of glasses as a function of composition, if one has data on both single-phase and multi-phase glasses, the usual practice is to only use the single-phase data (e.g., Piepel, et al. (2002)). Then a OIR occurs under conditions in which the glass produced is multi-phased.
- Let Y be the average tension on a roll of paper being wound. Then a OIR occurs under conditions in which the paper does not wind at all or when the paper breaks as it is being wound.
- Let Y be film thickness. Then a OIR occurs under conditions such that the film cannot be produced. This may occur when the melt index of a combination of resins used to make the film is either too low or too high.
- Let Y be a measure of a liquid. A OIR occurs if the material produced was a gas or solid.

To quantify the effects that OIR's have on an experiment, we formulate the failed-run problem into a mathematical one by making the following assumption.

***Restricted-Information Region Assumption.* If a treatment combination causes the response Y to change by more than $\pm R$ from its value under the current conditions (at 0, the center of the design) the run contains no information to the experimenter.**

The value of R is typically not observed, but we can still imagine $\pm R$ to be the boundary values that correspond to a OIR. (For simplicity here, we assume a “ \pm ” symmetry.) We will sometimes act as if we could actually measure these results outside of $\pm R$ and use them in the analysis. If we can use such readings, we will call this the *perfect world*. If these results produce OIR's, we will call this the *real world*.

So, the double-edged sword of FF's is that they explore much larger regions of design space than their 1FAT counterparts. This helps to produce the k -efficiency

numbers, but also increases the risk of a OIR. To make the comparison, it is reasonable to rescale one of the designs, say the FF, so that in the experiment the probabilities of a OIR—or more generally of an equal number of OIR’s—are equal.

5. AN EXAMPLE OF AN INEFFICIENT FF DESIGN

For concreteness, first consider the study of $k = 16$ factors in $n = 32$ runs. Assume, without loss, that $E[Y] = 0$ at the $\mathbf{0}$ conditions. Now, we must have $\sigma > 0$ so that efficiency is defined. On the other hand, when $\sigma > 0$ a run in the design may be a OIR only because of the random variation that a $\sigma > 0$ induces in the data. To avoid this complexity in this section—we will address it later—but to allow efficiency to be defined, let $\sigma \downarrow 0$. For this example, also assume that all 16 half-effects are +1. Let R be such that the most extreme of the 1FAT-design results are just inside the boundary of $\pm R$ —this is done simply to standardize on one of the designs. Then $R = 1^+$, where x^+ denotes a number slightly larger than x . Here, all the data are either +1 or -1, so *every* run in the 1FAT design is just inside the boundary.

On the other hand, say we used the principal generator of a standard FF design. This turns out to result, in the perfect world, in data with one value of -16, one value of +16, and 30 values of 0. But in the real world, we have 2 OIR’s and 30 0’s. In order to have information at every run, we would have needed to compress the design matrix from \mathbf{X} to $\mathbf{X}/16$. But this leads to $\text{Var}(b_{1FAT}) = \sigma^2/2$ and $\text{Var}(b_{FF}) = 8\sigma^2$. In the perfect world, the FF was 16 times *more* efficient. In the real world, it is 16 times *less* efficient. In the general case, the efficiency in the perfect world of k becomes $1/k$ in the real world.

This is a revealing example of how much the efficiency of the FF can be overstated. However, it is also unrealistically pessimistic, so we next consider how the efficiency of the FF compares to 1FAT under more realistic scenarios.

6. EFFECT-BASED EFFICIENCIES

We first list our assumptions:

1. Assume that the current operating conditions for the k factors are $\mathbf{0}$, and that $E[Y] = 0$ at these conditions.
2. Assume that only main effects exist and are linear. (Not all main effects need be active.)
3. Assume that, in the perfect world, we observe $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The $n \times k$ design matrix \mathbf{X} is, at least initially, a matrix whose elements are -1, 0, or 1, with an equal number of ± 1 ’s in each column (the intercept term is, without loss of generality, excluded); the design is orthogonal; $\boldsymbol{\beta}$ is the vector of half-effects; and $\boldsymbol{\varepsilon}$ is the vector of independent errors with mean 0 and variance σ^2 .

We will consider the following scenarios:

1. Let the sample size $n = 16, 32, 64$, with $k = n/2$.
2. Let the vector $\boldsymbol{\beta}$ of half-effects arise from one type of a certain set of distributions. Each type of distribution is defined so that the absolute value of the non-zero half-effects has a mean of 1. We have also defined each distribution to have a mean of 0. This is reasonable, because a zero-mean effect distribution corresponds to the reasonable practice of randomly deciding whether to allow the -1 level to refer to the high or low setting of a factor. Denote a generic element of $\boldsymbol{\beta}$ by β , and let each such element be sampled independently from one of the following types of distributions:
 - (a) pH (Half). With probability $p/2$, $\beta = 1$. With probability $p/2$, $\beta = -1$. With probability $1 - p$, $\beta = 0$.
 - (b) H (Half). With probability $1/2$, $\beta = 1$. With probability $1/2$, $\beta = -1$.
 - (c) pN. With probability p , $\beta \sim N(0, 1) \sqrt{2\pi}/2$. With probability $1 - p$, $\beta = 0$.
 - (d) N (Normal). $\beta \sim N(0, 1) \sqrt{2\pi}/2$.
 - (e) pChi. With probability $p/2$, $\beta \sim \chi^2(1)$. With probability $p/2$, $\beta \sim -\chi^2(1)$. With probability $1 - p$, $\beta = 0$.
 - (f) Chi. With probability $1/2$, $\beta \sim \chi^2(1)$. With probability $1/2$, $\beta \sim -\chi^2(1)$.
3. Let the design matrix \mathbf{X} correspond to either a 1FAT design, as described in Section 1, or a FF design. The generators for the FF designs correspond to those in Hamada and Wu (2000, pp. 194–198).
4. Let σ be either $\sigma \downarrow 0$ or $\sigma = 1$, as follows:
 - (a) $\sigma \downarrow 0$. For simulations, we simply set $\sigma = 0$, and allow values of $\mathbf{X}\boldsymbol{\beta}$ in $[-R, R]$ to be considered as containing information in the real world. But we also act as if σ were slightly greater than 0, so that efficiency measures can be calculated.
 - (b) $\sigma = 1$. With this value, the mean absolute value of the active half-effects is the same as σ . In the perfect world, this can easily be detected by the FF design, but not by the 1FAT, because $E[|Half-Effect|] / s.d._{Half-Effect}$ is $\sqrt{2}$ for 1FAT and \sqrt{n} for FF.

Additional values for σ could have been chosen as well, but we felt that these two would give a sense of the importance of excess variation.

5. Consider values for p of 0.2 and 0.4. This covers a reasonable range of values found in practice. Box and Meyer(1986) suggest that p is around 0.2.
6. Assume that the experimenter may be willing to accept the possibility of a small number of OIR values, and hence would need to drop these in the analysis. We will assume the experimenter is willing to drop either $D = 0$ or 2 values. (But note that being willing to have *any* OIR values may not be a good strategy in practice—in the FF example we used above, accepting 2 OIR’s meant that it would be impossible to see that *any* effects were active. See John (1979), who discusses this non-robust property of FF designs in a more general setting.) So $D = 0$ ($D = 2$) would require that the n^{th} ($(n - 2)^{\text{th}}$) order statistic of the $\{|Y_i|\}$ be less than or equal to R .
7. Assume that the experimenter wants the probability to be $q = 0.80$ or 0.95 that D OIR values or fewer will occur in the experiment.

We will study 9 types \times 2 σ ’s \times 2 D ’s \times 2 q ’s \times 3 n ’s= 144 combinations, under each of the FF and 1FAT designs. The nine types are pH2 (i.e., pH with $p = 0.2$), pH4, H, pN2, pN4, N, pChi2, pChi4, and pChi. For each *type- σ - n* combination, we generated 10,000 simulations of the experiment under the FF design, and 10,000 simulations under the 1FAT design. For each simulation, we found the n^{th} order statistic and the $(n - 2)^{\text{th}}$ order statistic of the $\{|Y_i|\}$, corresponding to dropping either $D = 0$ or $D = 2$ data points. These 10,000 simulated values generated an estimate of the distribution of these order statistics, from which we estimated their 80th and 95th percentiles, corresponding to the two q values of interest. These percentiles provided estimates of what R would need to be to satisfy the requirements for a particular design combination.

Only the relative values of R matter when we study efficiency comparisons, so we can study each combination as follows:

1. Find what R value is needed by each of the two designs, say $R.1FAT$ and $R.FF$, to ensure that there is a probability q of accepting D values or less of OIR’s.
2. Find how much the FF design needs to be scaled down, from \mathbf{X} to $\mathbf{X}/scale$, where $scale = R.FF/R.1FAT$, so that these probabilities are equal.
3. Measure the efficiency of FF relative to 1FAT as $k/(scale)^2$. This is the real-world efficiency, compared to the perfect-world efficiency of k .

We illustrate the H and N scenarios for $n = 32$, $k = 16$ with figures. Figures 3 and 4 show the type H scenarios for $\sigma = 0$ and $\sigma = 1$, while Figures 5 and 6

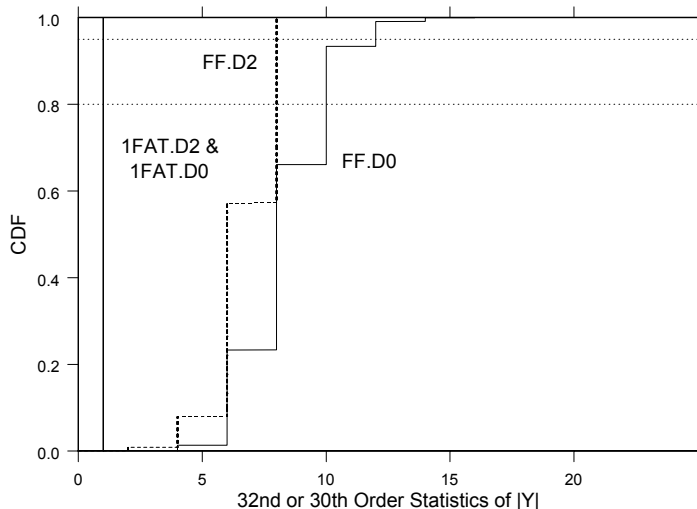


Figure 3: Type H, $\sigma = 0$, Scenarios.

show the corresponding type N scenarios. Each figure graphs the distribution of the 32nd (corresponds to $D = 0$) and 30th ($D = 2$) order statistics of the $\{|Y_i|\}$, based on the 10,000 simulations of the scenario. Consider Figure 4, and look only at the $D = 0$ cases for 1FAT and FF. For the 80th percentile, the $q = .80$ case, we find that $R.1FAT \approx 3.5$ and $R.FF \approx 10.7$, so $scale \approx 3.1$ and the efficiency is about $16/(3.1)^2 = 1.7$. So, the FF design is still more efficient, but the real-world efficiency of 1.7 is far less than the perfect-world efficiency of 16.

Table 1 shows the efficiency results for all the scenarios. Consider the $D = 0$ results for $n = 32$. Aside from the extreme type H scenario for $\sigma = 0$, the FF design performed at least as well as the 1FAT. However, the real-world efficiencies are quite less than the perfect-world efficiency of 16. By comparison, if we had formally used the reduced-model efficiency formula $\sqrt{k/k'}$, the efficiencies would be based only on the proportion of active effects, which on average are $p = 0.2, 0.4,$ and 1. This leads to efficiency values of 2.2, 1.6, and 1, respectively.

An ANOVA of the efficiency values in Table 1 reveals that, essentially,

1. All factors except q were active.
2. Only *type* interacted with the other factors.

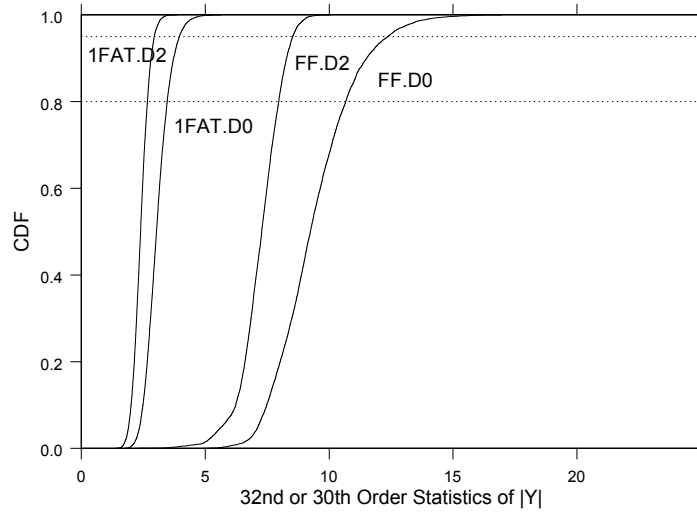


Figure 4: Type H, $\sigma = 1$, Scenarios.

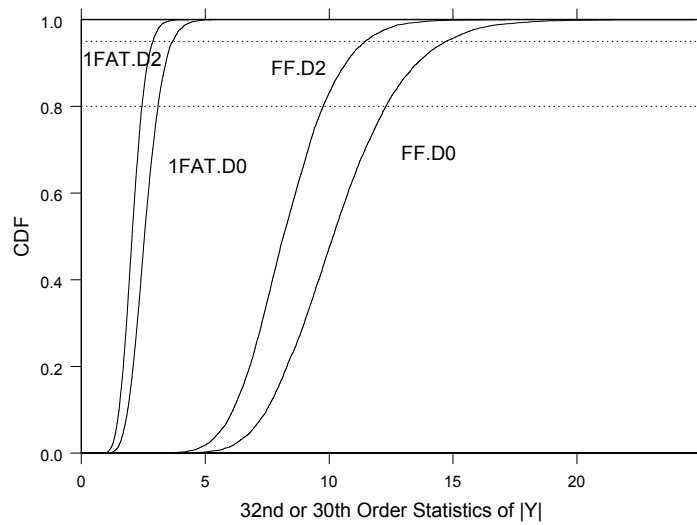


Figure 5: Type N, $\sigma = 0$, Scenarios.

| n | k | σ | D | q | Type | | | | | | | | | |
|-----|-----|----------|-----|------|------|-----|-----|-----|-----|-----|-------|-------|-----|--|
| | | | | | pH2 | pH4 | H | pN2 | pN4 | N | pChi2 | pChi4 | Chi | |
| 16 | 8 | 0 | 0 | 0.80 | 2.0 | 0.5 | 0.2 | 3.7 | 2.0 | 1.0 | 4.8 | 3.3 | 2.0 | |
| | | | | 0.95 | 0.9 | 0.3 | 0.1 | 2.9 | 1.8 | 1.0 | 4.5 | 3.3 | 2.3 | |
| | | | 2 | 0.80 | 2.0 | 0.9 | 0.5 | 0.7 | 1.0 | 1.0 | 0.2 | 0.5 | 0.9 | |
| | | | | 0.95 | 0.9 | 0.5 | 0.5 | 1.0 | 1.1 | 1.0 | 0.4 | 0.7 | 0.9 | |
| | | 1 | 0 | 0.80 | 3.9 | 2.7 | 1.7 | 3.8 | 2.8 | 1.6 | 4.3 | 3.2 | 2.1 | |
| | | | | 0.95 | 3.8 | 2.7 | 1.6 | 3.5 | 2.6 | 1.6 | 4.6 | 3.4 | 2.3 | |
| | | | 2 | 0.80 | 3.3 | 2.6 | 2.0 | 2.6 | 1.9 | 1.5 | 2.4 | 1.3 | 1.1 | |
| | | | | 0.95 | 3.1 | 2.6 | 2.2 | 2.0 | 1.8 | 1.4 | 1.2 | 1.0 | 1.1 | |
| 32 | 16 | 0 | 0 | 0.80 | 1.0 | 0.4 | 0.2 | 3.6 | 2.0 | 1.0 | 6.3 | 3.8 | 2.3 | |
| | | | | 0.95 | 0.6 | 0.3 | 0.1 | 3.2 | 1.9 | 1.0 | 6.5 | 4.1 | 2.7 | |
| | | | 2 | 0.80 | 1.0 | 0.6 | 0.3 | 1.5 | 1.4 | 1.0 | 1.0 | 1.1 | 1.2 | |
| | | | | 0.95 | 1.0 | 0.4 | 0.3 | 1.6 | 1.4 | 1.0 | 1.2 | 1.3 | 1.3 | |
| | | 1 | 0 | 0.80 | 4.8 | 3.2 | 1.7 | 4.8 | 3.2 | 1.7 | 6.1 | 4.2 | 2.6 | |
| | | | | 0.95 | 4.6 | 3.1 | 1.6 | 4.5 | 3.0 | 1.7 | 6.8 | 4.6 | 3.0 | |
| | | | 2 | 0.80 | 4.2 | 2.9 | 1.8 | 3.3 | 2.3 | 1.5 | 2.7 | 1.8 | 1.5 | |
| | | | | 0.95 | 4.0 | 2.9 | 1.9 | 2.6 | 2.1 | 1.5 | 1.9 | 1.6 | 1.5 | |
| 64 | 32 | 0 | 0 | 0.80 | 0.7 | 0.3 | 0.1 | 3.5 | 2.0 | 1.0 | 7.2 | 4.5 | 2.5 | |
| | | | | 0.95 | 0.5 | 0.3 | 0.1 | 3.3 | 2.0 | 1.0 | 7.9 | 5.2 | 3.1 | |
| | | | 2 | 0.80 | 0.9 | 0.5 | 0.2 | 2.1 | 1.6 | 1.0 | 2.1 | 2.0 | 1.6 | |
| | | | | 0.95 | 0.7 | 0.4 | 0.2 | 2.1 | 1.6 | 1.0 | 2.4 | 2.2 | 1.8 | |
| | | 1 | 0 | 0.80 | 5.7 | 3.4 | 1.7 | 5.6 | 3.4 | 1.8 | 8.0 | 5.3 | 3.0 | |
| | | | | 0.95 | 5.5 | 3.3 | 1.7 | 5.5 | 3.3 | 1.8 | 8.9 | 6.0 | 3.5 | |
| | | | 2 | 0.80 | 4.9 | 3.1 | 1.8 | 3.9 | 2.6 | 1.6 | 3.3 | 2.5 | 1.9 | |
| | | | | 0.95 | 4.6 | 3.0 | 1.7 | 3.4 | 2.5 | 1.6 | 3.1 | 2.6 | 2.0 | |

Table 1: Real-World Efficiencies of FF Designs Relative to 1FAT Designs.

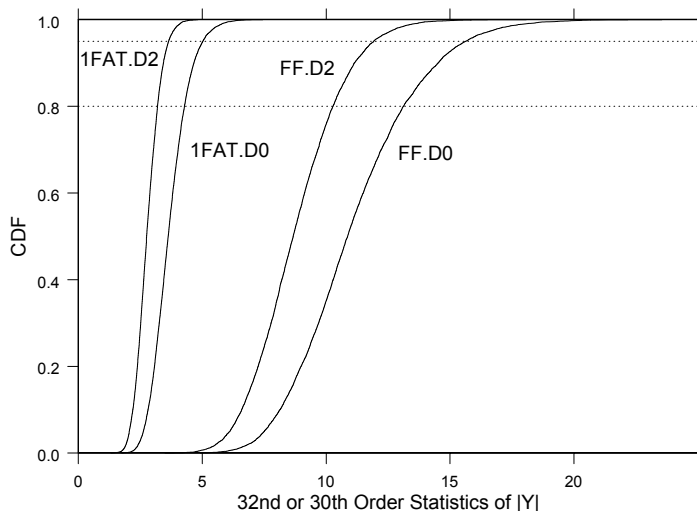


Figure 6: Type N, $\sigma = 1$, Scenarios

Based on this, we generated the interaction plots shown in Figures 7 to 9. (Use of log-efficiency means is mathematically preferred, but leads to the same conclusions here while being less intuitive.) Not surprisingly, Figure 7 shows that the FF design performed relatively worse as n decreased or as the number of active effects increased (e.g., from pN2 to pN4 to N). It also performed relatively better as the effects became more skewed away from 0 (H to N, N to Chi), because this skewness affects R.1FAT more than R.FF. In Figure 8, this also explains why, when $D = 2$, the efficiency for pChi2 and pChi4 dropped so much. But, as we noted earlier, the $D = 2$ results should be viewed cautiously—we believe that the $D = 0$ features should be the ones to guide the experimenter who is trying to estimate the real-world efficiency. From Figure 9, most of the $type \times \sigma$ interaction is due to the extreme case, where $type = \text{pH2, pH4, or H}$ and $\sigma = 0$. This figure also shows that the efficiency increases with σ .

Finally, note from Table 1 that the real-world efficiencies for $type = \text{N}$, $\sigma = 0$, cases are all equal to 1.0 in the simulations. In fact, these efficiencies are exactly 1.

Theorem 2. *For the $type = \text{N}$, $\sigma = 0$, case, the real-world efficiency is equal to 1.*

The proof is contained in the Appendix.

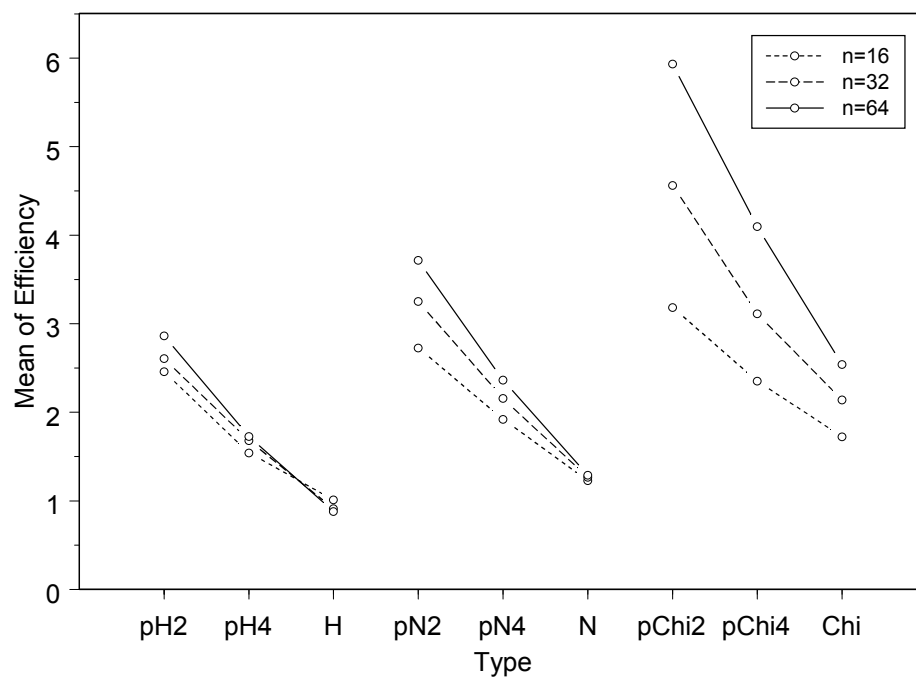


Figure 7: Interaction Plot for Efficiency: Type \times n

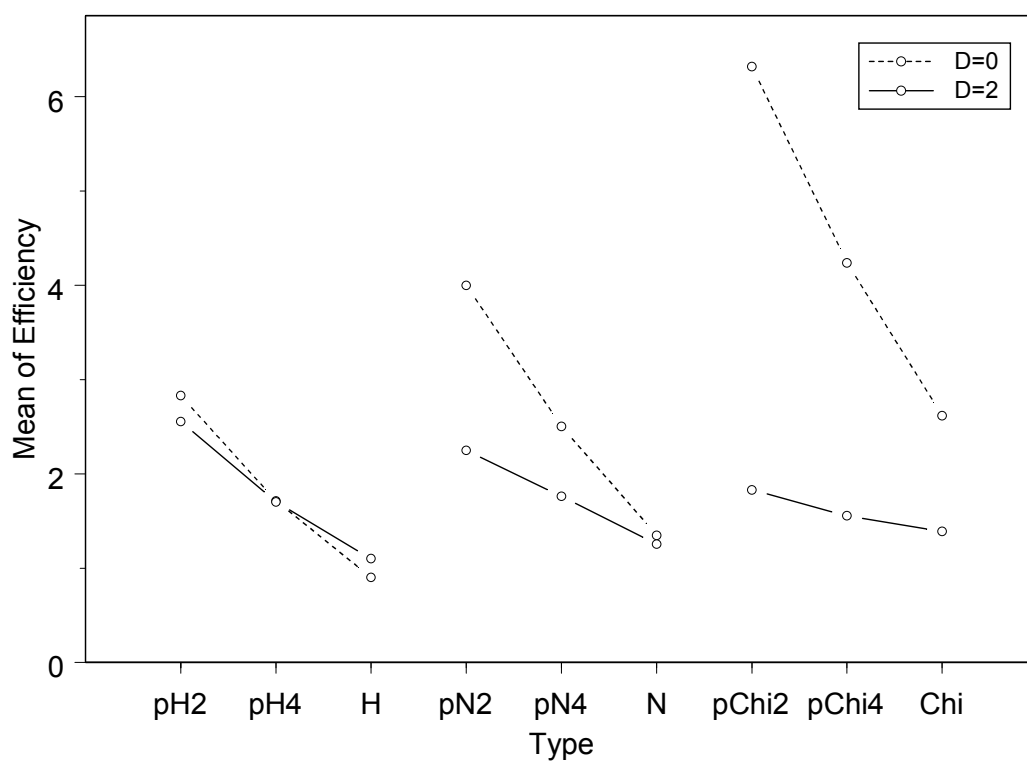


Figure 8: Interaction Plot for Efficiency: Type \times D

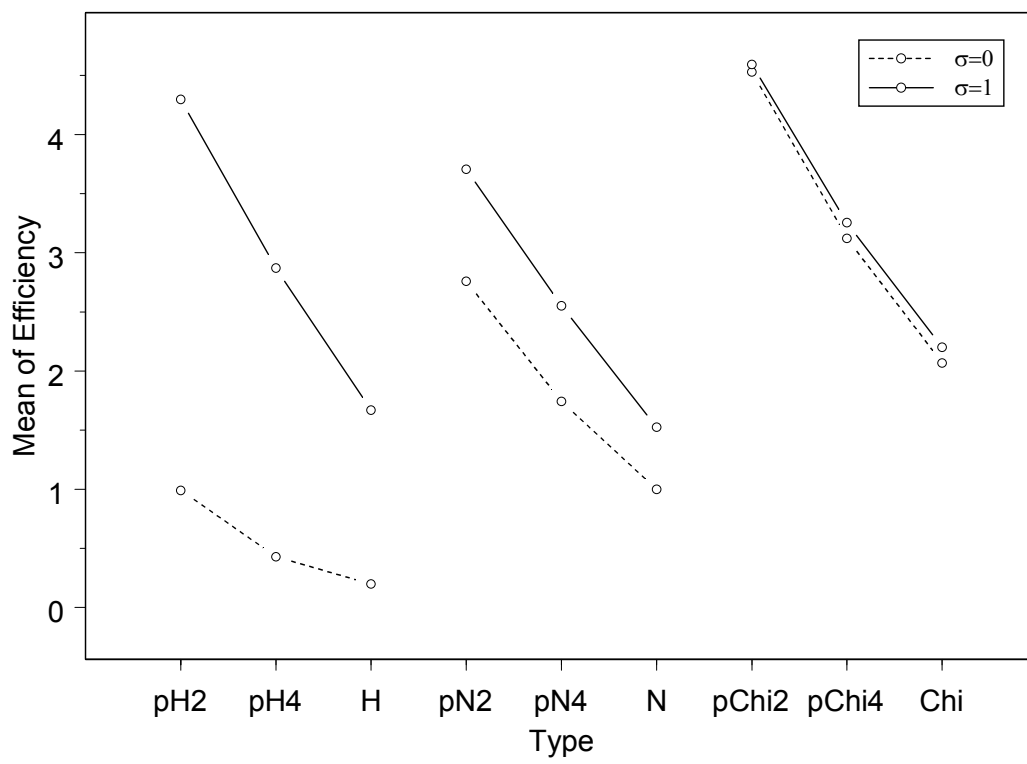


Figure 9: Interaction Plot for Efficiency: Type \times σ

| n | k | σ | D | q | Type | | | | | | | | | | |
|-----|------|----------|------|------|------|------|------|------|------|-----|-------|-------|-----|-----|-----|
| | | | | | pH2 | pH4 | H | pN2 | pN4 | N | pChi2 | pChi4 | Chi | | |
| 16 | 8 | 0 | 0 | 0.80 | 2.0 | 4.0 | 6.0 | 1.5 | 2.0 | 2.9 | 1.3 | 1.6 | 2.0 | | |
| | | | | 0.95 | 3.0 | 5.0 | 8.0 | 1.6 | 2.1 | 2.8 | 1.3 | 1.5 | 1.9 | | |
| | | 1 | 0 | 0.80 | 2.0 | 3.0 | 4.0 | 3.3 | 2.9 | 2.8 | 7.0 | 3.9 | 3.1 | | |
| | | | | 0.95 | 3.0 | 4.0 | 4.0 | 2.8 | 2.7 | 2.8 | 4.3 | 3.4 | 3.0 | | |
| | | | 2 | 0.80 | 1.4 | 1.7 | 2.2 | 1.4 | 1.7 | 2.2 | 1.4 | 1.6 | 1.9 | | |
| | | | | 0.95 | 1.5 | 1.7 | 2.2 | 1.5 | 1.7 | 2.2 | 1.3 | 1.5 | 1.9 | | |
| 32 | 16 | 0 | 0 | 0.80 | 4.0 | 6.0 | 10.0 | 2.1 | 2.8 | 4.0 | 1.6 | 2.1 | 2.7 | | |
| | | | | 0.95 | 5.0 | 7.0 | 12.0 | 2.2 | 2.9 | 4.0 | 1.6 | 2.0 | 2.4 | | |
| | | | 2 | 0.80 | 4.0 | 5.0 | 8.0 | 3.2 | 3.4 | 4.0 | 4.0 | 3.8 | 3.6 | | |
| | | | | 0.95 | 4.0 | 6.0 | 8.0 | 3.2 | 3.4 | 4.0 | 3.6 | 3.5 | 3.6 | | |
| | | | | 1 | 0.80 | 1.8 | 2.2 | 3.1 | 1.8 | 2.2 | 3.1 | 1.6 | 2.0 | 2.5 | |
| | | | | | 0.95 | 1.9 | 2.3 | 3.1 | 1.9 | 2.3 | 3.1 | 1.5 | 1.9 | 2.3 | |
| | | 2 | 0.80 | 1.9 | 2.3 | 3.0 | 2.2 | 2.7 | 3.2 | 2.4 | 3.0 | 3.3 | | | |
| | | | 0.95 | 2.0 | 2.4 | 2.9 | 2.5 | 2.8 | 3.3 | 2.9 | 3.1 | 3.3 | | | |
| | | 64 | 32 | 0 | 0 | 0.80 | 7.0 | 10.0 | 16.0 | 3.0 | 4.0 | 5.7 | 2.1 | 2.7 | 3.5 |
| | | | | | | 0.95 | 8.0 | 11.0 | 18.0 | 3.1 | 4.0 | 5.7 | 2.0 | 2.5 | 3.2 |
| | | | | | 2 | 0.80 | 6.0 | 8.0 | 12.0 | 3.9 | 4.4 | 5.7 | 3.9 | 4.0 | 4.5 |
| | | | | | | 0.95 | 7.0 | 9.0 | 14.0 | 3.9 | 4.4 | 5.7 | 3.7 | 3.8 | 4.3 |
| 1 | 0.80 | | | | | 2.4 | 3.1 | 4.3 | 2.4 | 3.1 | 4.2 | 2.0 | 2.5 | 3.3 | |
| | 0.95 | | | | | 2.4 | 3.1 | 4.4 | 2.4 | 3.1 | 4.3 | 1.9 | 2.3 | 3.0 | |
| 2 | 0.80 | | | 2.5 | 3.2 | 4.3 | 2.9 | 3.5 | 4.5 | 3.1 | 3.6 | 4.1 | | | |
| | 0.95 | | | 2.6 | 3.2 | 4.3 | 3.1 | 3.6 | 4.5 | 3.2 | 3.5 | 4.0 | | | |

Table 2: Scale Reduction of FF Design Relative to 1FAT Design

7. SCALING BACK THE FF DESIGN

Table 2 lists the scale values that generated Table 1 . This can be used to give an idea for how much to scale back the settings in a FF design versus where they would be for a 1FAT design.

Based on this information, the following strategy seems reasonable, at least for the $k = n/2$ situations considered in this paper:

1. Let the experimenter make best guesses for the actual settings to be used in the experiment if only one factor at a time would be changed, to try to ensure reasonably well that all the runs produce information.
2. Scale the settings down by using results such as those in Table 2 . If we use $D = 0$, $\sigma = 1$, and average the results for the pH2, pH4, pN2, pN4, pChi2, and pChi4 types, this table suggests that the settings should be scaled back by about a factor of 1.5 for $k = 8$, about 2.0 for $k = 16$, and about 2.5 for $k = 32$.

If we compare these scale-back values to the model-reduced approach of scaling back by $\sqrt{k'}$, these values correspond to a belief that the number of active factors will be 2.25 (28% of $k = 8$), 4 (25% of $k = 16$), and 6.25 (20% of $k = 32$), respectively.

This method can be used to compare any other FF designs to 1FAT designs, not simply the $n = 2k$ designs studied here. This strategy does not take interactions into account, but we do not anticipate that the small number of interactions that may surface should affect the strategy by much.

8. APPENDIX

8.1. Proof of Theorem 1. Result 1 of the proof is well-known and has already been noted in the article.

We will show Results 2, 3, 4, by first assuming Result 3 is correct and then using it on the 1FAT design matrix \mathbf{X} to prove Results 2 and 4.

First decompose \mathbf{R} into $[\mathbf{1} \mid \mathbf{R}_1] / \sqrt{k}$. By assumption, \mathbf{R}_1 is a saturated design of $k - 1$ factors in k runs. Such designs are well known—in particular they exist whenever k is a power of 2, and they are resolution III. Now, \mathbf{X} can be written as $\mathbf{X}' = [-\mathbf{I}_k \mid \mathbf{I}_k]$, where \mathbf{I}_k is the identity matrix of dimension k . A resolution III design of $k - 1$ factors in k runs can be folded over by running the design $-\mathbf{R}_1$, and a k^{th} factor may be added to the resulting $2k$ -run design by running that factor at one level for the first k runs, and the other level for the last k runs. This produces a resolution IV FF design. But this operation is the same as $\mathbf{X}[\mathbf{1} \mid \mathbf{R}_1] = \mathbf{X}\mathbf{R}\sqrt{k}$. We want to use the operation $\mathbf{X}\mathbf{R}$ instead, because \mathbf{R} is a rotation. This proves Result 4.

To prove Result 2, note that $\mathbf{X}\mathbf{R}\sqrt{k}$ has values of $+1$ and -1 in each column, so the rotated design $\mathbf{X}\mathbf{R}$ has values of $+1/\sqrt{k}$ and $-1/\sqrt{k}$. Thus, its relative efficiency of $1/k$ with respect to the FF design becomes 1 if the design is scaled up by the \sqrt{k} factor.

Result 5 follows immediately from the construction of the resolution IV FF given in the proof.

8.2. Rotation of a 3-Factor 1FAT. The original 1FAT design was rotated in various ways, based on code written in MathCad 2001i Professional to find a rotation that would yield a minimum-volume hypercube hull. A number of starting values of the rotation matrix was used. The initial design, and resulting optimal rotation and rotated design are, respectively,

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 2/3 & 1/3 & 2/3 \\ 1/3 & 2/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix} \quad \begin{bmatrix} -2/3 & -1/3 & -2/3 \\ -1/3 & -2/3 & 2/3 \\ 2/3 & -2/3 & -1/3 \\ 2/3 & 1/3 & 2/3 \\ 1/3 & 2/3 & -2/3 \\ -2/3 & 2/3 & 1/3 \end{bmatrix}$$

8.3. Proof of Theorem 2. We want to show that the real-world efficiency is 1 in the $type = N$, $\sigma = 0$, cases that we have considered. Let β denote the random vector of effects, and recall that the elements of β are *i.i.d.* normal with mean 0 and with a variance that we will call σ_β^2 . First, by the structure of the $1FAT$ design, the response may be written at $Y'_{1FAT} = [-\beta' \mid \beta']$. In particular, the first k rows of Y_{1FAT} , say $Y_{1,1FAT}$ are *i.i.d.* $N(0, \sigma_\beta^2)$, and the last k rows are simply $-Y_{1FAT}$. (Then, for example, the largest order statistic of the absolute value of the data Y_{1FAT} is equal in distribution to that of the largest order statistic of $\{U_i \sim i.i.d. \mid N(0, \sigma_\beta^2)\}$, $i = 1, 2, \dots, k$.)

Construct the FF design design \mathbf{X} with $k = n/2$ factors from a saturated $n/2$ -run design \mathbf{X}_0 in $k - 1 = n/2 - 1$ factors. Do this by first prefixing \mathbf{X}_0 with a column of 1's to get $\mathbf{X}_1 = [\mathbf{1} \mid \mathbf{X}_0]$, next folding over \mathbf{X}_1 to get $-\mathbf{X}_1$, and then joining the two designs together as $\mathbf{X}' = [\mathbf{X}'_1 \mid -\mathbf{X}'_1]$. In particular, the response Y_{FF} has its first k rows equal to $Y_{1,FF} = \mathbf{X}_1\beta$. Now, note that \mathbf{X}_1 is a square, orthogonal, matrix, with $\mathbf{X}'_1\mathbf{X}_1 = k\mathbf{I}$, which means that the k elements of $Y_{1,FF}$ are *i.i.d.* $N(0, k\sigma_\beta^2)$, and the last k rows are simply $-Y_{1,FF}$. This means that Y_{FF} will have the same distribution as Y_{1FAT} if we rescale the FF design matrix from \mathbf{X} to \mathbf{X}/\sqrt{k} . But doing so changes the original efficiency of k to an efficiency of 1.

This result depends both on the normality of the elements of β as well as the foldover properties associated with the particular designs we have chosen to study, so this result should not be expected to extend to more general cases.

9. BIBLIOGRAPHY

REFERENCES

- [1] Box, G. E. P., and Cox, D. R. (1964), "An Analysis of Transformations," *Journal of the Royal Statistical Society, Ser. B*, 26, 211–252.
- [2] Box, G. E. P., and Hunter, J. S. (1961), "The 2^{k-p} fractional factorial designs. Part I (Corr: V5 p417)," *Technometrics*, 3, 311-351.
- [3] Box, G. E. P., and Meyer, R. D. (1986), "An Analysis for Unreplicated Fractional Factorials," *Technometrics*, 28, 11–18.

- [4] Cheng, Ching-Shui (1998), "Some hidden projection properties of orthogonal arrays with strength three," *Biometrika*, 85 , 491-495.
- [5] Cheng, Ching-Shui (1995), "Some projection properties of orthogonal arrays," *The Annals of Statistics*, 23 , 1223-1233.
- [6] Czitrom, V. (1999), "One-Factor-at-a-Time versus Designed Experiments," *The American Statistician*, 53, 126–131.
- [7] Buckner, J., Huang, R., Monnig, K. A., Broadbent, E. K., Chin, B. L., Henri, J., Sorell, M., and Venor, K., "Chapter 16–Modeling a Uniformity Bulls-Eye Inversion," in Czitrom, V., and Spagon, P. (1997), *Statistical Case Studies for Industrial Process Improvement*, Philadelphia: SIAM.
- [8] Daniel, C. (1973), "One-at-a-Time Plans," *Journal of the American Statistical Association*, 68, 353-360.
- [9] Hamada, M., and Wu, C. F. J. (2000), *Experiments: Planning, Analysis, and Parameter Design Optimization*, New York:Wiley.
- [10] John, Peter W. M. (1979), "Missing points in 2^n and 2^{n-k} Factorial Designs," *Technometrics*, 21 , 225-228.
- [11] Lucas, J. M. (1976), "Which Response Surface Design is Best: A Performance Comparison of Several Types," *Technometrics*, 18, 411–417.
- [12] Morris, M. D. (1991), "Factorial Sampling Plans for Preliminary Computational Experiments," *Technometrics*, 33, 161–174.
- [13] Phadke, M.S. (1989), *Quality Engineering Using Robust Design*, Englewood Cliffs:Prentice-Hall.
- [14] Piepel, G. F., Cooley, S. K., Peeler, D. K., Vienna, J. D., and Edwards, T. B. (2002), "Augmenting a Waste Glass Mixture Experiment Study with Additional Glass Components and Experimental Runs," *Quality Engineering*, 15, 91-111.
- [15] Tang, Boxin (2001), "Theory of J -characteristics for fractional factorial designs and projection justification of minimum G_2 -aberration," *Biometrika*, 88, 401-407.