AIAA 2002-0746

Fractional Factorial Experiment Designs To Minimize Configuration Changes in Wind Tunnel Testing

R. DeLoach and D. L. Cler
NASA Langley Research Center
Hampton, VA

A. B. Graham
ViGYAN, Inc.
Hampton, VA

40th AIAA Aerospace Sciences Meeting & Exhibit
14-17 January 2002
Reno, NV

FRACTIONAL FACTORIAL EXPERIMENT DESIGNS TO MINIMIZE CONFIGURATION CHANGES IN WIND TUNNEL TESTING

Richard DeLoach* and Daniel L. Cler†
NASA Langley Research Center, Hampton, VA 23681-2199

Albert B. Graham‡
ViGYAN, Inc., Hampton, VA 23666-1325

Abstract

This paper serves as a tutorial to introduce the wind tunnel research community to configuration experiment designs that can satisfy resource constraints in a configuration study involving several variables, without arbitrarily eliminating any of them from the experiment initially. The special case of a configuration study featuring variables at two levels is examined in detail. This is the type of study in which each configuration variable has two natural states – "on or off", "deployed or not deployed", "low or high", and so forth. The basic principles are illustrated by results obtained in configuration studies conducted in the Langley National Transonic Facility and in the ViGYAN Low Speed Tunnel in Hampton, Virginia. The crucial role of interactions among configuration variables is highlighted with an illustration of difficulties that can be encountered when they are not properly taken into account.

The application of a sequential test strategy is illustrated for configuration testing, in which a single, large-scale test matrix is replaced with a coordinated series of smaller tests. Information obtained earlier on informs decisions made in subsequent tests in the same series. Each test typically features a small subset of all the possible combinations of configuration variable levels, judiciously selected to quantify main effects and likely lower-order interactions. This information is obtained at the expense of reduced information about higher-order interactions that are less likely to exist, and less likely to be important if they do exist. Substantial cost and cycle-time savings are achieved at the expense of somewhat reduced precision in estimating main and interaction effects. This reduced precision can cause some ambiguity in the estimated magnitude of those effects, especially for the case in which significant higher-order interactions exist. Techniques are discussed for augmenting the experiment design in order to efficiently resolve these ambiguities. This augmentation phase enables higher precision to be achieved for the case of main effects and/or interactions identified as especially important in earlier fractional factorial testing, without requiring an unnecessarily large number of additional configurations to be set.

* Senior Research Scientist
† Test Engineer
‡ Research Engineer

Nomenclature

alpha	angle of attack
CMXS	stability-axis roll moment coefficient
defining relation	a device for determining aliasing patterns in fractional 2-level factorial designs
design generator	one of the candidate effects in a two-level factorial design used to determine the assignment of individual data points to specific design fractions.
design matrix	a test matrix augmented with columns for higher-order terms and a column of +1 values
design space	a Cartesian coordinate system in which each axis represents an independent variable. Every point in this space corresponds to a unique combination of independent variables.
factor	an independent variable
GWB	Generic Winged Body
interaction	a condition in which the effect on system response of a change in one independent variable depends on the level of another
LEX	leading edge extension
main effect	change in response due to change in specified independent variable

MDOE	Modern Design of Experiments
OFAT	One Factor At a Time
table of signs	a design matrix for a two-level factorial experiment in which low and high values of each factor are represented by ±1 values.
test matrix	an array of numbers in which each column corresponds to a factor and each row corresponds to a data point. The elements of the array are factor levels for that data point.

Introduction

A major category of wind tunnel testing that is broadly described as "configuration testing" is especially resource intensive because it typically involves numerous changes to the wind tunnel model itself. Unlike changes in model attitude (e.g., angle of attack, angle of sideslip) and changes in flow state (e.g., Mach number, dynamic pressure) that can often be performed remotely, configuration changes generally require access to the model. Accessing the model to change a configuration variable involves several labor-intensive and time-consuming procedures that are not required to change other types of variables. In major wind tunnel facilities, tunnel ingress and egress involves detailed procedures dictated by safety considerations, for example. The configuration changes often involve the construction and subsequent dismantling of scaffolding around the model. The configuration changes themselves typically entail the removal and insertion of dozens of fasteners and connectors of one type or another. Often the changed response surfaces must be re-dressed with grit or trip dots, which must then be inspected. Extra care and inspection are required to assure that intended changes in a configuration variable do not result in unintended changes to other elements of the model such as misalignments or surface blemishes, a risk that is incurred any time the model is physically handled during a test.

Other factors can be associated with changing configuration variables in special circumstances, which are not encountered with changes in other types of variables. For example, when configuration changes are made in a cryogenic wind tunnel, it is necessary to wait several hours before the model temperature stabilizes around a level that it can be safely handled. Several hours are usually then required to re-acquire cryogenic test temperatures. There is also the risk in any cryogenic tunnel entry that trace levels of moisture can be introduced into the circuit, which can condense as frost when the temperature is subsequently reduced to cryogenic levels. If the frost condenses on the model, it can directly affect the forces and moments, and if it condenses elsewhere, it can affect such factors as flow quality. Even if the tunnel is not cryogenic, a certain amount of time is required after flow is re-established following a tunnel entry, to stabilize testing conditions as much as possible.

A further complication in configuration testing is the fact that there is often a bewildering array of combinations of configuration variables. Imagine that the researcher is interested in examining two basic wing geometries, say, each with two aileron designs, two trailing edge flaps, two leading-edge slats and two sizes of leading edge extension (LEX), with each LEX either relatively inboard or relatively outboard of some reference location on the wing. In addition, there is interest in the effect of including canards or not, and of including strakes or not. There is also interest in all combinations of these configurations when the speed brake is deployed and when it is not, and for landing stability the gear up and gear down configurations are of interest. Finally, the researcher would like to examine a range of deflection angles for the flap and aileron for each combination of the other configuration variables.

The scale of this hypothetical specification of configuration variables is by no means extreme compared to real configuration testing. Yet even if the researcher were content to limit control surface deflections to only two levels ("low" and "high", say), this configuration test as described would entail over 4,000 configuration changes to test all combinations. Conditions vary from tunnel to tunnel but even under the best of circumstances this many configuration changes would be expected to take a lot of time. For example, while veteran researchers would probably consider four such configuration changes per eight-hour shift as ambitious for planning purposes, even at that rate this test would require approximately two years of wind tunnel time, even in a two-shift/day operation. Beyond the enormous direct operating expense and prolonged cycle time that such a comprehensive plan of test would entail, there are also subtle adverse impacts on quality that are easy to overlook. Over such a prolonged time period there can be systematic variations due to unexplained seasonal effects, for example, or long-term wear of facility subsystems. There can also be technology advances that result in subtle unexplained variations between sets of data acquired earlier and those acquired later in prolonged test programs. For reasons of cost, quality, and time, therefore, it is in the best interest of all concerned to reduce the scale of configuration testing as much as possible.

Faced with the practical necessity to scale the configuration test plan to meet resource constraints of time and money, the researcher has few options other than to select a subset of all the possible configurations for examination, postponing other configurations until another time. This is often accomplished by dropping certain variables from the test plan that in the researcher's judgment are of secondary interest to certain other variables that have a high priority. Perhaps this time we will not include the canards or the strakes, for example, and we will save the gear and speed brake studies for another time. We may also only focus on one of the two candidate wing shapes. These changes would reduce the configurations in the original plan to a number that could be examined in 3-6 weeks in a two-shift operation, assuming 2-4 configurations per shift. This would still be an ambitious plan, but a plausible one.

There are obvious shortcomings to this type of concession to resource constraints, and there are also other shortcomings that may be less obvious but potentially even more serious in that they can lead to improper inferences. Clearly the test will suffer from a lack of information obtained about the main effects of the variables deleted from the plan. However, in addition to the absence of information on the main effects of individual variables, there is also lost information on interaction effects. In the hypothetical example we are considering, we have decided to retain the LEX for investigation. However, it is entirely possible that the performance of the LEX would be different if we included canards in the configuration, for example. So by dropping the canards, not only do we lose information about their direct effects on the forces and moments, but we also lose information on how they impact the performance of the leading edge extension. A decision to select one LEX geometry over another might be made differently, for example, if the influence that canards and strakes have on LEX performance was better understood, or if any of the other deleted configurations had been examined in conjunction with the LEX. The same is obviously true for interactions involving other configuration variables as well.

We will examine a strategy for selecting a subset of configurations that does not require the wholesale deletion of independent variables from the test plan. It is possible to retain in the test all of the original independent variables, but to judicially select only a relatively small subset of the possible combinations we would have set with a conventional full-scale test plan. Notwithstanding how small it is, the subset is chosen in such a way that it provides essentially all of the information that the full experiment would have yielded, but with a substantial reduction in cycle time and associated expenses. As an important bonus, this method also quantifies the interaction effects, and eliminates from the unexplained variance a large component attributable to systematic variations that persist over time. Such long-term unexplained systematic variation can be due to instrumentation drift, temperature or other environmental effects, and even operator fatigue or learning effects (a common phenomenon in which the operator's performance improves with practice). Such systematic variation generally accounts for considerably more uncertainty than can be attributed to the ordinary chance variations in data that comprise the main focus of conventional uncertainty analysis.

We assert that a great deal of wasted effort occurs in configuration testing that can be attributed to our reliance upon an especially inefficient experimental methodology known as one factor at a time (OFAT) testing. OFAT practitioners systematically vary one independent variable at a time while holding all other variables constant. The standard wind tunnel procedure of varying angle of attack systematically while holding constant all other variables such as Mach number and model configuration is an example of how OFAT methods are routinely used in experimental aeronautics. This same impulse to hold all other variables constant while examining only one at a time has an especially deleterious effect on the productivity of configuration testing. We will propose an alternative test strategy known as "factorial" testing, in which all independent variables are subject to change with each new data point. For example, where an OFAT configuration test plan might call for a series of points in which all other variables are held constant while the deflection angle of an aileron is changed, a factorial design would typically call for all other independent variables to be changed as well as the aileron. Factorial designs thus attack all independent variables in parallel, not serially, one at a time. Not only does this increase productivity significantly, requiring many fewer total configuration changes, it also facilitates the study of interactions among the independent variables that OFAT methods are not well-suited to quantify.

Efficient factorial designs have been widely used in industrial engineering disciplines for many years, especially those that focus on process and product optimization. They are less well known in experimental aeronautics, however, notwithstanding their considerable potential in such application as configuration testing. This paper is therefore intended primarily as a tutorial introduction of factorial methods that are not widely practiced at this time in the aerospace industry, but that nonetheless have significant potential for cost savings as well as for insights into the interaction effects that govern how independent variables operate jointly to influence system response. As noted, they can also improve the quality of experimental results substantially.

The Role of Interactions

Factorial experimentation has an important advantage over conventional OFAT testing in that it is capable of detecting and quantifying interactions among independent variables. This advantage can best be illustrated with an example. We will consider a two-factor configuration test that was recently conducted in the National Transonic Facility at Langley Research Center using an advanced extension of the basic factorial techniques we will introduce in this paper. These advanced techniques taken as a whole are referred to as the Modern Design of Experiments (MDOE), which is being developed at Langley as a proposed replacement for the weaker OFAT techniques in common use in the aerospace industry at the end of the 20th century.

MDOE extends factorial design to include various tactical quality assurance measures that enable the researcher to assume a greater role in ensuring quality through the design of the experiment.[1,2] MDOE methods also focus on matching resources to the specific objectives of an experiment to ensure that ample resources are planned for the tasks at hand but that resources are not wasted, as by the gratuitous acquisition data in volumes that far exceed what is needed to control inference error risk. (The MDOE productivity doctrine dictates that *inferences be drawn* at high rate, not that *data be collected* at a high rate.[3] Excess resources beyond what is needed for a particular purpose are therefore applied to achieving additional insights, rather than acquiring additional data to drive the inference error probability for one specific objective even lower, after risk levels declared acceptable by the principal investigator in the formal planning process have already been achieved.) The full application of MDOE methods also entails advanced analysis techniques, in which unexplained variance and the associated uncertainty in experimental results is precisely characterized. These methods also focus on forecasting the response of the system to a general population of potential independent variable combinations, based on the responses to a sample of independent variable combinations observed in the experiment.

The entire spectrum of MDOE methods is beyond the scope of this paper, which nonetheless does focus on a key MDOE concept; namely the advantage of factorial methods over OFAT techniques. Consider the configuration study referenced above, in which one of the objectives was to define the combination of flap deflection and gap between the flap and the trailing edge of the wing that maximized lift for the approach configuration of a commercial jet transport.[4] The conventional OFAT approach to this objective would be to hold one of the variables constant at some nominal level – gap, let us say – while systematically varying the other (flap deflection) to find where lift is greatest. Having thus determined the optimum flap deflection, the flap would be held constant in this position while the gap was varied, again seeking the greatest lift. (We emphasize that OFAT methods such as these were *not* used in this experiment. We cite them here simply to illustrate a typical OFAT approach to such problems. The MDOE method that was actually used resulted in a response model of lift as a function of flap deflection and gap that enables us to forecast lift within specified precision intervals for all combinations of flap deflection and gap in the design space. Here we simply use this response model to simulate the results that a typical OFAT approach would have generated, as a pedagogical demonstration.)

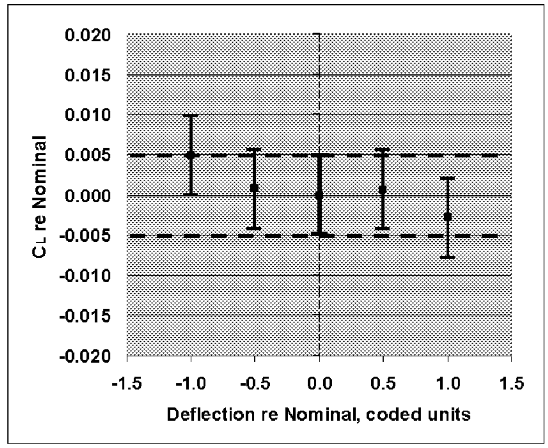

Figure 1. OFAT method reveals no C_L improvement when deflection varied while holding normalized gap constant at nominal level.

Figure 1 represents five measurements of approach lift at different flap deflections uniformly distributed over the full range of deflection angles of interest in this experiment. Error bars represent a "two-sigma" precision interval half-width of 0.005 in lift coefficient as specified in this experiment. The data are presented as change in lift coefficient relative to that which is achieved with the nominal flap deflection angle, plotted against departures from the nominal flap deflection

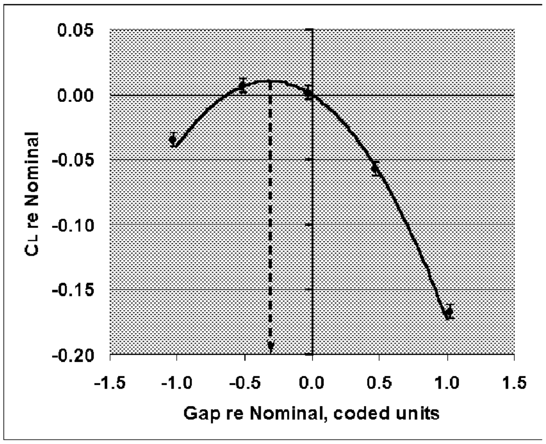

Figure 2. OFAT method reveals C_L improvement of 0.0106 when normalized gap reduced by 0.32 in coded units, deflection held constant at nominal level.

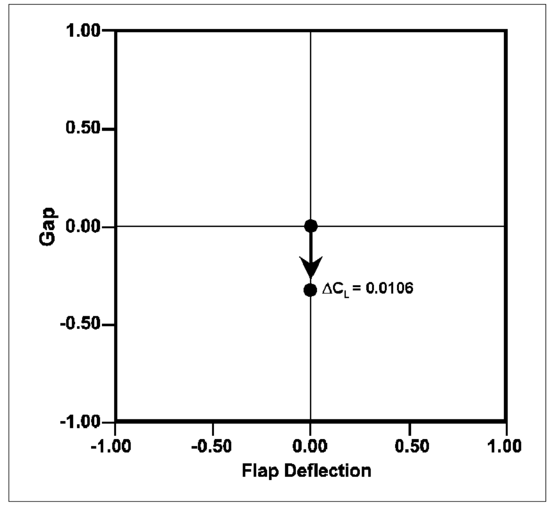

Figure 3. OFAT lift optimization. Changing one factor at a time starting with flap deflection angle.

angle in coded units, where 0 represents no change from nominal flap deflection. There is no obvious trend in the data and in fact the 95% confidence interval for the slope of a least-squares straight line fit to these data extends from -0.0072 to +0.0032. Since this interval includes 0, we conclude (with an inference error risk of no more than 5%) that with a nominal gap setting, changes in flap deflection over the range investigated provide no significant increase in lift. We therefore have no basis for recommending a flap deflection other than the nominal setting when the gap is nominal.

Having now established that the nominal flap deflection is optimal over the range we are considering, the OFAT practitioner would then hold that factor constant at the nominal (optimum) level and vary gap. Figure 2 shows what the result would be. It suggests that some increase in lift coefficient (0.0106) could be achieved by reducing the size of the gap by 0.32 in the coded units of the figure. The manufacturer would have to decide if the tooling costs and other production expenses associated with reducing the gap would be justified for an improvement in lift of only a factor of two times the experimental error budget. If not, then the decision would be to retain the nominal set points for flap deflection and gap. Figure 3 shows where the optimized set point appears within the design space when the factors are varied one at a time, starting with flap deflection.

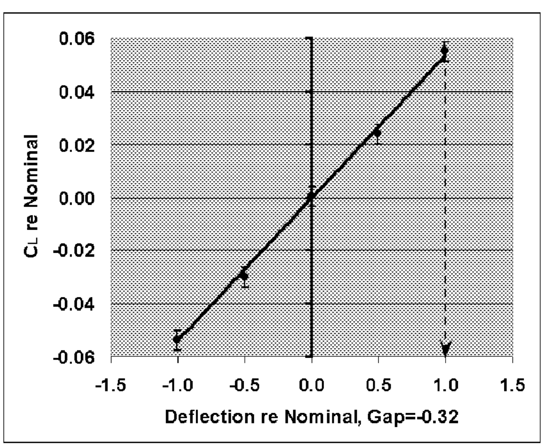

Figure 4. OFAT method reveals C_L improvement of 0.0538 when normalized gap held constant at -0.32 in coded units, and deflection increased by 1.0 in coded units.

One of the weaknesses of the OFAT approach to optimization can be illustrated by repeating the experiment with the order of the variables reversed; that is, with the deflection first held constant at its nominal level while the gap is optimized, followed by holding the gap constant at the optimum level while the deflection is optimized. Figure 2 corresponds to the first step, which suggests that a gap setting in coded units of -0.32 would maximize lift at nominal flap deflection, as before. Figure 4 is a plot of change in lift as a function of change in deflection from nominal, at the OFAT optimal gap of -0.32 in coded units. Clearly, there is a trend of increasing lift with increasing deflection at this gap setting. Note also that the flap deflection that maximizes lift apparently lies outside the range of deflection angles tested. That is, figure 4 suggests that the performance of the wing could be improved by increasing flap deflection beyond the largest value set in the experiment. However, within

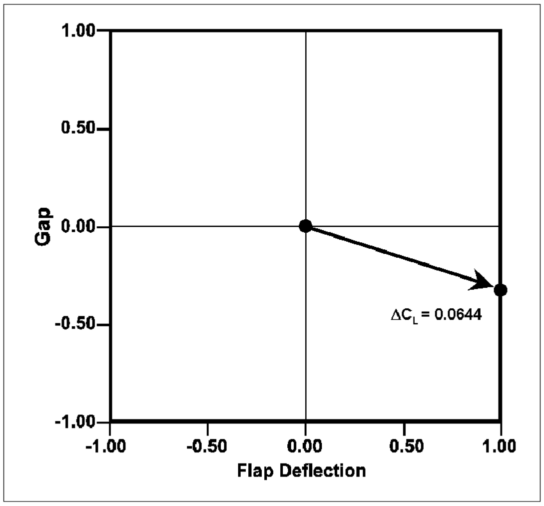

Figure 5. OFAT lift optimization. Changing one factor at a time starting with gap.

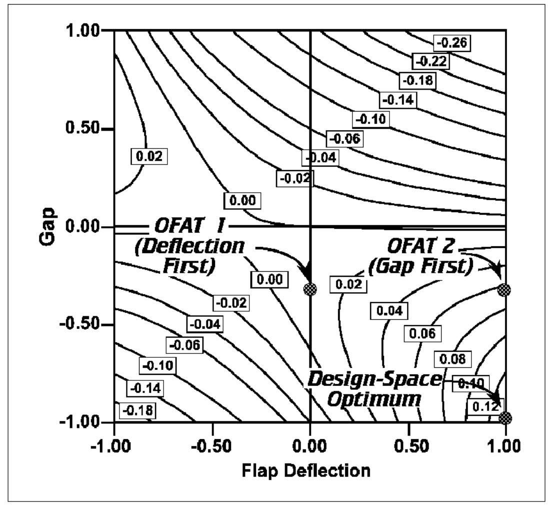

Figure 6. Changing both gap and deflection results in a greater improvement in lift than changing either one alone.

the design space of the experiment, the gap and deflection set points that appear to maximize lift with this ordering of the variables are, respectfully, -0.32 and +1.0 in coded units. Optimizing the deflection at a gap setting of -0.32 adds 0.0538 to the maximum lift coefficient. This is on top of the 0.0106 achieved by optimizing the gap, for a total OFAT improvement of 0.0644 when gap is optimized first and deflection is optimized second. Figure 5 shows where the optimized set point appears within the design space when the factors are varied one at a time, starting with gap.

The essence of Figure 5 can perhaps be most succinctly captured by noting that it is not figure 3. That is, the OFAT technique produces two substantially different results, depending on the order in which one factor is held constant while the other is varied. Since the order is determined arbitrarily by the researcher, this is a troubling development. However, it is not the *most* troubling development, as figures 6 and 7 reveal. The most troubling aspect of the OFAT approach to optimization is not so much that the results depend on the order in which the researcher decides to investigate the variables, unsettling as that is. The most serious difficulty is that under commonly occurring circumstances in which the independent variables interact, these OFAT optimization procedures produce the wrong answer no matter *which* variable is examined first.

Figure 6 reveals the source of the difficulty. It displays contours of constant lift coefficient throughout the design space. Each contour represents a change in lift relative to the original nominal combination of flap defection and gap that corresponds to the coordinates (0,0) in the coded units of this figure. It is clear from this figure why the OFAT procedure is inadequate in this situation. A path that traverses the design space in a direction that is parallel to either axis (which corresponds to holding one factor fixed while varying the other) will sometimes be more parallel to the contour lines and sometimes more perpendicular. Relatively little change in lift occurs on paths that are nearly parallel to the contour lines, while paths that are perpendicular encounter relatively more change in lift. For example, it is clear from figure 6 why the initial OFAT procedure resulted in an impression that with a nominal gap setting, changes in flap deflection had little effect on lift over the range of changes in this experiment. There is a contour of constant lift that runs through the original nominal configuration set point at (0,0) that is essentially parallel to the deflection axis *at that gap setting*. Therefore changes in deflection at this gap setting do little to change lift. However, a change in deflection clearly does result in a change in lift at *other* gap settings. For example, increasing deflection with the gap set at -1 in coded units results in a rapid increase in lift with flap deflection.

We have a situation in which the change in lift induced by a given change in flap deflection is different at one gap setting than another. That is, there is an interaction between flap deflection and gap. One of the interesting consequences of this interaction is that even though changing deflection from 0 to 1 in coded units has a negligible effect on lift with the gap held constant at 0, and changing the gap from 0 to -1 with the deflection held constant at 0 actually *decreases* the lift, when both changes are made together the lift increases

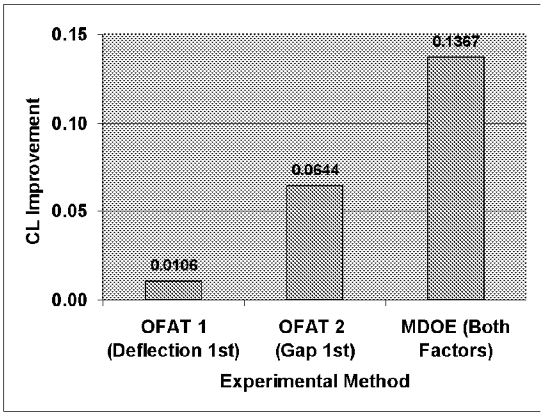

Figure 7. MDOE achieves greater lift improvement by accounting for interaction effects. OFAT results depend on the order that the variables are examined, because interactions are not taken into account.

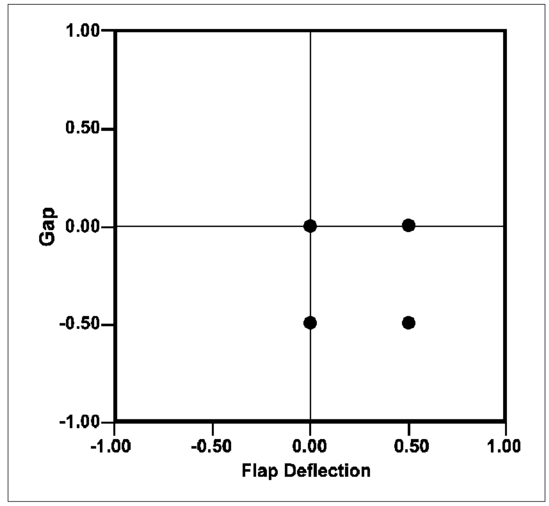

Figure 8. Factorial lift optimization. Changing both factors at the same time.

substantially. Figure 7 compares the two different OFAT results with the improvement that could be realized if interaction effects were taken into account by varying both factors at the same time rather than one at a time. Accounting for interactions in this way results in considerable improvement over either OFAT result.

Interaction effects such as this are not uncommon in experimental aeronautics and in fact they are the general rule; it is rare to encounter independent wind tunnel variables that do not interact with each other. Furthermore, interactions among more than two variables are not unusual. If the degree of interaction between flap deflection and gap is different at one angle of attack than another, say, then there would be a three-way interaction involving deflection, gap, and alpha. If this three-way interaction changed with Reynolds number, there would be a four-way or 4^{th}-order interaction, and so on. Understanding such interactions provides rich insights into the underlying physics of the process under study. Conversely, a failure to appreciate interactions can produce the kinds of results illustrated in figures 3 and 5 – results that are inconsistent, and erroneous. The fact that OFAT methods are so maladroit at illuminating interaction effects is one of their principal weaknesses, especially for applications as rich in interactions as configuration testing.

Two-Level Factorial Designs

Factorial experiment designs offer an attractive alternative to OFAT testing in those ubiquitous circumstances when interaction effects are important. Figure 8 illustrates how a factorial experiment might have been applied to the lift optimization problem, for example. This figure represents the layout for a particularly efficient class of factorial designs called *two-level* factorials, which we will treat in this paper. In a two-level factorial design, each factor is set at only two levels. These levels are typically assigned the descriptors "low" and "high" to distinguish one from the other, although the labels are arbitrary and apply equally well to quantitative variables that may actually have relatively low and relatively high levels such as flap deflection and gap in this example, or to categorical two-level variables that may have values such as "gear up, gear down", "canards on, canards off", etc.

The reader may be uneasy about an experiment design that features only two levels of important independent variables, because two levels permit only first-order pure effects to be described. Effects such as within-variable curvature require more than two points. However, early in a configuration experiment it may be premature to consider relatively high-order effects. Often the initial objective of a complex configuration study is to arrive quickly at insights that enable is to focus subsequent resources on variables that most significantly impact the response of the system. The two-level factorial designs that facilitate these early insights can be augmented efficiently later in the investigation, to provide additional levels of the variables for which higher-order effects are of interest.

Table 1 is a test matrix for the two-variable, two factor experiment represented schematically in Figure 8. The factors A and B represent flap deflection and gap, respectively, and the "high" and "low" levels are represented by "+1" and "-1", as is a common convention. For the design of figure 8, these low and high values refer to the coded values 0 and +0.5, respectively, for factor A (deflection). They refer to coded values -0.5 and 0, respectively, for factor B (gap). The third column lists the change in lift coefficient at each point relative to the nominal setting for deflection and gap, denoted by (0,0) in coded units.

Table I. Factorial *test* matrix, with response vector

A	B	ΔC_L
-1	-1	0.0070
-1	1	0.0000
1	-1	0.0500
1	1	-0.0018

We define the "main effect" for factor A as the change in response that results from changing factor A from its low level to its high level. Note in Table I, however, that we can make this change at either of the two levels of factor B. With factor B at its low level, changing factor A from low to high results in a change in system response from 0.0070 to 0.0500, or an A effect at low B of 0.0500 – 0.0070 = 0.0043. Similarly, the A effect at high B is -0.0018 – 0.0000 = -0.0018. The overall A effect is defined as simply the average of these two values, or (0.0043 – 0.0018)/2 = 0.0206. The B effect is defined analogously.

The definition of these main factor effects and the layout of Table I suggest a general algorithm for computing the main effects. The column of factor levels is simply multiplied term by term times the column of responses, and the products are summed. The result is then divided by the number of "+1" values in the factor column – always half the total number of rows in designs such as this. For example, using this algorithm to compute the main A factor effect, we see from Table I that it is [-0.0070 -0.0000 +0.0500 + (-0.0018)]/2 = 0.0206. Similarly, the B effect is -0.0294.

Note that the A effect is computed at both levels of the B factor and then averaged. No matter how many factors there are in the experiment, the general algorithm for estimating main effects computes them at all combinations of all levels of the remaining factors and then averages those results. We say because of this that the two-level factorial experiment design enjoys a "wide inductive basis". A large main effect is therefore a more reliable indicator that this variable is important than if it had been estimated for only a limited combination of the remaining variables.

Interaction effects can be estimated much like the main effects. We define the AB interaction as half the difference between the A effect at high B and the A effect at low B. (The factors can be reversed in this definition so that the AB interaction is also half the difference between the B effect at high A and the B effect at low A. Both give the same numerical result.) The A effect at high B was -0.0018 and at low B it was 0.0043. The AB interaction is therefore (-0.0018 – 0.043)/2 = -0.0224. The minus sign indicates that in this region flap deflection and gap are compensating variables. That is, the effect on lift of a decrease in flap deflection can be offset by an increase in gap.

Table II. Factorial *design* matrix, with response vector

Constant	A	B	AB	ΔC_L
1	-1	-1	1	0.0070
1	-1	1	-1	0.0000
1	1	-1	-1	0.0500
1	1	1	1	-0.0018

The same algorithm for computing main effects can be applied to compute the interaction effect if a suitable column of ±1 values is provided for the interaction. That column is easily obtained by multiplying corresponding levels in the main factor columns, producing the AB column in Table II. When a conventional test matrix such as in Table I is augmented with columns for higher order (interaction) terms and a column of +1 levels is added on the left side as in Table II, the resulting matrix is called the *design* matrix. The role of the column of +1's will be described shortly. For now, note that multiplying term by term the ±1 levels in the AB interaction column by the corresponding response values, summing, and dividing by the number of plus signs, yields [(+0.0070 -0.0000 -0.0500 +(-0.0018)] = -0.0224, the value we computed above from the definition of the interaction. That is, the same algorithm that worked for the main effects also works for the interactions. The design matrix, also called the table of signs, is therefore a convenient structure for computing all main effects and interactions.

The design matrix concept expands in a straightforward way to accommodate any number of factors. A three-factor two-level design would have columns for factors A, B, and C, say, with two-way interaction columns for the AB, AC, and BC interaction effects. There would also be a column for the ABC three-way interaction, generated by multiplying ±1 values in the A, B, and C columns term by term just as for the two-way interactions. The ABC three-way interaction is defined analogously to the two-way interactions, in that it is half the difference between the

AB interaction at high C and the AB interaction at low C. (As before, the definition is the same no matter which of the main effects is labeled A, B, or C.) However, it is generally more convenient to estimate the three-way interaction using the table of signs algorithm than to apply the definition directly.

Two-Level Factorials for Optimizing Response

In an optimization problem, it is convenient to envision the response as a function of the independent variables that can be represented as a surface over the design space. In this case of the lift optimization problem, the height of the response surface above the design space is proportional to lift, and we seek the location in the design space where this surface peaks. The geometric model is useful for distinguishing between locations that are near the peak and those that distant from it. Points that are not in the immediate vicinity of a peak are often on more or less planar slopes of the response surface. The researcher has an opportunity early in a factorial experiment to determine whether or not the design is centered near an extremum such as a peak.

One technique for estimating proximity to a response surface peak is to compare the average response at all points in the design (the four corners indicated in figure 8, say), with a point at the center of the design. The center point for the design represented in figure 8 would be at (+0.25, -0.25) in coded variables. If the response at the center point is roughly the same as the average of the responses at the corner points, it suggests that the response is relatively planar in this region and that the peak is therefore not in the local region. In the case of the lift optimization problem, the average of the corner responses can be estimated from the responses in Table I. The average is 0.0138. That is, the average deviation from the lift coefficient at nominal deflection and gap settings measured at the four corners of this design is 0.0138. The center-point value was 0.0204 in this case. The difference is only 0.0066, which is not large compared to the 0.005 error budget for this experiment. (Assuming an 0.005 two-sigma uncertainty in each individual estimate of lift coefficient, a difference of 0.0066 between the center point and a four-point average of the corners is too small to resolve with 95% confidence, for example.) We conclude therefore that because the response measured at the center of the design in figure 8 is roughly in the same plane as the responses measured at the four corner points, the design is located on some slope of the response surface and is not centered near the peak. We therefore will seek another more interesting region in the design space to

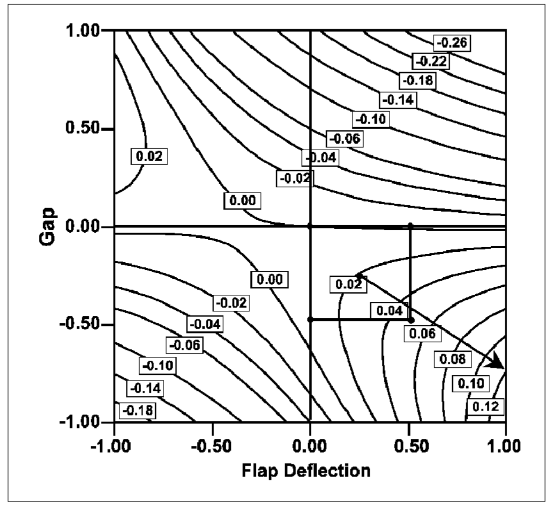

Figure 9. The path of steepest ascent estimated from the initial design of figure 8.

explore. This illustrates one way that two-level factorial designs can save resources. In this case a relatively small number of configuration changes has indicated that we are not in a very profitable region for further exploration, and therefore we will not invest further resources exploring this region for a peak in the response surface for lift. A conventional high volume data collection strategy might have invested substantial resources just "getting data" throughout this region and others, notwithstanding the fact that the peak of interest is nowhere nearby.

The initial two-level factorial experiment revealed that the peak in the lift response is somewhere else in the design space. It also indicates the direction in which to search for the peak, in that the main factor effects can be used to compute the direction cosines for a unit vector that points in this direction. The direction cosines are $A/\sqrt{A^2+B^2}$ and $B/\sqrt{A^2+B^2}$, where A and B are the main effects for flap deflection and gap, computed above as 0.0206 and -0.0294, respectively. The direction cosines for the A and B factors are thus 0.5738 and -0.8190. They define what is called the path of steepest ascent, which can be represented by a line with one end at the center of the design and the other end at any point for which the displacements from the center parallel to the A and B effects axes within the design space are in the ratio of 0.5738 to -0.8190. Such a line points at an angle of -55° relative to the flap deflection axis of the design space, as indicated in figure 9.

Note that the path of steepest ascent is approximately perpendicular to the contours of constant lift in the region of the design, indicating that it represents a relatively direct path toward the peak in the response surface. We would explore along this path, acquiring a relatively coarse sample of points intended simply to bracket the peak. We would then translate the design to our updated estimate of the peak's location, perform another two-level factorial experiment centered there, and again test for curvature by comparing the lift at the center of the design with the average lift at each of the corners as before. Another path of steepest ascent could be estimated if the new region still appeared planar. If the center-point test revealed curvature, however, the two-level factorial could be augmented with additional levels of flap deflection and gap to more adequately capture nonlinearities in the behavior of the response surface in this region. By limiting these additional configuration settings to the region of interest – near where the response surface is believed to peak – considerable time and effort can be saved that would otherwise have been devoted to a detailed exploration of relatively uninteresting regions of the design space. If these other regions are also of interest and the resources exist to explore them, then this can still be done. However, simple two-level factorial designs can provide the researcher with objective information for prioritizing the expenditure of resources. This process of beginning with relatively limited experiments and building toward more complex investigations is described as "sequential assembly."

As an epilogue to the lift optimization problem we note that sequential assembly was *not* employed when this experiment was actually conducted. We simply cite it to illustrate how these techniques might have been used to efficiently zero in on the peak in the lift response surface. In the actual experiment, the subject matter specialists felt that substantial interaction between flap deflection and gap was sufficiently unlikely to justify an assumption based on OFAT-only information that the peak was located near the center of the design space. The experiment therefore proceeded immediately to relatively expensive and time-consuming multi-level settings of flap deflection and gap in order to characterize the response surface throughout the design space and to quantify the peak lift. This lead to the definition of the contours in figure 9, which provide hindsight evidence that not all the initial assumptions were entirely justified.

For example, not only was the response surface peak not located near the center of the design space, it was not even within the design space at all. Furthermore, far from being negligible, the interaction effects in this experiment dominated even the main effects. Subject matter expertise correctly suggested that if flap deflection was important over the design-space range, it would be an increase in deflection that caused lift to improve. Likewise, if a change in gap was to be effective, it would be through a reduction in the gap. Therefore an alternative OFAT plan of test called for three single-point measurements; one at design-space coordinates (0,0) corresponding to nominal flap deflection and gap setting, one at coordinates (0,-1) corresponding to nominal flap deflection and minimum gap, and one at coordinates (1,0) corresponding to maximum deflection at nominal gap. The strategy of this classic OFAT design was to select among these three design-space locations the one that maximized lift.

If interactions could be safely neglected, this strategy would identify potential improvements attributable to changing deflection or gap, while minimizing the number of configuration changes. While this design could describe the effects of changing deflection OR gap, it could not describe the effects of changing deflection AND gap. Because of the strong interaction, deflection and gap were not simply additive effects as they would have been absent the interaction. Changing deflection alone would have produced an improvement too small to clearly resolve within experimental error and reducing gap alone actually would have caused a substantial decrease in lift. Changing them both together resulted in over 1000 counts of lift improvement.

It is a source of some frustration to advocates of formal experiment design that such results tend to be anecdotal. In this particular experiment, an interaction effect proved to be important. However, had the alternative OFAT experiment been conducted, the conclusion would have been that there is no evidence to suggest a set-point other than the original nominal level at (0,0) in the design space. (Indeed, such OFAT methodology no doubt lead to the original specification of nominal set-point, with its lift penalty of 1000+ counts.) Without the factorial element of the experiment imposed as part of a special MDOE evaluation in this experiment, the effect of ignoring interactions could well have gone undiscovered. There are no doubt many such interaction effects in wind tunnel testing today that will continue to remain undiscovered as long as factorial methods are ignored in favor of more familiar OFAT techniques.

Multiple Factors

Two-level factorials have advantages that make them well suited for experiments involving more than the two factors considered in the lift optimization experiment. We now consider a more involved two-level design that involved a total of six factors, which is by no means an unusually complex example of configuration testing.

A landing stability test was recently conducted at Langley Research Center in which a generic winged body used in spacecraft landing studies was tested with a number of configurations. This study was designed as a two-level factorial with six variables, which are described in Table III. A table of signs could be easily constructed for this experiment and each of the main effects and interactions computed using the algorithm introduced above. In practice, such experiments are usually analyzed with software designed explicitly for this purpose. Responses were recorded for the six principal body axis and stability axis forces and moments when the model was in a high angle of attack approach attitude with zero sideslip, and a Mach number appropriate for landing. Similar analyses were conducted for all responses but we will use roll moment as an illustration.

Table III. Variables in landing stability study.

Factor	Symbol	Low Level	High Level
Left Elevon Deflection	A	-20^{o}	0^{o}
Right Elevon Deflection	B	-20^{o}	0^{o}
Body Flap	C	-10^{o}	0^{o}
Canards	D	Off	On
Landing Gear	E	Up	Down
Speed Brake	F	Not Deployed	Deployed

Blocking

A full factorial two-level six-factor design requires $2^6 = 64$ total configurations. Before we discuss the details of such an experiment, note that it may be necessary to distribute this many configuration changes over a fairly extended period of time. If we set the ambitious goal of averaging one configuration change every 15 minutes, we could only make half of the 64 configuration changes in one eight-hour shift, even in the unlikely event that there were no unanticipated delays so that 60 minutes were available for testing out of every hour. In a single-shift operation, this would entail an overnight delay between the two halves of the experiment. In a two-shift operation, it means that different crews would acquire the data in one shift than another. In either case, there is some potential for a bias shift from one subset of the data to the other. In an

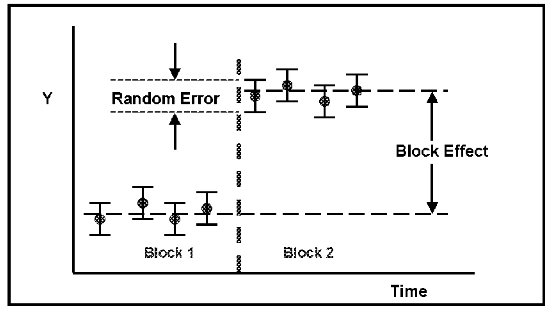

Figure 10. Schematic representation of a block effect.

overnight shutdown, for example, there can be bias shifts in the instrumentation induced by thermal effects and other reasons, or there could be a myriad of subtle shutdown/restart discontinuities which, while possibly small in an individual absolute sense, could nonetheless combine to form a net error that is some substantial portion of the entire error budget in a precision wind tunnel test. We use the term “block” to describe in this case a block of time in which greater commonality may exist among points within the block than between different blocks. A net difference in response across blocks that is too large to attribute to ordinary chance variations in the data is called a “block effect”. Figure 10 is a schematic representation of a block effect.

While block effects do not necessarily occur every time, they can have a serious impact and they can also difficult to detect. The prudent researcher is therefore advised to defend against them. One of the great advantages of two-level factorial designs is that they can provide a convenient and effective defense against block effects, simply by assigning points blocks in a particular way, as we will now describe.

To illustrate the impact of block effects on our ability to properly estimate the main and interaction effects in a two-level factorial experiment, consider the data in Table IV. These points are a subset of the full 64 configurations that were set in the GWB landing stability test. They represent all the points for which elevon and body flap deflections were all zero. That is, Table IV represents a two level factorial experiment in only three factors: speed brake, canards, and landing gear. We use such a subset simply to save space and simplify the example, but the results we will present apply no matter how many factors there are.

Imagine that for some reason it is necessary to subdivide the eight configurations represented in Table IV into two blocks. Perhaps there was some unforeseen delay that caused only the first four configurations to be set before close of business on one day, for example, so that the last four points were acquired after an intervening overnight shutdown. Assume further that some unknown source of error has biased all of the estimates of roll moment made in the first block so that they are each 0.0005 too high, and that in the second block some equally unrecognized error source has biased the roll moment estimates too low by, say, 0.0003. If the acceptable uncertainty in roll moment is represented by a standard deviation of 0.0001, say, then these errors are quite significant. In Table IV-a we represent the data acquired in what is called "standard order", which most clearly represents the layout of the design. Columns of data feature roll moment with and without the hypothetical bias errors described above. Table IV-b is identical to Table IV-a except for the order of the points. The same points are acquired, but a different subset is acquired on the first day than on the second day.

Table IV. Block Effects.

Day	D	E	F	CMXS		
				Unbiased	**Bias Error**	**Biased**
1	-1	-1	-1	0.00482	+0.0005	0.00532
1	-1	-1	1	0.00485	+0.0005	0.00535
1	-1	1	-1	0.00371	+0.0005	0.00421
1	-1	1	1	0.00498	+0.0005	0.00548
2	1	-1	-1	0.00413	-0.0003	0.00383
2	1	-1	1	0.00487	-0.0003	0.00457
2	1	1	-1	0.00439	-0.0003	0.00409
2	1	1	1	0.00468	-0.0003	0.00438

a) Data acquired in standard order.

Day	D	E	F	CMXS		
				Unbiased	**Bias Error**	**Biased**
1	-1	-1	-1	0.00482	+0.0005	0.00548
1	-1	1	1	0.00498	+0.0005	0.00537
1	1	-1	1	0.00487	+0.0005	0.00489
1	1	1	-1	0.00439	+0.0005	0.00475
2	-1	-1	1	0.00485	-0.0001	0.00361
2	-1	1	-1	0.00371	-0.0001	0.00403
2	1	-1	-1	0.00413	-0.0001	0.00458
2	1	1	1	0.00468	-0.0001	0.00000

b) Data orthogonally blocked.

Let us now estimate the main effect for factor D, using the table of signs algorithm introduced earlier. The D effect for roll moment is the change in roll moment caused by changing from a configuration without canards to one with canards. To estimate this effect, we multiply the ±1 values in the D column times the roll moment response values, sum algebraically, and divide by the number of "+" signs in the D column – 4 in this case. Let us first apply the algorithm to the unbiased data in table IV-a. We get D = (-0.00482 - 0.00485 - 0.00371 - 0.00498 + 0.00413 + 0.00487 + 0.00439 + 0.00468)/4 = -0.00007. This value is within the acceptable uncertainty for roll moment in this experiment, represented as a standard deviation of 0.00010, and we would conclude therefore that canards have no significant net effect on roll moment. This conclusion is in harmony with expectations based on symmetry. The same numerical result would be obtained if we used the unbiased data of Table IV-b, of course.

Now apply the table of signs algorithm to compute the D effect to the biased data of Table IV-a. We get D = (-0.00532 -0.00535 -0.00421 -0.00548 +0.00383 +0.00457 +0.00409 +0.00438)/4 = -0.00087. Not surprisingly, the substantial bias errors that occurred on both days have introduced a significant error in the result. In this case, we would have estimated a roll moment of -0.00087, which is over eight standard deviations away from zero. An effect of eight standard deviations is too great to reasonably attribute to simple chance variations in the data, and we would be forced to conclude from the data that the addition of canards induces a significant roll moment, implausible as that might seem from symmetry considerations and other discipline specialist knowledge.

Let us now estimate the canard effect with the same bias errors in play, but with the data arranged in the order indicated in Table IV-b. Again applying the table of signs algorithm, we get D = (-0.00532 - 0.00548 -0.00537 - 0.00489 + 0.00475 + 0.00361 + 0.00403 + 0.00458)/4 = -0.00007. This is precisely the same result we obtained without the bias errors. That is, despite the identical bias errors that produced a roll moment error a factor of eight times the acceptable standard deviation in unexplained variation, and despite the fact that precisely the same combinations of configuration variables were acquired, the re-ordering of the data in Table IV-b produced exactly the same estimate of roll moment we would have achieved had there been no bias errors at all! This is a most remarkable result. It implies, among other things, that not all test matrices are created equal, and that some set-point orderings are apparently preferable to others. In particular, the set-point ordering that is the fastest or the most convenient is not guaranteed to produce the highest quality result. On the contrary, there is generally a tradeoff between speed and convenience on the one hand, and quality on the other. For example, the set-point ordering of Table IV-b requires a somewhat greater number of individual configuration changes than the set-point ordering of Table IV-a. The reward for this extra effort is freedom from the kinds of bias errors featured in this example.

This example illustrates a quality assurance tactic known as orthogonal blocking, which is especially convenient to implement with two-level factorial designs. If we imagine that the columns in the design matrix represent vectors, then the vectors of a two-level factorial are all mutually orthogonal. That is, the sum of term-by-term cross-products of any pair of columns is zero. This is proportional to the cosine of the angle between the vectors which implies that all the vectors are at right angles, or *orthogonal* to each other. If a block effect confounds one of the vectors, it will have no effect on any of the others for that reason. The technique used in this example, then, was to cross-multiply the elements of the D, E, and F factors in the table of signs to produce a column of signs for the DEF three-way interaction. Data points were assigned to the two blocks according to whether the signs in this column were positive or negative. Points with positive signs in the DEF column were assigned to one block and those with negative signs in the DEF column were assigned to the other. Bias errors would confound the estimate of the DEF three-way interaction, but because of the orthogonality property all the other columns would be unaffected by the block effect. Thus, by sacrificing a higher-order interaction that was not likely to be very large compared to the main effects or lower-order interactions, all of those potentially more significant effects were protected from block effects. Further details of the quality enhancements that can be achieved by judicious set-point ordering is beyond the scope of this paper, but is discussed in standard references on experiment design.[5-7] Specific applications to aerospace testing are also described in the literature.[1,2] The chief point for the purposes of this paper is that factorial designs such as this one can be structured easily to eliminate what otherwise could be significant components of unexplained variance in the data. This, in turn, relieves the pressure to acquire data in higher volume (i.e., to specify more configuration changes) as a prerequisite for seeking higher precision in a configuration test.

The Sparsity of Effects Principle

Let us return to the full six-factor two-level design, which requires $2^6 = 64$ total configuration changes. We use the symbols in Table III to represent the main effects, and we use combinations of those symbols to represent interactions. For example, "A" and "B" represent the main effects for left and right elevon, respectively, "AB" represents the two-way interaction between left and right elevon, and so forth.

There are N!/[n!(N-n)!] possible n-way interactions involving N factors. In this experiment, then, there are thus 6!/(1!5!) = 6 "one-way interactions" (i.e., main effects, and by symmetry there are also six possible five-way interactions), there are 6!/(2!4!) = 15 two-way interactions and again by symmetry there are 15 candidate four-way interactions, there are 6!/3!3! = 20 possible three-way interactions, and there is 6!/6!0! = 1 six-way interaction. There are thus a total of 6 + 15 + 20 + 15 + 6 + 1 = 63 total possible effects. The 64 data points provide a single degree of freedom for each of the 63 candidate effects, plus one degree of freedom that is consumed in estimating the mean of the data.

Notwithstanding the substantial number of possible candidate effects, in practical situations there are often relatively few effects of significant magnitude. Furthermore, there is a natural hierarchy in which main effects and relatively low-order interactions tend to be more significant than higher-order interactions. We call this general tendency the "sparsity of effects" principle, which we will exploit presently. We can illustrate the sparsity of effects principle for the roll moment analysis of the current six-factor experiment that we are considering. The left and right elevon main effects completely dominate all the remaining 63 possible effects, as is expected from the fact that it is elevon deflection that provides the primary roll command authority for this vehicle. That is, the vehicle was designed explicitly to have significant elevon effects on roll moment. Figure 11 is a bar chart that displays the magnitude of the remaining 61 effects, with elevon main effects deleted from this figure simply to enable greater graphical resolution of the remaining effects. (The main left and right elevon roll moment effects, A and B, are +0.04871 and -0.04495, respectively, which are approximately 50 times larger than the largest interaction effect displayed in figure 11.)

The variance in estimates of the main effects and interactions can be computed using a formula that is easy to derive from a general formula for error propagation provided in standard texts on uncertainty analysis[8,9] and it is also available in standard texts that treat two-level factorial experiments.[6,7] The formula is:

$$\sigma^2_{effect} = \frac{4\sigma^2}{N} \rightarrow \sigma_{effect} = \frac{2\sigma}{\sqrt{N}} \tag{1}$$

where N is the number of points (64 in this case) and σ with no subscript is the standard deviation in the response variable that we are studying. The standard deviation in roll moment was 0.00041 in this experiment so by equation 1 the standard deviation in roll moment effects was 2 x 0.00041/8 = 0.00010. A 57 degree of freedom estimate of the roll moment standard deviation was based on the data, from which a 95% confidence interval half-width was computed as

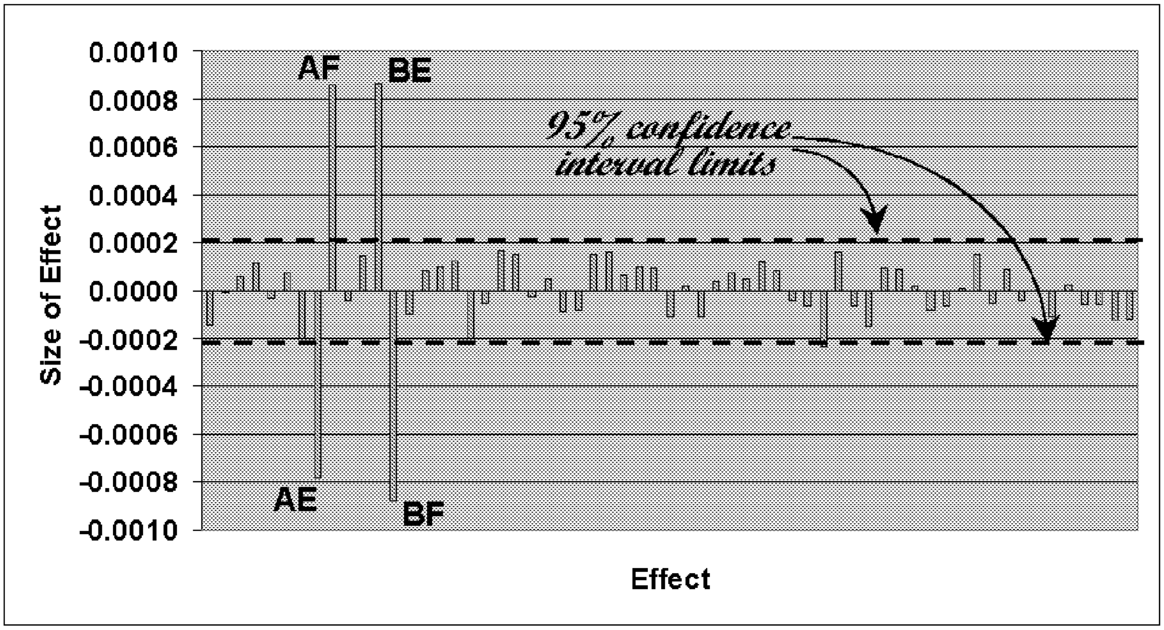

Figure 11. Main and interaction roll moment effects for GWB *full factorial* experiment. Main elevon effects, A and B, excluded to enhance resolution.

$t_{(0.05/2,57)}$ x 0.00010 = 2.0025 x 0.00010 = 0.00021. Note that since each main effect and interaction is a linear combination of the same number of response measurements, each with the same uncertainty, the variance for all effects is the same, regardless of whether they are main effects, low-order interactions, or high-order interactions. All effects therefore share the same 95% confidence interval.

The upper and lower limits of such an interval centered on zero are marked in figure 11. Note that of the 61 main effects and interactions represented on this figure, 57 are within the 95% confidence interval centered on zero. That is, for each of these 57 effects, we have no basis for rejecting a null hypothesis that their true value is zero and that their specific non-zero values are attributable simply to chance variations in the roll moment data. It is only the main elevon effects, A and B, and their two-way interactions with factors E and F that are significant. From Table III we see that factors E and F are landing gear and speed brake, respectively.

A certain degree of clarity has emerged from a potentially large number of complex candidate effects for roll moment. Apparently, out of 63 possible effects that could be in play, the only significant ones are the main elevon effects and the two-way interaction of each elevon with the speed brake and the landing gear. That is a total of only six effects that are not in the noise, or less than 10% of the 63 total possible effects. We can use this information in a number of ways. In the first place, knowledge of the important main effects and interactions, as well as knowledge of which effects play no significant role, enhance our insights into the basic underlying mechanics of the process. While the role of elevons in inducing roll moment is not unexpected, the influence of the landing gear and speed brake was not necessarily anticipated. The fact that these were important while the body flap and canards made no difference provides subject matter specialists with interesting insights into the general stability and control problem.

It is especially useful to know which factors are important and which are not if we decide to employ a steepest ascent strategy to seek independent variable combinations that maximize or minimize roll moment. In this case we can base the direction cosine computations and local curvature estimates on $2^4 = 16$ corner-points rather than $2^6 = 64$, a savings of 48 configuration settings at each stage of this process. We not only know that we can drop 75% of the configuration settings, we know which 75% to drop.

Fractionating the six-factor design

Recall that we used the column of signs for the highest-order interaction as a basis for blocking the

Run Number	Variable						ABCDEF	Block
	A	B	C	D	E	F		
1	-	-	-	-	-	-	+	I
2	-	-	-	-	-	+	-	II
3	-	-	-	-	+	-	-	II
4	-	-	-	-	+	+	+	I
5	-	-	-	+	-	-	-	II
6	-	-	-	+	-	+	+	I
7	-	-	-	+	+	-	+	I
8	-	-	-	+	+	+	-	II
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
63	+	+	+	+	+	-	-	II
64	+	+	+	+	+	+	+	I

Figure 12. Using the 6-way interaction to blocking a six-factor 2-level design.

experiment in such a way that block effects confounded the high-order interaction (which is not usually expected to be significant by the sparsity of effects principle), and in this way defended all main effects and all other interactions from block effects. If we were to block this six-factor experiment into two blocks in this way, we would sacrifice the ABCDEF six-way interaction in order to ensure that all other effects are clear of block effects. Figure 12 illustrates how we would use the ABCDEF interaction to assign points to blocks in this way, where the signs alone are used as implied ±1 values to conserve space and in keeping with another common convention for representing high and low levels in a two-level factorial design. Since only six degrees of freedom are consumed in estimating the six significant roll moment effects and one more is needed for the mean, it is reasonable to ask why the 32 data points in the first block would not be ample for our purposes. Why press on to acquire 32 additional data points beyond the 32 we already have acquired in the first block, if the sparsity of effects principle suggests that with 32 degrees of freedom already acquired we probably have more than enough data to quantify the number of effects that are likely to be significant? One reason might be that additional degrees of freedom are needed via equation 1 to satisfy inference error risk tolerance specifications. Another reason might be that we simply fear there are more significant effects in this specific situation than the sparsity of effects principle would suggest for the general case. However, in many commonly occurring circumstances we can do with a significantly smaller volume of data (which means significantly fewer configuration settings in a configuration study) than is necessary to estimate every single main effect and every single higher-order interaction that is theoretically possible.

Note also that the number of configuration settings required to estimate all theoretically possible effects doubles with the addition of each new configuration variable. The Blended Wing Body is a proposed transport concept with 16 trailing edge control surfaces, for example. With power on/off as a 17^{th} variable, over 131 *thousand* configuration changes would be needed for a full-factorial experiment in these 17 variables, even if each was set at only two levels. However, it is highly unlikely that any of the 17!/(8!9!)=24,310 eight-way interactions will be large compared to the main effects and lower-order interactions, for example, and likewise for the 23,310 nine-way interactions, the 19,448 ten-way interactions, the 12,376 11-way interactions, and so on. Experience suggests that significant interactions much beyond about 3^{rd}-order are unlikely, and that not even all effects of order three or less will be significant. So in large problems such as this a full factorial design seems especially wasteful. This leads to the concept of "fractionating" the design to more closely match the number of configurations examined with the number of effects that are likely to be real. We will use the six-variable GWB landing stability test to illustrate an efficient approach to fractionating the full factorial design, and to illustrate the penalty associated with the substantial cost savings that accrue from eliminating configuration settings.

Imagine that we have blocked this design as indicated in figure 12, by using the six-way ABCDEF interaction column to assign configurations to blocks. Assume further that we have executed the first block only. Such a design is called a *half-fraction* of the full factorial design, for obvious reasons. Figure 13 represents the table of signs for this half-fraction design. This table of signs has almost exactly the same structure as the full factorial design. The only difference is that the ABCDEF six-way interaction column contains all values of the same sign (the sign used to define this block as in figure 12), and there are only 32 rows instead of 64, reflecting the factor of two cost savings we are attempting to achieve. We could proceed to use the table of signs algorithm with this table to estimate all the effects exactly as before (except the six-way interaction effect that we have sacrificed to achieve this cost savings), and we would have set only 32 configurations, a great savings of time and money.

The wary reader may sense the onset of a free lunch proposition about which he is entitled to be skeptical. Is it really reasonable to assume that 32 degrees of freedom are sufficient to unambiguously estimate the mean plus any of up to 63 additional effects? No, of course it is not. We "pay for our lunch" through a phenomenon known as "aliasing", which can be illustrated by comparing the column of signs for the

F factor main effect in figure 13 with the column of signs for the ABCDE five-way interaction, for example. Both columns of signs are identical. This means that when we use the column of signs algorithm to estimate the main effect for factor F, we actually compute the sum of this effect plus the five-way interaction effect. If we adopt a common convention by using square brackets to represent an aliased estimate of an effect and no brackets to represent the true effect, then we can write [F]=F+ABCDE to indicate the aliasing. The term "alias" derives from the fact that the same column of signs now caries two labels or names – F and ABCDE.

Note that the F effect is not the only aliased main effect. The column of signs for the E main effect is identical to the column for the ABCDF interaction in figure 13 so we also have [E]=E+ABCDF, and likewise [D]=D+ABCEF. It turns out that all main effects and every interaction effect is aliased in this way. (It is just as proper to say that the ABCDE five-way interaction is aliased by the F main effect as to say the main effect is aliased by the interaction, so we can also express this as [ABCDE]=ABCDE+F.) That is, every estimated effect is actually the sum of that effect plus one additional effect. A little reflection reveals why this must be. There are a total of 63 possible effects plus the mean, which must now be estimate with 32 degrees of freedom. Each degree of freedom must therefore do double duty, representing two effects instead of one.

At first glance, aliasing may seem to represent a hopeless confounding of effects. If every estimated effect is really the sum of two different effects, how can we determine how much is contributed by one effect and how much by the other? Without performing the entire full factorial experiment, we cannot tell definitively how much each of the two aliased effects contributes to the estimate computed from their common column of signs. However, we can invoke the sparsity of effects principle again, to surmise that the higher-order interactions are likely to be either non-existent or relatively small compared to the main effects and lower-order interactions. So in cases such as we have illustrated here, where main effects are aliased by five-way interaction effects, there is reason to believe that the aliased estimate of the main effect may not be much different from the true main effect. That is, we have some reason to believe that the enormous savings we have achieved in time and money may have been secured with a relatively small price in the quality of our result. If we can arrange it so that the effects that are likely to be important are aliased by effects that are likely to be insignificant, then we may be able to enhance productivity substantially.

The full-factorial roll-moment data set for the GWB experiment was reanalyzed, using only the half-fraction component of figure 13. The main left and right elevon effects for roll moment, A and B, were aliased in this case by the BCDEF and ACDEF five-way interaction effects, respectively. Nonetheless, at +0.04858 and -0.04501 they were still on the order of 50 times larger than the next largest effect in the half-fraction design, just as in the full factorial case, and their significance as important roll moment drivers is therefore still unambiguous despite the aliasing. Figure 14 displays all of the other (aliased) effects, with the

Run	Main Effects							Two-factor Interactions					Five-factor Interactions			Six-Factor Interaction
Number	A	B	C	D	E	F	...	BF	CF	DF	EF	...	ABCDE	ABCDF	ABCEF	ABCDEF
1	-	-	-	-	-	-	...	+	+	+	+	...	-	-	-	+
4	-	-	-	-	+	+	...	-	-	-	+	...	+	+	-	+
6	-	-	-	+	-	+	...	-	-	+	-	...	+	-	+	+
7	-	-	-	+	+	-	...	+	+	-	-	...	-	+	+	+
10	-	-	+	-	-	+	...	-	+	-	-	...	+	-	-	+
11	-	-	+	-	+	-	...	+	-	+	-	...	-	+	-	+
13	-	-	+	+	-	-	...	+	-	-	+	...	-	-	+	+
16	-	-	+	+	+	+	...	-	+	+	+	...	+	+	+	+
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
64	+	+	+	+	+	+	...	+	+	+	+	...	+	+	+	+

Figure 13. Table of signs for one block representing half of a six-factor 2-level design.

(aliased) A and B effects removed from the plot as before to enhance graphical resolution.

We know what the benefit is of conducting the six-factor two-level configuration experiment as a half fraction: We save 32 potentially arduous and time-consuming configuration changes. Even under the ambitious assumption that we could make such changes every 15 minutes, this would have reduced cycle time by a full eight-hour day. A full cost accounting evaluation of the impact of a day's cycle time reduction transcends the scope of this paper and the expertise of the authors. We note, however, that it is not uncommon for a major aircraft manufacturer to have several billions of dollars committed to bringing a new aircraft design to market. At historical market rates, the cost of capital (interest) on this much money can approach the million-dollars-per-day level. There is also the weighted expectation value cost associated with the risk that a competitor will bring some similar design to market a day earlier. Clearly, saving a day of cycle time now and again on a major program can translate into substantial benefits on a full cost accounting basis. This is likely to be true even on smaller projects.

The cycle-time savings afforded by fractionating two-level factorial designs do not come without costs, as the above discussion of aliasing indicates. We can compare figure 14 to figure 13 to assess the impact of such aliasing for the six-factor GWB landing stability test we have been examining. This comparison is quite revealing. Again, as in figure 13, horizontal dashed lines indicate the upper and lower limits of a 95% confidence interval centered on 0, which is computed under the assumption that chance variations in the data are the only sources of error. This represents the minimum uncertainty that could be achieved with a 32-point data sample given the variance in individual roll moment measurements, per half-width computations based on the standard error formula in equation 1. The 95% confidence interval half-width for the fractionated design is 0.00032, versus 0.00021 for the full factorial design, reflecting the smaller sample through equation 1 plus the slightly enlarged t-statistic associated with the difference between 57 degrees of freedom to estimate the unexplained variance for the full factorial case and only 25 degrees of freedom that would have been available to assess the unexplained variance in the half-fraction.

The essence of figure 14, which represents the half-fraction case, is that it virtually indistinguishable in its important details from the full-factorial case of figure 13. The magnitudes of four interaction effects are again substantially larger than all other effects, which are seen to be statistically indistinguishable from zero just as in the full-factorial case. The same (albeit aliased) interaction effects are clearly distinguishable from the noise for the half-fraction case as for the full factorial case. That is, the half-fraction inferences are identical to those drawn from the full factorial case, notwithstanding the fact that they are based on half the data.

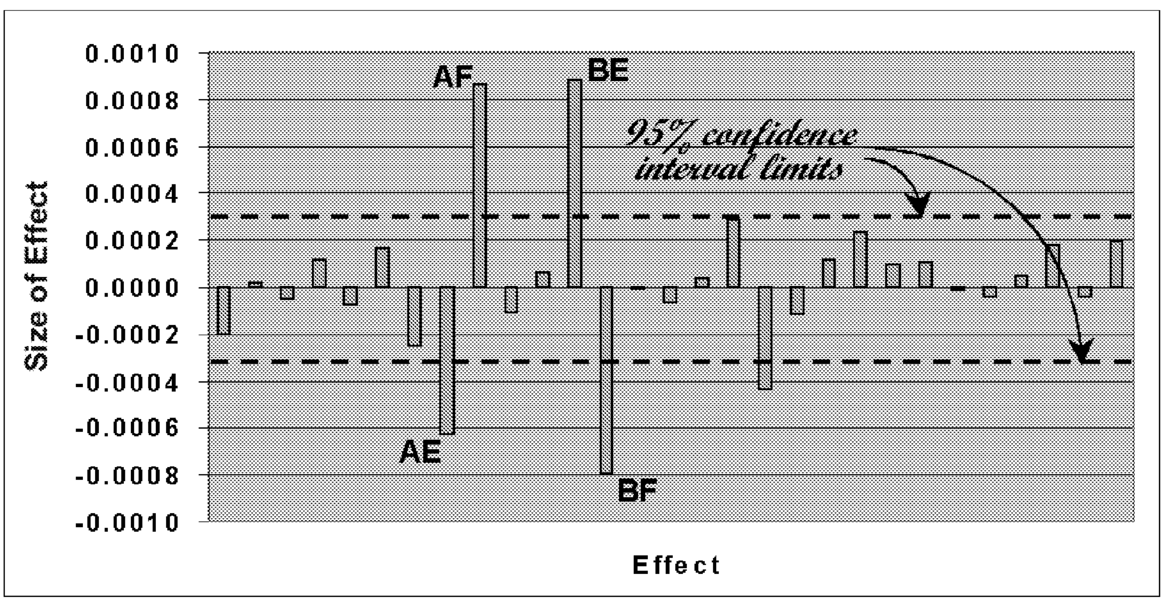

Figure 14. Main and interaction roll moment effects for GWB *half-fraction* experiment. Main elevon effects, A and B, excluded to enhance resolution.

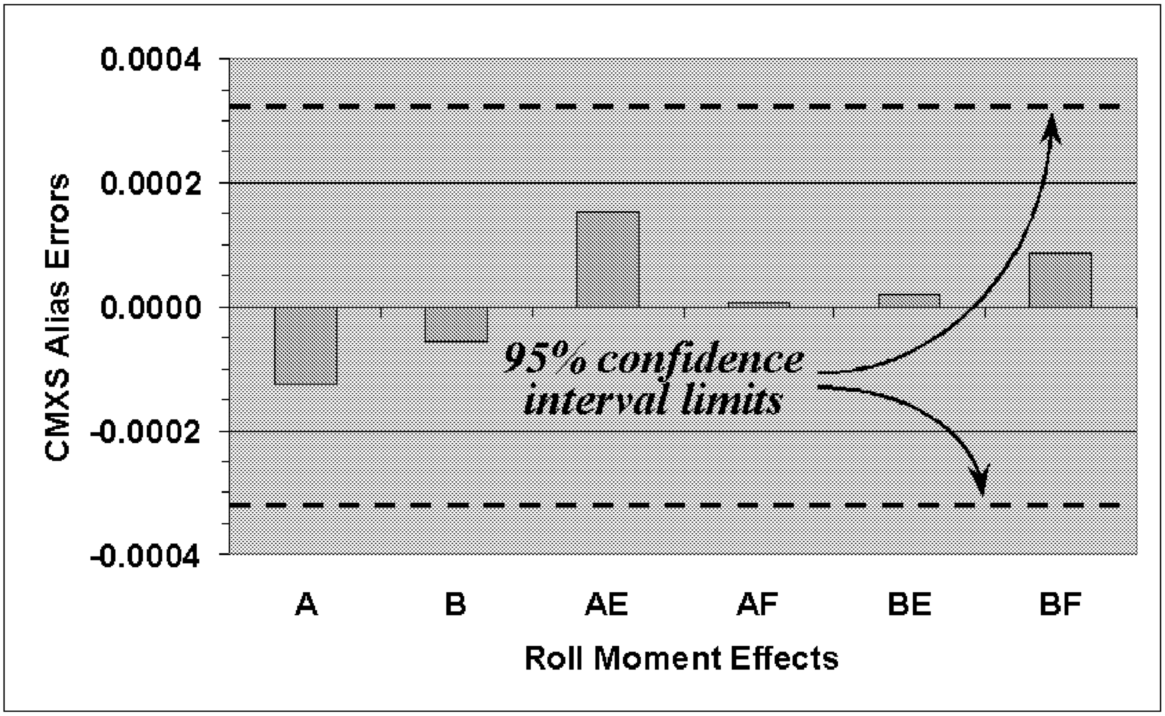

Figure 15. Half-fraction alias errors for roll moment.

We have already noted that the main effects are aliased with five-way interactions in the six-factor half-fraction design. All the two-way interactions are aliased with four-way interactions, as it happens. Specifically, for the four significant two-way interactions we have that [AE] = AE + BCDF, [AF] = AF + BCDE, [BE] = BE + ACDF, and [BF] = BF + ACDE. The sparsity of effects principle suggests that the four-way interactions will be small and that the two-way interaction estimates will not be substantially different with or without aliasing by four-way interactions. It is always a good idea to examine the specific aliasing effects to see if there is any special reason to believe that they constitute an exception to the general sparsity of effects principle, but if there are no such special cases, we have the option of accepting the extra uncertainty associated with the aliased estimates of the significant effects, as the price we pay for a substantial savings in cycle time.

In this pedagogical example, we have the luxury of examining both the aliased and unaliased estimates of the significant main effects and two-way interactions. Figure 15 shows the difference for each of the two significant main effects and four significant two-way interactions. These differences represent an additional component of uncertainty introduced by the aliasing with higher-order interactions. (In this case they represent the magnitude of the high-order aliasing effects themselves.) The 95% confidence interval associated with errors in the 32-point effects estimates induced by ordinary chance variations are indicated in the figure. It is clear that the errors due to aliasing are no greater than ordinary random experimental error in this case, and that the cost/benefit tradeoff associated with fractionating the design would have been a favorable one. (In the actual experiment, the full factorial case was executed as part of an in-house demonstration/evaluation of the benefits of fractional factorial designs, but a fractional factorial would almost certainly have been executed in the case of a six-factor experiment, given the likelihood that only low order effects would be significant and that they would be aliased with higher-order effects.)

Fortunately, it is easy to forecast the aliasing patterns in advance, through a simple mechanism known as the "defining relation" that will now be described. We begin by defining a column containing all plus signs as an identity column, denoted by "I". For example, the ABCDEF column in figure 13 is an identity column. We can therefore write in this case that I=ABCDEF. This equality is the "defining relation" for this half-fraction design.

We introduce a special algebra for the columns of signs in which multiplication of two columns is meant to imply the term by term multiplication of corresponding elements in each vector, in exactly the same way that columns for the interaction terms were formed from the main effects columns in the table of signs. The columns algebra features only two rules. The first rule is that since the identity column is a column of plus signs only, multiplying "I" times any other column results in that same column. The second rule is that multiplying any column times itself produces the identity column. This is because +1 time +1 is +1, and -1 times -1 is also +1. Armed with these two rules and the defining relation, we can now determine the aliasing patterns for any main effect or interaction. For example, to determine the aliasing pattern for the "A" main effect, we simply multiply both sides of the defining relation by A and invoke our two rules involving the identity column:

I = ABCDEF
A x I = A x ABCDEF = (A x A) x BCDEF
A = I x BCDEF
A = BCDEF

This tells us that A and BCDEF are aliases of each other.

The same procedure can be sed for any other effect. For example, B x I = B = B x ABCDEF = A x (BxB) x CDEF = AICDEF = ACDEF. So B and ACDEF are aliases. The general algorithm when multiplying a word in the defining relation by some other word is to drop all letters common to both words and retain only those that are unique to one word or the other, then to drop the "I" factors.

We can see from the defining relation that the AE interaction we saw was significant is aliased by the BCDF interaction, the AF interaction is aliased by the BCDE interaction and so forth. There were no significant three-way interactions for roll moment but we could easily have determined their aliasing patterns had any been significant. Multiplying I=ABCDEF through by ABC, for example, yields ABC = DEF. These two three-way interactions are therefore aliases of each other and share the same pattern of signs in their respective effects columns in the half-fraction design.

A pattern has emerged in which we see that main effects are aliased with five-way interactions, two-way interactions are aliased with four-way interactions, and three-way interactions are aliased with other three-way interactions. That is, *n-way* interactions are aliased with *m-way* interactions, where in this case $n+m=6$. We use the term "resolution" to describe this situation and say in this specific case that a half-fraction of a six-factor two-level factorial is of "resolution VI", where it is a convention to use Roman numerals to indicate the resolution. Note that the resolution of the design equals the number of characters in the longest word of the defining relation.

If we can neglect 3rd-order interaction terms and higher, we can be confident that the aliasing that occurs in a fractional factorial design of resolution V or higher will not be a serious problem. This is often a safe bet in practical circumstances, but since the aliasing patterns are known in advance, we can also assign the physical variables to factor letters in the design in such a way as to exploit our subject matter expertise to cluster unrelated effects into the interaction terms that alias the effects that most interest us. A certain amount of judgment is required to do this effectively, not to say some plain good luck. Nonetheless, it is possible to "stack the deck" in the researcher's favor by the judicious exploitation of aliasing patterns in advance.

Quarter-fraction designs

Having achieved such a good cost/benefit trade through a half fraction of the original full factorial design, it is natural to ask if we could divide the half-fraction in half again, achieving an even greater reduction in the number of configurations we would need to set. After all, even a half fraction requires 32 configuration settings, which may still seem excessive since in our example we know that only six effects are important. In a real application we would not know in advance exactly how many significant effects to expect, of course, but the sparsity of effects principle gives us the same general answer in all cases: "Not very many." On that basis we may be tempted to explore further fractionation.

Recall that we began the development of the half-fraction design by using the column of signs for the highest-order interaction term in the full factorial to break the design into two half fractions, according to which points had plus signs in that column and which had minus signs. We call the ABCDEF interaction term used for this purpose a "design generator".

A single design generator was used to create the half fraction design. Two design generators can be used to create a quarter fraction. We select two columns of signs from the full factorial design and divide the points (the rows in the table of signs) into four 16-point fractions according to whether both signs in the two defining generators are positive for a given point, both are negative, the first is negative and the second is positive, or the first is positive and the second is negative. We can use whichever of the four quarter fractions is most convenient. In this example we will consider what is called the "principal fraction", which is formed from the points with plus signs in the same location of the two design generators. For this case we used as design generators the ABCE and BCDF interaction terms from the full table of signs. The criterion for selecting design generators is that we wish to develop the highest-resolution design possible to minimize aliasing effects. In reality, this is a solved problem for designs featuring a wide range of numbers of factors and for various levels of resolution. Computer software exists for designing fractional factorials in which the number of factors and desired number of points are specified (six and 16, respectively, for the quarter-fraction design we are currently considering) and the design generators that produce the highest resolution design are then selected by the computer via a table lookup scheme involving tabulated results of years of trial and error design attempts. The reader can also consult the literature for these tables of optimum design generators.[6,7,10,11]

Because we use the principal fraction, both of the design generators for our quarter fraction design will be identity columns and we will have I = ABCE = BCDF. There is an interaction effect associated with any pair of effects that can be determined by column multiplication of the kind we have been using all along. In this case, the interaction of the ABCE and BCDF effects is ABCE x BCDF = A (BB) (CC) DEF = AIIDEF = ADEF. Since both ABCE and BCDF are identity columns their product is also, and we have the following defining relation for this quarter fraction design:

$$I = ABCE = BCDF = ADEF \tag{2}$$

We see immediately that this will be a resolution IV design, because the longest word in the defining relation has only four characters. That means that if we can neglect interactions of order three and higher, we can still expect to have reasonably clear estimates of the main effects of all six factors, since main effects are only aliased with interactions of order three or higher in a resolution IV design. The two-way interactions will be a bit more problematic, however, because two-way interactions are aliased with other two-way interactions in a resolution IV design, and one would seldom be justified in assuming no significant two-way interactions. Still, it is possible to gain important insights in a relatively short period of time with highly fractionated (i.e., relatively low resolution) designs such as this, as we can demonstrate by further fractionating our roll moment experiment design.

We are in the fortunate position of having already analyzed a full factorial version of this experiment, so we know which factors are significant without any ambiguities introduced by aliasing. They are the A and B main effects for left and right elevon deflection, and the two-way interactions of each of these main effects with variables E and F, the landing gear and speed brake, respectively. We can use the defining relation for our quarter-fraction design to determine the aliasing pattern for these significant effects. They are as follows:

$$\begin{aligned}
[A] &= A + BCE + ABCDF + DEF \\
[B] &= B + ACE + CDF + ABDEF \\
[AE] &= AE + BC + ABCDEF + DF \\
[BE] &= BE + AC + CDEF + ABDF \\
[AF] &= AF + BCEF + ABCD + DE \\
[BF] &= BF + ACEF + CD + ABDE
\end{aligned}$$

Note that each estimated effect is actually the sum of four effects, consistent with the fact that we have 16 degrees of freedom available to represent a total of 64 potential effects (63 plus the mean). Each available degree of freedom must therefore carry four effects.

The main left and right elevon effects are as unambiguously part of the solution in this quarter-fraction design as they were in the full factorial and half fraction. Again, the three-way and five-way interactions are not likely to be large compared to the main roll moment effects, especially when the main effects are elevon deflections.

Figure 16 presents the quarter fraction effects except for A and B, again deleted to increase graphical resolution for the remaining effects. The aliased versions of the four effects we know to be significant again protrude above the noise, although less so because the 95% confidence interval is now wider, reflecting the additional factor of two reduction in sample size and the aliasing effects. Interactions higher than third order have been deleted from the labels of the effects in figure 16.

We already know which effects are real and which are noise from the full factorial results, but we must put ourselves in the position of an experimenter who has conducted only the quarter-fraction result. What can be said about these results without the benefit of the extra knowledge we possess in this special case? One plausible line of reasoning might develop along lines similar to the following: We start with the two significant effects the furthest to the right in figure 16. These are:

AF + DE
BF + CD

We already know that the main A and B effects dominate the roll moment. We know this because of the relatively unencumbered estimate that a resolution IV design provides for main effects, which indicate that they are about 50 times larger than the next largest significant effect, and also from our own subject-matter expertise, which informs us that the elevons provide the main roll control authority for the vehicle. Since the A and B effects are so important for roll moment, it is more likely that a true two-way interaction would include A or B than any other factor. Thus, AF is more likely than DE to be a real effect, and likewise BF is more likely to be real than CD. Taken together, if we tentatively accept that either AF or BF is real, then it is more likely by simple symmetry that the other is also real. That is, if the *left* elevon interacts with the speed brake (factor F), then it seems likely that the *right* elevon would also interact with it. If instead we assume that DE dominates AF and CD dominates BF, we are left with a model in which the canards (factor D) interact with the body flap (factor C) and the landing gear (factor E), and that these interactions are greater for roll moment than the AF and BF interactions that involve the main roll moment control surfaces. While this is not impossible, it seems rather unlikely. It is reasonable to assume that the analyst would correctly infer that the AF and BF interactions were significant, despite the relatively substantial aliasing in the quarter fraction design.

The two significant effects on the left side of figure 16 are somewhat more ambiguous. They are as follows:

BE + AC
AE + BC + DF

We could perhaps eliminate the DF interaction from contention on the grounds that it does not involve either of the elevons and all four of the remaining candidate interactions do. However, there is no rational basis for favoring a model in which both elevons interact with the landing gear (factor E) over a model in which both elevons interact with the body flap (factor C). So the real portions of the above two aliased pairs of effects could just as easily be AE and BE, or AC and BC.

Here we see that aliasing has introduced an ambiguity that is irresolvable without some additional information. We are presented with two outcomes that are equally plausible – that the dominant elevon effects

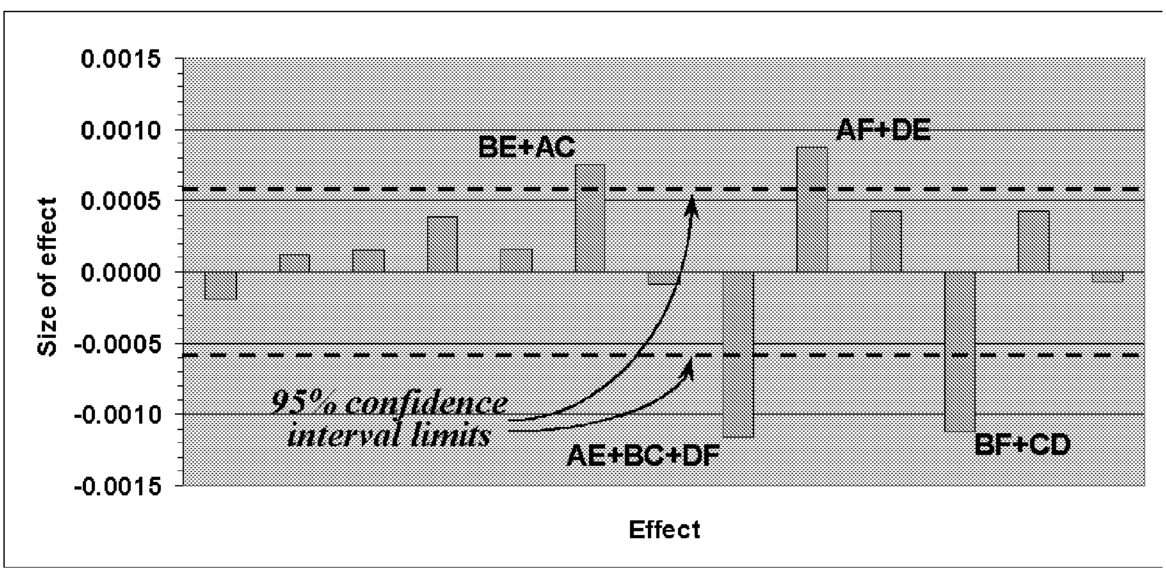

Figure 16. Main and interaction roll moment effects for GWB *quarter-fraction* experiment. Main elevon effects, A and B, excluded to enhance resolution.

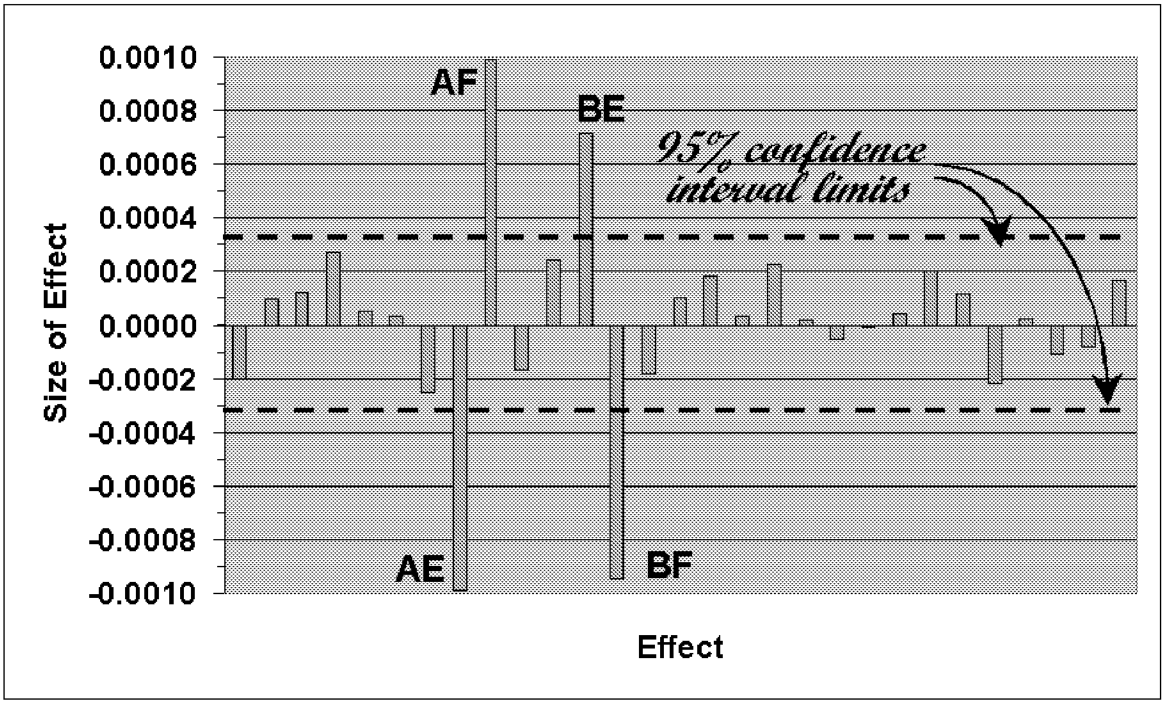

Figure 17. Main and interaction roll moment effects for GWB experiment *after factor-C augmentation*. Main elevon effects, A and B, excluded to enhance resolution.

can be influenced either by the influence of the landing gear deployment or by the deflection of the body flap, a trailing-edge control surface mounted on the axial centerline of the vehicle aft of the a propulsion pod and halfway between the left and right elevons. This suggests that eliminating 48 of the 64 candidate configurations may have been too ambitious in this case, and that a resolution VI half fraction – which nonetheless eliminates half of the configurations without having to drop any of the variables from the design – might be the best selection. However, assuming that we have just executed this quarter-fraction design in six variables, we would like to be able to augment it methodically so that we can proceed to a half fraction without having to throw away any of the results obtained so far. Fractional factorial designs permit us to do exactly tha, whereas if a small scale OFAT design obtains insufficient information, the researcher is quite frequently in the position of starting substantively from scratch.

There are numerous augmentation strategies available for fractional factorial designs but many of them involve techniques for adding fractions judiciously. For example, for a 16-point quarter fraction of a six-variable two-level factorial, these strategies would typically involve adding an addition 16-point fraction. One way of doing this is called a "foldover" augmentation that in this case would add 16 more points to the design with all the plus signs and minus signs reversed in the table of signs. This strategy has the effect of clearing main effects of certain low-order interaction effects, providing a higher resolution estimate of all the main effects. For example, the foldover of a resolution III design, in which all main effects are confounded by two-way interactions that cannot be neglected out of hand, results in a resolution IV design with all main effects clear of interactions less than 3rd-order.

Another strategy is to create a second fraction identical to the first, except that all signs in the column for one main effect are reversed. This clears the main effect of that variable and all of its two-way interactions of effects less than 3rd-order. In our case, neglecting 3rd-order interactions or higher, we have two aliased terms for which either alias is plausible. We have an effect that is equal to BE + AC and another effect that is equal to AE + BC (+ a third candidate interaction effect for roll moment that does not involve the elevons, which we are temporarily discounting). We cannot say whether the elevon effects interact with the E factor (landing gear) or the C factor (body flap). We decide to relieve this ambiguity by reversing the signs in the column for the C main effect in an otherwise identical table of signs for the quarter fraction. This will free the main effect of the body flap from low-order interactions and likewise its two-way interaction with all other main effects and two-way interactions.

Table V. Augmentation of quarter-fraction design

Quarter-Fraction Significant Terms	
A + BCE + DEF + ABCDF	0.04845
B + ACE + ABDEF + CDF	-0.04497
BE + AC + CDEF + ABDF	**0.00075**
AE + BC + DF + ABCDEF	**-0.00115**
AF + BCEF + DE + ABCD	0.00088
BF + ACEF + ABDE + CD	-0.00112
Significant Terms of Augmented Design	
A + DEF	0.04866
B + ABDEF	-0.04501
BE + ABDF	**0.00071**
AE + DF	**-0.00099**
AF + DE	0.00099
BF + ABDE	-0.00094
Body-Flap Terms After Augmentation	
C	-.000020
AC	**0.00003**
BC	**-0.00017**
CD	-0.00018
CE	0.00010
CF	1.85E-04

Table V compares the significant main effects from the quarter fraction design with the corresponding effects of from the augmentation. Note that the ambiguities involving the C factor are resolved. In particular, all of the interaction terms involving the C factor are now eliminated from the aliasing pattern of the augmented design. Furthermore, unaliased estimates of C and its two-way interactions with each of the other five factors are now clear of interaction terms that were present in the quarter-fraction design. This permits us to see how small they are. Recall that no roll-moment effect from a 32-point sample with the experimental error of this investigation can be resolved at 95% confidence unless its magnitude is at least 0.00032. The AC and BC effects that were causing so much confusion in the quarter-fraction design are now seen to be +0.00003 and –0.00017, respectively – well below the noise floor. So we know now that these terms were not responsible for the effects being significant, and that it is now much more clear that it is the landing gear (factor E) that is interacting with the elevons, not the body flap. Figure 17 displays the effects estimated after the factor-C augmentation to the quarter-fraction design. It is similar to figures 11 and 14, showing effects for the full factorial and original half-fraction, respectively.

The effects we have been estimating represent the change in response (e.g., roll moment) associated with a change in coded variable from −1 to +1. The effects therefore represent twice the change in response due to a unit change in the factor. The change in response due to a *unit* change in a factor represents the coefficient of a response model. We can therefore use the significant factors we have discovered to build a response model that can be used to predict responses at other combinations of the independent variables besides the ones we set in the experiment. We estimate the constant or y-intercept for this model by applying our table of signs algorithm in the usual way, but using the identity column (all plus signs) that forms the left-most column of the design matrix. The algorithm requires us to apply the signs to the response column, compute the algebraic sum, and divide by the number of plus signs, which simply results in the sample mean for the entire data set. This, then, is the y-intercept.

Table VI. Coefficients of roll moment response model from full factorial design (half the effects estimates).

Factor	**Coefficient Estimate**	**95% CI Low**	**95% CI High**
Intercept	0.00267	0.00257	0.00277
A	0.02435	0.02425	0.02446
B	-0.02248	-0.02258	-0.02237
AE	-0.00039	-0.00050	-0.00029
AF	0.00043	0.00033	0.00053
BE	0.00043	0.00033	0.00054
BF	-0.00044	-0.00054	-0.00034

Table VI displays the roll moment response model coefficients for the full factorial design. These are half the effects values we estimated by applying the table of signs algorithm, and the intercept is the average of all 64 roll moment estimates. We can use these coefficients to construct a roll moment model as follows:

$$\begin{aligned} CMXS = {} & 0.00267 + 0.02345A \\ & - 0.02248B - 0.00039AE \\ & + 0.00043AF + 0.00043BE \\ & - 0.00044BF \end{aligned} \tag{3}$$

where from Table III we know that A and B are left and right elevon deflection angles, E is landing gear and F is speed brake. The landing gear and speed brake factors are categorical values, assuming only the values of ±1. The elevon variables are numeric, and can assume a range of values between ±1.

A close examination of equation 3 and the 95% confidence interval limits in Table VI reveals something remarkable. *The coefficients of all four 2-way interact terms are equal within experimental error!* Note also that the magnitudes of the two main elevon effects are also similar, although not indistinguishable within experimental error. If we use the average magnitude of the four two-way interaction terms (0.00042) to represent the magnitude of all four terms, we can simplify equation 3 considerably. Within experimental error, equation 3 reduces to

$$\begin{aligned} CMXS = {} & 0.00267 + 0.00097A \\ & - 0.02248(A - B) \\ & + 0.00042[(A - B)(F - E)] \end{aligned} \tag{4}$$

This has the general form of

$$CMXS = b_0 + b_1 A + (b_2 + b_3 x_2)x_1 \tag{5}$$

where the b_i values are numerical constants, A is left elevon deflection as before, x_1 is a new variable representing the differential elevon deflection – left elevon minus right – and x_2 is another new variable representing F-E, the difference between the categorical variables describing speed brake and landing gear deployment, respectively. The variable x_2 takes on three discrete values: 0 when either the landing gear and the speed brake are both deployed or neither deployed, +2 when the speed brake is deployed and the landing gear is retracted, and –2 when the landing gear is deployed and the speed brake is not deployed. The picture that emerges, then, is that for a given deflection of the left elevon, A, the roll moment is a simple linear function of differential elevon setting with a slope that depends on the speed brake and landing gear deployment. A given change in differential elevon setting has the most effect on roll moment when the speed brake is deployed and the landing gear is retracted, the least effect when the gear is deployed and the brake is not, and an intermediate effect when both are deployed or both are retracted. Gear and brake therefore serve as "gain factors" to alter the sensitivity of the elevon's roll authority on approach. With gear down for approach (E = 1), deploying the speed brake desensitizes roll moment with respect to differential elevon setting, making that approach more stable in that the vehicle will be less sensitive to differential elevon displacements. Retracting the speed brake would augment the roll command authority, if additional maneuverability was required on short notice during approach, for example.

We emphasize that considerable insight into underlying flight mechanisms can be achieved through these simple two-level factorial designs, and if additional detail is needed to describe curvature in the response functions for example, it is easy to augment the two-level design with additional levels as needed. In the meantime, the two-level full and fractional factorials can efficiently separate a myriad of candidate factors into what Professor George Box famously distinguishes as the "vital few" and the "trivial many". This, then, provides a rational basis for further study that concentrates on the most important configuration variables without wasting resources on those that have relatively little influence.

Concluding Remarks

Fractional factorial experiment designs appear to have a number of attractive features that are especially relevant to configuration testing in wind tunnels. They illuminate interactions among multiple configuration variables that conventional OFAT methods can overlook. In so doing, they reveal opportunities to combine variable levels in ways that do not assume that individual factor effects are strictly additive. This can lead to substantial and unanticipated system performance improvements.

Factorial designs also serve as useful building blocks in more complex experiments. They can serve as the basis of steepest-ascent optimization procedures that rapidly identify design-space regions where specific responses occur as extremums. Time-consuming investigations geared to identify detailed response behavior in the vicinity of important maxima or minima can be reserved for design-space regions where such maxima or minima have been efficiently localized using factorial designs.

Factorial designs present the researcher with an attractive alternative to the wholesale elimination of candidate factors that is often otherwise required to conform with inevitable resource constraints when there is a large number of candidate configurations variables in play and insufficient time to set every combination. Retaining all candidate variables has the obvious advantage that more factor effects can be examined, but also the subtle advantage that effects that may not have a large direct influence on system response can nonetheless influence the sensitivity of system response to changes in some other variable. Factorial designs increase the probability of detecting such effects by retaining all candidate factors in the study while simply eliminating combinations of them that contribute relatively little to the understanding of the most important main effects and interactions.

Factorial designs provide a path for efficiently adding additional data when initial data sets are inadequate to unambiguously resolve effects of interest. This saves the resources that might otherwise be wasted if it were necessary to start completely from scratch in such circumstances.

There is a direct relationship between the effects that are quantified in factorial configuration studies and the coefficients of response models that can be used to predict responses for other combinations of variables besides those physically set in the experiment. These response models can often yield surprising insights into the underlying physical mechanisms that govern the response of the system.

Acknowledgments

This work was generously supported by the Langley Wind Tunnel Enterprise. The cooperation of the Langley Aerothermodynamics Branch, the Langley Vehicle Dynamics Branch, Orbital Sciences Corporation, and the Boeing Company in evaluating factorial experiment designs for configuration testing is gratefully acknowledged. The staffs of the National Transonic Facility and the 16-Foot Transonic Wind Tunnel at Langley Research Center, and the ViGYAN Low Speed Tunnel in Hampton, VA, provided invaluable assistance and support of this work.

References

1) DeLoach, R. "Improved Quality in Aerospace Testing Through the Modern Design of Experiments". AIAA 2000-0825. 38th AIAA Aerospace Sciences Meeting and Exhibit. Reno, NV. Jan 2000.

2) DeLoach, R. "Tactical Defenses Against Systematic Variation in Wind Tunnel Testing". AIAA 2002-0879. 40th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV. Jan 2002.

3) DeLoach, R. "Tailoring Wind Tunnel Data Volume Requirements through the Formal Design of Experiments". AIAA 98-2884. 20th AIAA Advanced Measurement and Ground Testing Technology Conference, Albuquerque, NM. June 1998

4) DeLoach, R. Hill, J.S., Tomek, W.G. "Practical Applications of Blocking and Randomization in a Test in the National Transonic Facility" (invited) AIAA 2001-0167. 39th AIAA Aerospace Sciences Meeting and Exhibit. Reno, NV. Jan 2001. (Replace "et al." with names)

5) Cochran, W. G. and Cox, G. M. (1992). *Experimental Designs. 2nd ed.* Wiley Classics Library Edition. New York: Wiley.

6) Box, G. E. P., Hunter, W. G., and Hunter, J. S. (1978). *Statistics for Experimenters. An Introduction to Design, Data Analysis, and Model Building*. New York: Wiley.

7) Montgomery, D. C. (2001). *Design and Analysis of Experiments, 5th ed.* New York: Wiley.

8) Coleman, H. W. and Steele, W. G. (1989). *Experimentation and Uncertainty Analysis for Engineers*. New York: Wiley.

9) Bevington, P.R. and Robinson, D.K. (1992, 2nd Ed). *Data Reduction and Error Analysis for the Physical Sciences.* New York: McGraw-Hill.

10) National Bureau of Standards (1957). *Fractional Factorial Experiment Designs for Factors at Two Levels*. Applied Mathematics Series No. 48. Washington D.C: U.S. Printing Office.

11) McLean, R.A. and Anderson, V. L. (1984). *Applied Factorial and Fractional Designs.* New York: Marcel Dekker