SAFETY, EFFICACY, AND EFFICIENCY:

DESIGN OF EXPERIMENTS IN FLIGHT TEST

Lt Col Aaron A. Tucker, United States Air Force

ABSTRACT

Modern flight test tends to be a complex, expensive undertaking so any increases in efficiency would result in considerable savings. Design of experiments is a statistical methodology which results in highly efficient flight test where only the samples needed are collected and analyzed. Design of experiments uses objective, verifiable statistical analysis to plan, execute, and analyze test. This planning process ensures that the variables of interest are measured to provide a specified confidence in the analysis. Statistical confidence provides critical information on both the magnitude and significance of system performance. Further, system response models are continually generated as test matrices are executed. These models enable the tester to understand and evaluate system performance as the test progresses. This increases test safety and flexibility. A conceptual model is developed and applied to a case study to illustrate how design of experiments assists in safely, effectively, and efficiently planning and executing flight test.

1. Introduction

Design of experiments uses objective, verifiable statistical analyses to plan, execute, and analyze tests. Its planning process ensures that the response variables and their interactions of interest are measured to provide a specific, statistical confidence in the analysis. Further, design of experiments provides a specified statistical power to detect changes from specified treatments.

The specialized equipment, facilities, and skills associated with flight test require a continuous expenditure of resources. A large defense industry can afford to maintain such an infrastructure if it continually produces new aircraft and upgrades to those weapon systems. However, as financial resources tighten in the face of sustained global combat operations, flight test leadership to must have the tools to decide exactly which test resources are to be invested. That decision is quantified in the statistics of confidence and power that is the *raison d'être* of a designed experiment. Design of experiments can be applied to the flight test planning process by providing analyses to a specified confidence and power which enables decisions based on objective, programmatic factors instead of subjective, political considerations. Often, political considerations arbitrarily set by a single stakeholder, dominate technically valid flight test objectives.

1.1. Flight Test Overview

The old practice of testing a new aircraft's performance for the first time in flight killed a lot of great aviators, particularly after World War II in the initial days of jet aircraft.[1] Since then, the mission of the flight test enterprise has evolved to emphasize safe, efficient execution of test

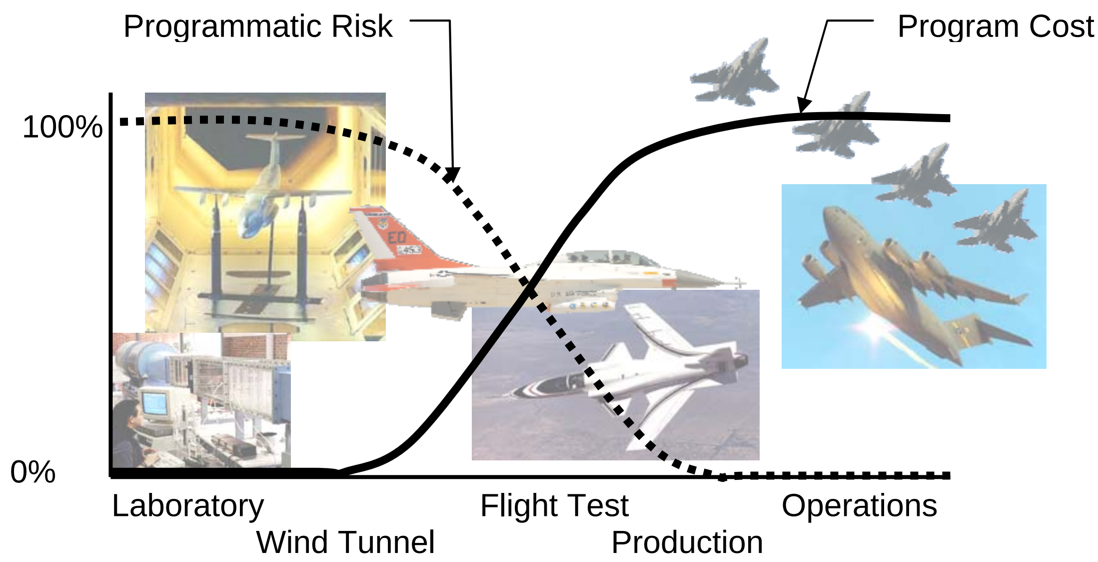

Figure 1. Flight Test at the Cross-over of Cost and Risk

missions to support the needs of warfighter operations. The current model of Defense Acquisition programs delivers capability to the warfighter quickly through continuous system improvements in a spiral development model.[2] The aerospace industry deals extensively with complex systems that are also required to act within a larger system.

The role of the flight test enterprise occurs after the laboratories and wind tunnels as a means of risk reduction before fielding a system to general production and operations (Figure 1). Flight test works in conjunction with modeling in simulation and usually benefits from informed simulations as a means of risk reduction for systems entering flight test. Models and simulations, however, suffer from the inherent limitations of any engineering model based on theory and assumptions. Flight test in operational conditions is necessary to prove a systems' capability before delivery to the warfighting customer.

The flight test enterprise serves not only to reduce the physical risk of crashing airplanes but also the programmatic risk of running over cost, over schedule, and under performance. This risk-reduction role occurs at a critical point in the systems' life cycle as flight test takes place just as program costs increase dramatically into production and operations. Performance risk is controlled since flight test proves a flight vehicle before large scale production and detects problems before they surface in operations after production. Cost and schedule risk are reduced since fixing problems in development is orders of magnitude less expensive than retrofits to systems in production.

Testing over the entire lifecycle of the system, including military utility and operational availability, reliability, and maintainability (sustainment testing) is important. Also, improving the sustainment of the system saves ever-shrinking valuable resources, including the deployment of people.[3] Doing the right job right is the new imperative.

1.2. Stakeholder Viewpoints

The key to success in lean thinking is the proper selection of stakeholders. The idea of the stakeholder was developed as a "group or individual who can affect or is affected by the

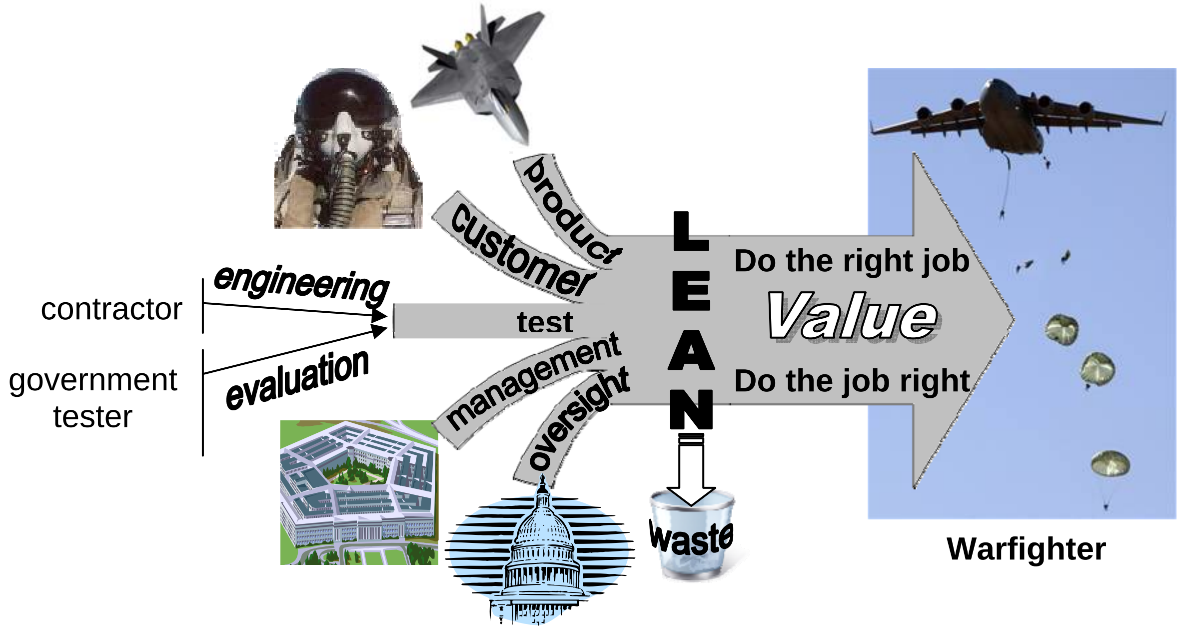

Figure 2. Stakeholders in Flight Test

achievements of the organization's objective."[4] Everything that stakeholders want can be represented by a value stream. Similarly, everything that stakeholders don't want is represented by waste, the anti-value. Various stakeholders must be considered in flight test (Figure 2):

1. Product value (User): the capability to the user over the product's lifetime.[5]
2. Customer value (Production, User): delivery of an aircraft on schedule, within budget, and with full confidence in its safety, performance, and reliability characteristics.[6]
3. Test team value (Contractor, Government Tester): a well-planned and stable program with timely access to adequate support equipment and personnel.[7]
4. Management value (System Program Office (SPO)): completion of a test program with the aircraft meeting all technical specifications within the prescribed time and budget.[8]
5. Oversight agency value (Congress): a testing program with minimal risk that ensures that the aircraft is fully compliant with safety and performance requirements.[9]

Initially excluding some stakeholders can undermine any planning and work completed when they, inevitably, assert their requirements and concerns. Conversely, including the wrong stakeholders in planning will unnecessarily slow program development and add waste to the value stream. The Department of Defense's *Policy Addendum for Systems Engineering* includes independent subject matter expert participation to assess technical maturity and risk mitigation which serves to bring together stakeholders.[10] The customer must understand risks associated with a delivered product—it is impossible to identify unknown unknowns.[11] The flight tester evaluates the product against the requirement set and reports the performance of the aircraft along with the associated risks. Adequate and timely testing with good test plans makes for good products and reduces the risk of major problems in the field

2. Conceptual Model

Design of experiments is the application of statistics to analyze both the main effects of input variables and interactions of those effects on response (output) variables. The experimental design provides a measure of the experiment's ability to provide a measure of statistical confidence in the system response. Further, design of experiments gives the probability that a certain sample size was adequate to detect the response under test—statistical power.

Design of Experiments is the simultaneous solution to the twin challenges of flight test: testing both deeply and broadly while maintaining confidence and power.[12] It allows planning and conducting experiments then analyzing the resulting data so that valid and objective conclusions are obtained. Through each phase of flight test, design of experiments addresses the ultimate objective of achieving a robust flight test—a process affected minimally by external sources of variability.[13] Designed experiments enhance traditional flight test procedures and allow systems engineers and program managers to take advantage of objective, verifiable, and traceable empirical models that reveal both main effects and their interactions, if they exist.[14] Carefully designed fractional factorial designs work well because the data from the main effects are preserved while sacrificing information about higher-order interactions in the interest of saving experimental resources. Advantages of design of experiments as applied to flight testing:

1. Design of experiments provides a structured planning process that can be used to generate test and analysis plans that are comprehensive and efficient.
2. Sequential testing and analysis leads to immediate system discovery and understanding.
3. Empirical statistical models can be used for estimation and prediction which increase test safety.[15]

The principles of design of experiments add value by considering the needs and expertise of stakeholders throughout a disciplined structure. Design of experiments eliminates waste by planning the analysis of response variables and sets the number of test points to deliver the appropriate amount of statistical confidence and power to the stakeholders. It succeeds in extracting as much information about the system under investigation as possible.

A conceptual model has been developed to apply design of experiments techniques to the flight test planning, execution, and reporting processes (Figure 3).[16] Finally, the model will be applied to a case study to illustrate the application of design of experiments to an actual flight test experiment.

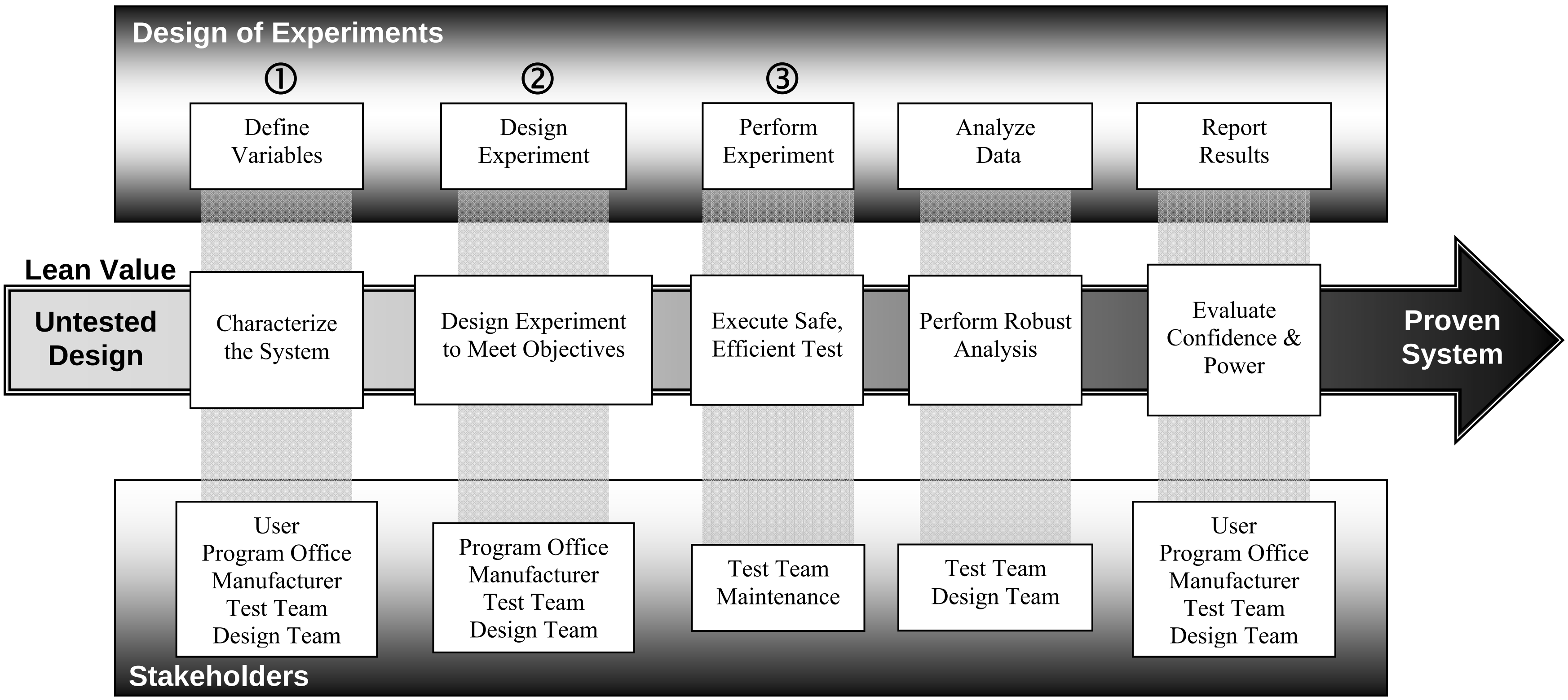

Figure 3. Conceptual Design Diagram

2.1. Define the Variables

One of the most important exercises within the conceptual model is starting with a solid understanding of the system under test. This is achieved by explicitly defining the variables (Figure 3①). Two goals in a lean flight test enterprise are reducing the test budget and completing the test schedule on time while meeting stakeholder needs. Streamlining test plans requires modifying the approach to defining the variables to be investigated. Generally, each investigated variable involves a substantial investment in flight test resources to instrument the test vehicle to record the data, bandwidth and storage to telemeter or record the data, and added cost during data reduction and analysis.

Subject matter experts are consulted in order to consider which variables are active in a system's response. Any variables that will be investigated, however, should be tied directly to a valid stakeholder requirement. A review of the data requirements is therefore considered a critical programmatic decision and should involve candid discussion between the subject matter experts within the test team and the program office. The program office can have an input on the definition of design variables from a programmatic viewpoint, but the focus of this step in the conceptual model should be technical.

2.2. Design the Experiment

Each of the stakeholders in the second step of the conceptual model, design the experiment (Figure 3 ②), is responsible for critical inputs to the success of the flight test program. The program office is responsible to deliver the system's capability to the user with some level of risk defined by cost, schedule, and performance and sets the programmatic scope and performance requirements of the test and evaluation program. The design team identifies the technical data requirements while the test team advises on test objectives that are practical to be achieved in flight test. The manufacturer adds production considerations. When defining the overall scope of the test program, the program office makes decisions based on expert technical input from the full range of stakeholders.

In order to meet the risks of test and evaluation, the needed personnel resources are quite extensive in order to properly plan and execute a test. In fact, up to half of the engineering staffing resources required in a system development program are for test engineering.[17] Among the skill sets accessible to the test team during test planning, the test team should have access to:[18]

1. Operational insight—how is the system designed to be used, in what environments, and against what targets?
2. Expertise in maintaining and repairing the system.
3. Good science and engineering—what are the technical risk areas of the technology under test, what processes stress these risk areas, and how can success or failure best be measured?
4. Test engineers who can identify the instrumentation and data storage requirements for the system, targets and test environment.
5. Specialists in data acquisition, retrieval, filtering and storage.
6. Able management and leadership to communicate its intent and results with good plans, reports, status messages, briefings, etc.

All of these personnel and resources then progress through distinct phases: strategic test plan, detailed test objectives, and planning flight test and analysis.[19] Each phase is linked to specific requirements. An old consultant's rule is that "getting the right design run gives 90 percent of the solution."[20] Design of experiments helps to ensure that the "right design" is selected through a structured, disciplined planning process. Previously, test design was an art. Each test team relied on intuition, past practice, or other strategies to determine each test condition, how the events were to be ordered, how many runs should be made, and how the resulting measurements were to be analyzed in order to draw conclusions.[21]

Engineers and scientists have traditionally generated new test plans by building a comprehensive set of test matrices in order to assess the aircraft in all conceivable operational flight conditions. The intent is to gradually expand the performance envelope and expose the aircraft to operationally representative flight conditions so that the majority of the system flaws can be identified and repaired. These test plans are developed using a combination of past experience and current system knowledge. Many flight tests still use traditional one factor at a time methods and limit analysis to graphical studies.[22] Varying one factor at a time does not reveal the true relationship between the inputs and the response variable and their interactions. Design of experiments actually produces more information about the system in significantly fewer runs than one factor at a time.[23]

One design of experiments technique is critical to constructing rigorous statistical models of system behavior, the test team should run a comparable set of runs each day. With these data, day to day differences can be readily detected or separated out.[24] Flight test is necessarily executed outside the laboratory so daily variations are unavoidable and expected. For example, atmospheric conditions are the quintessential dynamic, unpredictable experimental error. Such factors as airframe fatigue, maintenance status, minor fluctuations in fuel quality, and pilot technique will introduce variability to the test. A test team could plan to repeat a flight condition on each sortie. The results of a replicated flight condition should be the same, so any variance between those data can give a good estimate of the experimental error resulting in a rigorous, statistical model and stakeholder value.

As the flight test investigation progresses toward the edges of the design envelope, the test team can now use their mature, high fidelity models to increase safety. The test team can make operational decisions on which design of experiments factorial matrices to investigate, even between daily test sorties. For example, in a flutter experiment, the interaction between the aircraft's inertial properties and aerodynamic flight conditions can cause airframe oscillations often at unexpected points in the flight envelope. After an unexpected test result, the test team needs to decide how to proceed. Should the airspeed, altitude, mach number, g-loading, or structural properties be changed next? The rigorous analytical models developed in a designed experiment can be examined to tell the test team which variable (or interaction of variables) is most likely to be causing the flutter response. This information can allow the test team to safely continue envelope expansion flight test. The analysis could be performed shortly after touchdown so that the engineers would be more informed to make decisions regarding the next flight's profile.[25] A reduction in poor flight test execution decisions eliminates waste where it is most costly: flying hours with their attendant aircraft, fuel, and maintenance costs.

However, in some developmental tests, safety constraints require an incremental development of the test or flight envelope. In these cases, safety considerations override certain design of experiments techniques such as randomizing the run order.[26] Far from being hobbled by safety considerations, design of experiments can enhance the safety of the flight test enterprise by enabling rigorous simulations based on statistical system response models. Simulation can be used when system problems arise and testing can be performed in a safe environment to assess the level of success of various repair efforts.[27] Finally, as the initial data is returned from test sorties, engineers can incorporate the actual flight test data to increase the fidelity of their models and continue the pattern of providing value to the flight test enterprise.

Factorial matrices enable a simultaneous solution to the twin challenges of test—testing both deeply and broadly, while maintaining power and confidence.[28] A competent power analysis is a critical input to the value stream since it yields an objective basis for determining the number of runs required for a given test. This objective communication with the stakeholders is key when arguing for more resources in a given test effort, so as to achieve adequate power. It is also important to communicate to stakeholders that a proposed test project, if undersized, is likely to fail in its objectives.[29]

After careful consideration of stakeholder requirements, the test team has a number of options to reduce the total number of runs and still provide value to the stakeholders. Among these options are to:[30]

1. Eliminate a variable when not required by the value stream.
2. Reduce the number of levels of a variable if such resolution is not required by the planned analysis.
3. Set aside a less important variable to be run in a side matrix of smaller size as an excursion or demonstration.
4. Fractionating the design by ½, ¼, etc. Such fractional factorials are a powerful and often-used technique.

The most important rationale for the use of fractional designs is the separation of the vital few factors from the trivial many.[31] In order to discriminate between factors, one must make a value judgment with some attendant risk. Also, one needs to guard against the interactions among the various control factors.[32] It is this combination of statistical methods with knowledge of the physics of process that delivers lean value to the customer.[33]

A decision continuum guiding programmatic trade-offs is presented in Figure 4. Time, money, and performance are traded for experimental results with their accompanying statistical confidence and power. An increase in expended resources results in better information for stakeholders. A full factorial experimental design testing every possible combination of variables is very expensive and time consuming, particularly in flight test. However, it will provide the highest fidelity models with the highest statistical confidence in the main effects of test conditions and their interactions. Also, the statistical power to detect smaller effects will be improved. Partial factorials take advantage of the sparcity of effects principle which asserts that most physical systems are dominated by some of the main effects and low-order interactions and most higher-order interactions are negligible.[34] Assumptions are made on which effects and interactions will be confounded (statistically inseparable) with other effects. These assumptions must be made carefully with subject matter expert input because of the increased risk of either not detecting a performance problem or generating a false system model. Also, if conflicts between customer value streams occur, appropriate tradeoffs can be set using the quantitative models derived from the designed experimental analysis.[35] The benefit of a partial factorial, however, is significant savings in time and money for the program.

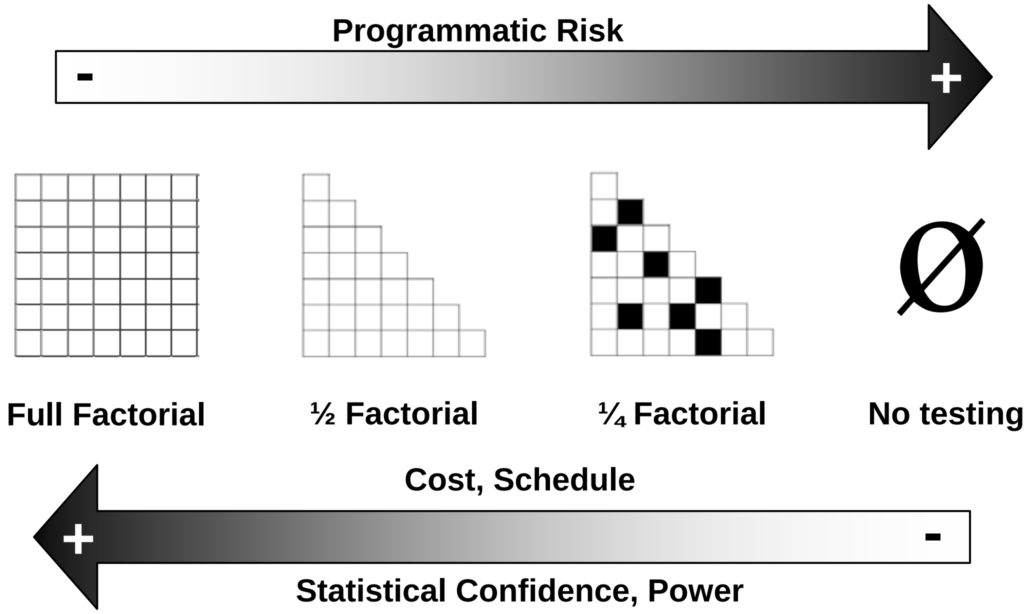

Figure 4. Risk/Performance Tradeoffs in Factorial Design

2.3. Perform the Experiment

Once well-defined test objectives are agreed upon by the stakeholders, the test team and test article maintenance can focus on test safety and efficiency in the third step of the conceptual model, perform the experiment (Figure 3 ③). Test professionals tend to reside in a highly specialized subset of their skill whether as aircrew, engineer, or instrumentation specialist. They are very skilled at accomplishing the test and have mature processes for ensuring minimal physical risk while accomplishing the test objectives. For example, the Air Force Flight Test Center has both technical and safety review processes designed to bring in subject-matter expertise external to the test team. The authority of these reviews is resident in their role to advise the commanders' decisions on the technical adequacy and risk assessment of a given series of flight tests. A candid, formalized process is critical to combating the omnipresent programmatic pressures on any flight test program. These processes are generally credited with vastly improving flight test's safety record.[36] Test efficiency generally follows classic learning curve theory as the team learns the best sequence of test events and becomes more adept at executing the various maneuvers with the desired data quality and within the specified tolerances.

2.4. Analyze the Data

Upon completion of the test execution, the test team proceeds to the fourth step in the conceptual model, analyze the data (Figure 3 ④). Engineers select an appropriate method to analyze the response variables (e.g., time domain, frequency domain, histogram, etc.). The test team then analyzes the data using statistical procedures such as an analysis of variance (ANOVA) to provide objective, verifiable conclusions on the main effects and their interactions.

These data include a statistical confidence and power. A critical part of the flight test analysis is determining the confidence and statistical power of the experiment. Confidence tells the stakeholders the probability that the experiment correctly measured the system effect. Power shows the probability that an effect of a given magnitude would have been detected.[37] Was there a significant system response, and did the experiment have an appropriate ability to detect a response of a given magnitude? Design of experiments provides the stakeholders with confidence in the objective, statistically rigorous analysis of flight test data by which to make major programmatic decisions on high-value acquisitions programs. In general, high power is more costly but is a primary lean value that the flight test enterprise offers the stakeholders, specifically the decision-makers. Also, without a responsible power analysis, the test may lack power to reveal true performance differences, or the test may be oversized, wasting time and resources.[38]

In a properly planned experiment, the data analysis is an explicit part of the experimental design. During the planning step, both input and response variables are identified which helps ensure the appropriate instrumentation is installed and flight test maneuvers are executed. Planning a designed experiment even goes so far as to lay out the data reduction and analysis algorithms. Laboratory and ground test results can be utilized to generate analytic models of the system response. Often, the flight test team will have even proven that the analysis will meet the stakeholders' requirements before actual flight test by injecting reasonable "dummy data" into the data reduction and analysis algorithms.[39] The output will exercise the flight test analysis procedures, give structure (if not actual values) to the analytical models, and give the test team experience with the analysis process. Time spent exercising the process will certainly save expensive flight test resources.

Preliminary analysis helps to diagnose faults in the test equipment since test equipment itself generates one-third of the discrepancies.[40] Without design of experiments, test engineers typically gather aircraft configuration information starting at the time that problems are first detected. However, by using a statistical approach to designing flight test experiments, flight test engineers use flight test data to continuously refine their system response models. When presented with a problem, the models can be pressed into service to diagnose unexpected test results.[41] If the model is sound and generates the appropriate responses for predicted inputs, the test team can rely on the data reduction process and isolate faults in the test apparatus.

Generally, the test team also provides flight test data to the design team for their analysis. The design team includes these data in their system models as the highest fidelity (and most expensive) data possible—actual operating conditions. After a well-designed experiment, the data can be used in future experiments as baseline data for modifications or system enhancements.

2.5. Report the Results.

As part of the fifth step in the conceptual model, the test team reports flight test results and conclusions to the user, program office, design team, and manufacturer (Figure 3⑤). Often the user is interested in flight test results as the test program progresses toward completion and a fielding decision. Another important function of flight test reporting is the preservation of a historical record. The flight test enterprise stakeholders could utilize past test teams' conclusions, recommendations, and lessons learned as more expert information to educate sound programmatic decisions regarding the assumption of risk in a test program.

The test team produces models of system behavior in order to understand the effects of input variables. These models support design changes that bring the capability of the system in line with the user's requirements. In a future design spiral, the same models can help design teams and test teams deliver enhanced capability to the user rather than having to recollect baseline data as in a classical, one factor at a time test.

From the reported conclusions, the test team will make recommendations for design changes to reduce performance risk, further flight test to investigate other areas of the design space, or fielding decisions with some level of known, accepted risk. The program office will use these recommendations along with other stakeholder input to make programmatic decisions. The final result of the application of design of experiments to create value and eliminate waste in the flight test enterprise is the delivery of a needed capability to the warfighter.

Finally, even with the disciplined use of statistics in design of experiments analysis, the test team should never neglect its own subject-matter knowledge in the interpretation of the flight test results. The use of such prior experience rather than ad hoc philosophy concerning the interaction of control-factor effects is the key to understanding test results.[42]

3. Case Study

The case study will utilize an actual flight test program to illustrate the principles contained in the conceptual design. The flight test program in question is the author's flight investigation into demonstrating the effect of active flow control on an F-16B targeting pod-induced buffet on a ventral fin (Figure 5).[43] The targeting pod was equipped with six synthetic jet actuators designed to inject high velocity jets of air into the flow field at a specific frequency. These jets of air were intended to actively control the buffet levels experienced by the F-16B's ventral fin approximately 13 feet downstream of the targeting pod. The experiment did not detect any significant change in buffet level across a wide-ranging flight envelope in 10 flight

Figure 5. Flight Test Experimental Setup

test sorties and 14 flight hours despite good statistical power.

In this case, the synthetic jet actuator designer and manufacturer were interested in refining their design to make it marketable as a practical aeroelastic load control technology. Because this was a research flight test program, the Air Force Research Laboratory was both the user of the technology while maintaining some programmatic control of the project. Also, the Air Force Research Laboratory conducted preliminary wind tunnel research on the effects of the synthetic jet actuators. The test team was comprised of student pilots and flight test engineers at the United States Air Force Test Pilot School.

3.1. Define the Variables

As a stakeholder in the flight test enterprise, the test team led the design and scope of the flight test investigation. During this first step, the test team solicited other stakeholder input to define the variables and characterize the system. The designer of the synthetic jet actuators participated in a telephone conference and gave excellent input toward defining the variables of interest in aeroelastic load control and gave insight into the physics of the synthetic jet actuators. The user, however, was not involved in the definition of the variables and offered the initial wind tunnel investigation only later in the test program. While the test team should lead the effort to define the variables of interest, subject matter experts are critical to this effort.

In this case, only partial understanding of the system mechanics was delivered to the stakeholders via the value stream. This resulted in an insufficient system characterization to make proper design decisions on which variables to concentrate. This left the test team unable to discriminate among all the identified variables: altitude, Mach number, angle of attack, angle of sideslip, and actuator operation. Further, in the absence of subject matter expertise, the test team defined the angle of attack data points relative to level flight due to operational considerations. Given subject matter expert input, however, the test team may have learned that absolute angle of attack was more appropriate to measuring the pod-induced buffet. In fact, defining angle of attack relative to level flight confounds the main effect of that variable with two other variables: altitude and Mach number.

The response variables were set by the design of the test equipment. The F-16B’s right ventral fin was instrumented with six pressure transducers, four strain gages, and two accelerometers. As discovered well into test execution, the accelerometer output signal was saturated by the aerodynamic buffet in the turbulent flow and was unusable. A stakeholder within the test team, the instrumentation engineer, would have been an excellent contributor to the value stream to eliminate the waste of flying with inadequate instrumentation.

3.2. Design the Experiment

The test objective was to demonstrate the effect of operating conditions (Mach number, altitude, angle of attack, angle of sideslip, and synthetic jet actuator) on pod-induced ventral fin buffet.[44] In the absence of a thorough understanding of design of experiments techniques, the test team, on the advice of the subject matter expert, elected to complete an altitude/Mach number survey across the entire flight envelope (Figure 6). Using subject matter knowledge, the test team added investigations into increased angles of attack and sideslip at areas of expected synthetic jet actuator effectiveness. This one factor at a time method is widely criticized as an inefficient use of experimental resources.[45]

Design of experiments provides a selection of alternatives. a better experimental design, a central composite design which combines a two-level factorial with a central point and four axial points that lie beyond and between the factorial points (Figure 7). This design enables a much greater fidelity in the analytical model due to the capability for a higher-order (quadratic) fit. This simple and highly effective experimental design is well-suited to a research investigation.[46]

As shown in Figure 7, the central composite design requires 81 flight conditions or one-half the flight conditions executed in the original case study (Figure 6). Not only does the designed experiment require fewer resources but will better characterize the flight envelope since each altitude/Mach number point is investigated through the full ranges of angle of attack and sideslip. The original case study concentrated flight test resources on the high, fast and low, slow regions and covered the rest of the envelope with only cursory data points.

Another benefit to factorial design is the flexibility of the design. For example, if the test team noted interesting results in the low altitude, low Mach number region of Figure 7, another two-level factorial could be analyzed by adding one point, the blue circle. Such flexibility should be anticipated by the flight test enterprise with test and safety planning that preserves the comprehensive management of risk but allows for flexibility in the placement of specific test points.

3.2.1. Baseline Experiment.

The case study applied a typical grid-based survey method to a mach-altitude flight envelope.[47] In Figure 6, blue dots indicate a typical survey grid placed at the edges of the flight envelope and at regular intervals within the flight envelope. These blue dots represent a test condition at the level-flight angle of attack, αlevel+{0°}, and zero angle of sideslip, β={0°}. The case study subject matter experts also selected specific areas of the flight envelope in which to further investigate the system response. Therefore, variations in angle of attack and angle of sideslip were applied in these flight regions. Red ✦s indicate that data were collected at αlevel+{0°, 2°, 4°}, β={0°}, and green triangles indicate that data were collected at αlevel+{0°, 2°, 4°} × β={–2°, 0°, +2°}. At each test condition, data were recorded with the synthetic jet actuators operating and turned off. The trials consisted of 14 flight hours in an instrumented F-16B flight test aircraft over 16 sorties for a total of 324 test points.

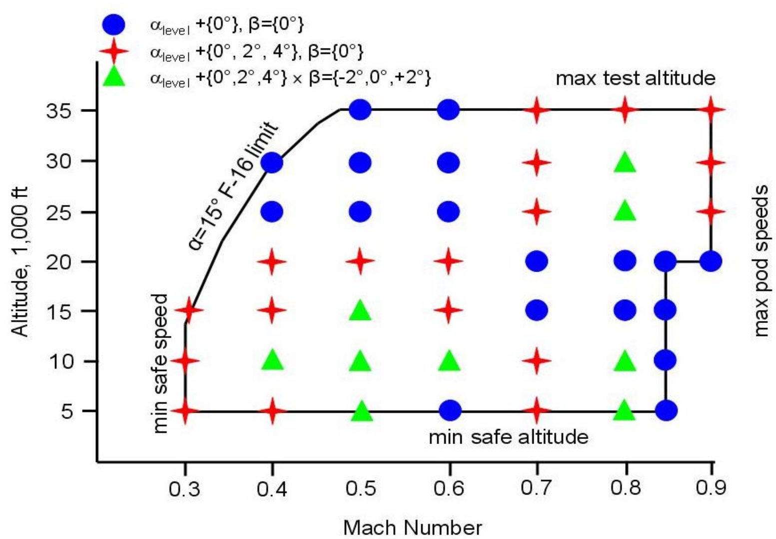

Figure 6. Baseline Experiment

3.2.2. Central Composite Design (CCD).

Figure 7 shows the central composite design (CCD) in altitude, Mach number, angle of attack and angle of sideslip. The CCD was invented in the 1950s by Fisher's son-in-law, Dr. George Box, to address the problems of chemical and industrial research and development testing.[48] It efficiently fits a second order polynomial with interactions to describe the effects of the control variables (Mach number, altitude, α, β, synthetic jet operation) on the vibration response. The CCD requires $2^k + 2k + 3$ center points (without fractionating the factorial portion). With four variables (k=4) this is 16 + 8 + 3 = 27 points. Multiplied by two for each on/off pod condition, the total number of test points is 54, 17% of the baseline experiment. A half-fraction Resolution III design requires 2^4 points, (8 + 8 + 3) × 2 = 38 test points. An effective approach to this problem would be to fractionate the CCD resulting in a factorial with center points then add axial points in factors which prove to be active. Not only does the designed experiment required fewer resources but will better characterize the main part of the flight envelope since each altitude/mach number condition is investigated through the full ranges of angle of attack and sideslip. A drawback to this design is that the axial points on the edges of the envelope are remote from the main factorial portion, and the corners of the envelope are not fully explored. Figure 8 illustrates, however, that the center of the envelope receives much more operational utility than the edges of the envelope.[49]

Another benefit to the CCD is the flexibility of the design, in particular, its ability to be efficiently augmented in numerous ways to explore interesting or unexpected behavior. Points can be added that enable the analysis of an entirely new orthogonal matrix, a CCD+*n* design. If the test team noted interesting results in the low altitude, low mach number region of Figure 7, another two-level factorial could be analyzed by adding a single altitude-mach test condition, the blue square, resulting in a CCD+1 design. The resultant system response model would have higher fidelity in an area of regular operational utility (Figure 7). If variations in angle of attack and sideslip warranted investigation, the entire α_{level}+{0°, 2°, 4°} × β={–2°, 0°, +2°} data set could be added to the experimental design with the addition of 9 test conditions, a CCD+9 design. Fractionating techniques could result in a specific number of additional test conditions (between one and nine) with a known loss of fidelity quantified by the entirety of the CCD design.

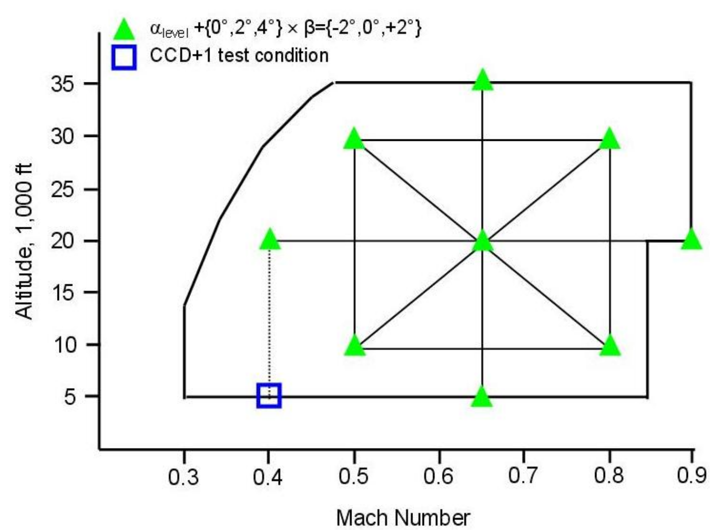

Figure 7. Central Composite Design

Such flexibility should be anticipated by the flight test team with test and safety planning that preserves the comprehensive evaluation of risk but allows for flexibility in the placement of specific test points. Consider that similar factorials could be created to further refine the model at the various edges of the envelope, often a necessity to releasing an operational capability to the warfighter.

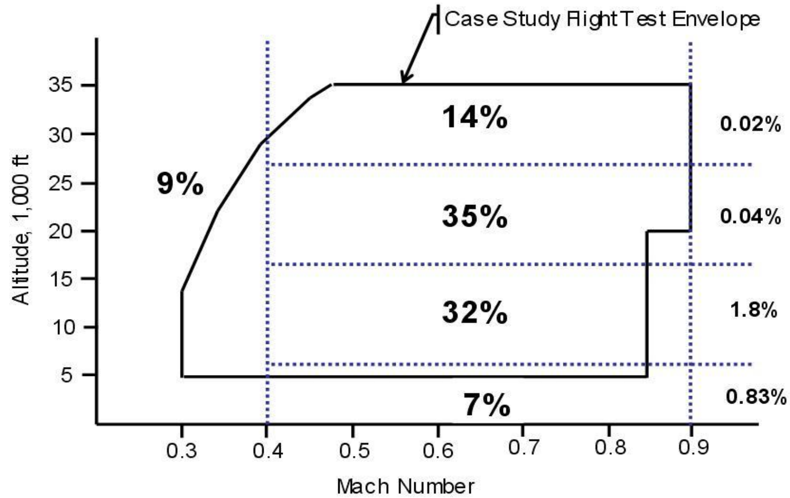

Figure 8. F-16CG operational employment spectra

3.2.3. Face-Centered Central Composite Design (FCD).

Another design (same number of runs) choice would be the face-centered CCD (FCD) where the axial points of the CCD are drawn into the face of the factorial (Figure 9). The FCD was developed by Box[50] to address the problem of physical limitations on the test article that did not allow the axial points to be placed remotely from the main portion of the design. It has the added benefit of efficiently exploring a hyper-cubic test region such as the Mach number/altitude envelope commonly encountered in loads, vibrations, limit cycle oscillation, and environmental characterization testing. While this approach degrades a mathematical property of the design known as rotatability, rotatability is not particularly important in this experiment. As with the CCD, the design allows an explicit model of quadratic responses, characterization and quantification of experimental noise and repeatability, and experimentation with confidence and power. If desired, smaller adjacent matrices can be constructed as shown in the low Mach number/altitude corner of the envelope (Figure 9 dashed lines). Each FCD again consists of 54 test conditions, 17% of the case study.

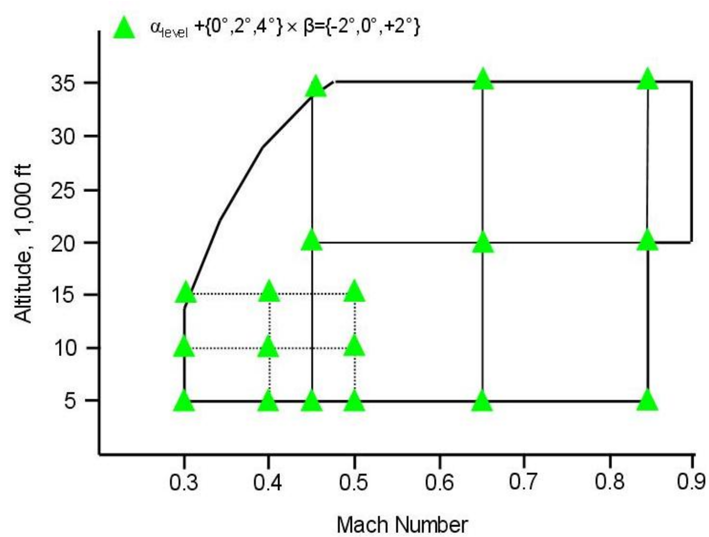

Figure 9. Face-Centered Central Composite Design

Consider that both matrices could be planned in a flight test program and undergo the required test and safety planning processes with flexibility in the placement of the points. The major FCD would be executed, and the data reduced and analyzed before placing the points for the minor FCD. This allows the test team to learn from the fully-modeled system response, better understand the system, and apply that knowledge to the minor FCD and realize the benefits in test efficiency, known risks, and confidence in system response.

3.2.4. Embedded FCD.

In the past two years, Dr Drew Landman of Old Dominion and NASA Langley, and Dr Jim Simpson from Florida State[51] have been considering a new class of designs for modeling highly nonlinear processes. As shown in Figure 10, one can consider embedding two FCDs – one within the other (blue and green triangles in Fig. 5). The resulting composite design allows up to third-order polynomials to be fit through the experimental region at modestly increased cost. The two 54-test-point FCDs can be run together for a total of 108 test points (minus 3 repeated common center points.)

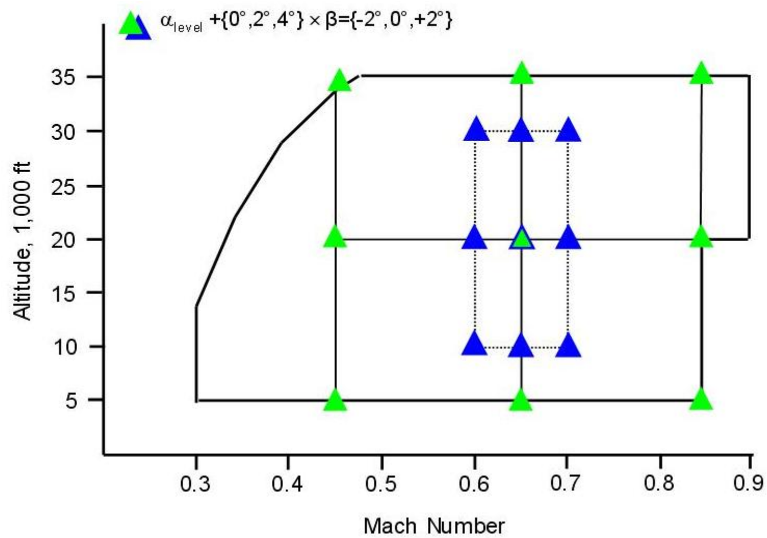

Figure 10. Embedded FCD

3.3. Perform the Experiment

When performing this experiment, the test team and maintenance will work to ensure disciplined, safe, and efficient flight test. The test team will design flight test maneuvers that safely collect data within tolerances specified by the subject matter experts. To increase safety, a build-up approach will be used as much as possible before approaching any flight envelope limits. Finally, the test conditions will be efficiently ordered to take advantage of aircraft energy and operational considerations.

Flight test maneuver tolerances recognize the fact that flight is dynamic and test pilots can only control so many factors. A set of tolerances give the test team a definition for data quality and should be traceable to specific data requirements. These tolerances should be derived from the value stream to create valuable data and eliminate needlessly restrictive tolerances that are more difficult to fly and can create waste. In the case study, the test team used their operational experience to set the tolerances to place priority on data collection and allow quick progression from one test condition to another. Fewer test points represented by a designed experiment would allow the test pilots to spend more time on a test condition and operate to tighter tolerances. Another benefit of the designed experiment is that replication allows for variability between two identical test points to measure the within-group variation in the analysis of variance.

In a designed flight test experiment, the techniques of randomization and blocking often oppose safety and efficiency. Randomization calls for the data points to be executed randomly. Blocking recognizes nuisance variability contained in a natural division of the experiment—aircraft sorties, in the case study.[52] The safety technique of build-up starts in the middle of the

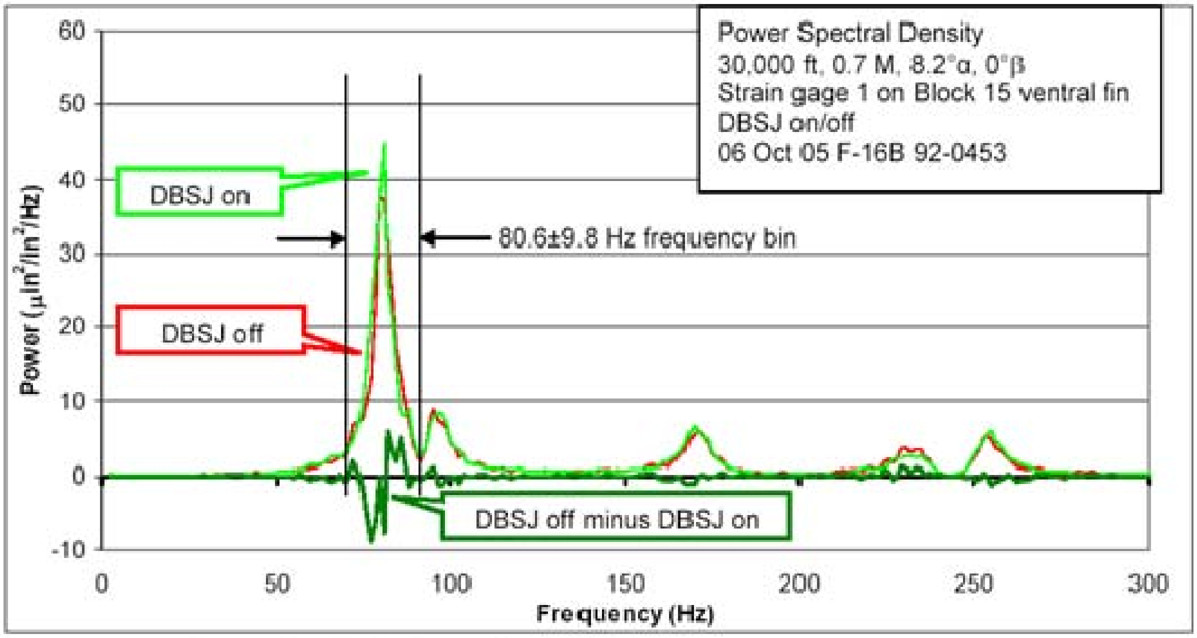

Figure 11. Power Spectral Density Chart for the F-16B Ventral Fin Strain

aircraft envelope and approaches the edges of the flight envelope incrementally as the test team gains experience with the system in flight. Also, test efficiency desires to collect data in similar energy states; time and fuel is wasted by needlessly changing altitude and airspeed. An acceptable compromise is to randomize points in the middle of the envelope and collect that data in the first sorties (blocks) then expand the investigation to different regions of the flight envelope. A direct measure of variability between sorties is achieved by collecting data at the center point (20,000 ft, 0.65 Mach in Figure 7) during each sortie before moving on the different test conditions.

Finally, the test team should be cognizant of the hidden effects of operational considerations on data variability. Often, detailed records of weather conditions, aircraft fuel state and system operations at the time of the data collection, even the names of the pilots and flight test engineers are important data. In the case study, replicates of a test condition should be planned early and late in a sortie to average out aircraft gross weight and test team qualitative comments should be recorded for examination during data analysis (e.g., turbulence, winds, pilot technique in flying the test condition, etc.).

3.4. Analyze the Data

The aerodynamic buffet on the F-16B's ventral fin was measured using strain gages and static pressure transducers, which produce very noisy data. A very effective method to view these data is in the frequency domain via a power spectral density (PSD) plot. Figure 11 is a typical power spectral density plot of the F-16B ventral fin's strain.[53] The physical significance of a power spectral density plot for aeroelastic testing is a direct representation of energy content at a specific frequency, which is absorbed by the structure's modes of flexion. Aeroelastic load control seeks to reduce these structural movements and, in turn, high-cycle fatigue.

The data analysis algorithm calculates the area under a power spectral density peak that is contained within a frequency bin. An effect of a flight condition or the synthetic jets (as shown in Figure 11) is a change in that calculated area. This datum is recorded in the factorial with other data, and an analysis of variance is performed on the factorial to determine if the effect of operating the synthetic jets on the ventral fin buffet is significant. A benefit of the design of experiments is that variance in the data from sources other than the main effects and their interaction is separated as random noise.

The effect of the synthetic jets was shown to be insignificant in the case study, so a power analysis was needed. A power analysis shows the experiment's probability of detecting an effect of a given magnitude. Replicated factorials have good to excellent power which means that even a small effect of the synthetic jets on ventral fin buffet would have most likely been detected.

In order to compare the efficiency and efficacy of four experimental methods, the data from the case study were formed into data sets as prescribed by each designed experiment. However, the original case study did not include some test conditions that a designed experiment would have prescribed, particularly in the $\beta=\{-2°, +2°\}$ data set. Therefore, the $\beta=\{0°\}$ data sets for the various designed experiments were completed using interpolation, extrapolation, and hand-fairing of system response curves. The case study's $\beta=\{-2°, +2°\}$ data set, however, did not have enough original data to be completed and could not be completed. Analyses of Variance (ANOVA) were used to analyze only the $\beta=\{0°\}$ data and statistical power was calculated for each design.

3.4.1. Statistical Confidence.

A tester can make one of two errors in a test program – determine that a factor makes a difference in the response when it does not (measured by statistical confidence) or to fail to detect a difference when such a difference really exists (measured by statistical power.) Unfortunately for aerospace testers, the probability of making the first type of error is referred to by statisticians as "alpha," normally reserved for angle of attack. The protection from making the alpha error is referred to as (1 – alpha) or confidence. The second type of error is again, unfortunately, referred to as "beta," usually reserved for angle of sideslip, and (1 – beta) is referred to as power. So, the tester would like to have high confidence in his conclusions, those conditions determined to make a difference really do make a difference, while preserving high power, the ability to detect real differences when they exist.

Table 1 presents the experimental designs and the ranges of the four independent variables in the data set: altitude, Mach number, angle of attack (α), angle of sideslip (β), and synthetic jet operation.[54] Note that there are two data points (synthetic jet on and off) for every test condition in a designed experiment. Table 1 includes a p value for the effect of the system under test on each system response: strain (80 Hz), pressure (158 Hz), and pressure (225 Hz). Note that none of the synthetic jet p values reach a level of significant effect ($p < 0.05$). This is important to the tester because it shows the synthetic jets were unable to affect the system response in the flight envelope of interest. The conclusion is that the synthetic jets did not significantly affect the vibration of the F-16 ventral fin. An analysis of the 324 data points gathered in the baseline experiment reached the same conclusion, with the same confidence as the designed experiments. Importantly, the designed experiments provided the same conclusion with 16% to 30% of the test resources expended.

Table 1. Results of ANOVA for Designed Experiments

Experimental Design	Number of points in data set (% baseline)	Response Variable	p value
Baseline	324 (100%)	Strain (80 Hz)	0.77
		Pressure (158 Hz)	0.45
		Pressure (225 Hz)	0.95
CCD	54 (17%)	Strain (80 Hz)	0.65
		Pressure (158 Hz)	0.98
		Pressure (225 Hz)	0.77
CCD+3	60 (19%)	Strain (80 Hz)	0.65
		Pressure (158 Hz)	0.94
		Pressure (225 Hz)	0.82
Face-Centered CCD	54 (17%)	Strain (80 Hz)	0.66
		Pressure (158 Hz)	0.89
		Pressure (225 Hz)	0.82
Embedded FCD	108 (33%)	Strain (80 Hz)	0.62
		Pressure (158 Hz)	0.88
		Pressure (225 Hz)	0.79

3.4.2. Statistical Power

Figure 12 compares the statistical power achieved by the several types of experiments.[55] The SME-designed baseline experiment has the statistical power to detect very small changes in ventral fin vibration. For example, a synthetic jet actuator effect of 0.50 standard deviation, 0.50σ, has a 90% probability of being detected by the baseline experiment. While a CCD only has a 40% probability of detecting a 0.50σ effect or would take a 1.0σ effect to have a 90% probability of detection. It should be noted that such small changes in the response (half the noise level or lower) are typically not of much practical interest in a flight test program. One might take the position that the SME-generated baseline was *too* powerful for the intended purposes (finding a solution to ventral fin buffet) and therefore wasteful of test resources.

Statistical power is a crucial commodity in flight test. One can make the case that power is the “gold standard” by which any test program should be judged. A test that lacks power will fail to detect the very behavior it was resourced to explore, while an over-powered test simply wastes resources. The baseline experiment can detect very small effects with a high probability of detection but comes at a cost of six times the invested resources of the CCD experiment. The test team can select a required statistical power for a given effect size and design and resource a test program with the statistical rigor to successfully investigate the system response. The alternative to a designed experiment is to blindly collect and inspect data with no insight into whether or not that test could have detected the effect of interest.

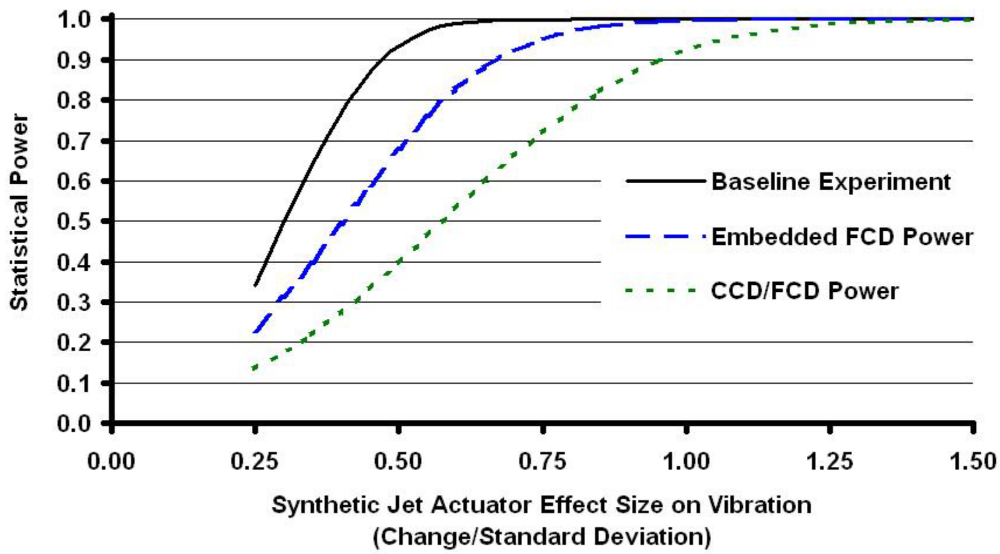

Figure 12. Experimental Statistical Power

3.5. Report the Results

The synthetic jet's effect on the F-16B ventral fin will be reported to the design team, program office, manufacturer, and user. The customer will receive data indicating the significance of the synthetic jet's effect bolstered by statistical confidence and power analyses. In this case study, the results were that the synthetic jets did not have an effect. As a result of the designed experiment, the design team now has a rigorous model of the flight conditions on an F-16B ventral fin with which to modify or further investigate the synthetic jets. Because the proposed central composite design only requires 54 flight conditions to fully execute, the test team has an enviable set of alternatives:

1. Stop testing and use the analytical model of the F-16B ventral fin response for future testing with an improved synthetic jet actuator configuration. The test program would have expended one-half of the flight test resources of the original test with substantially better information returned to the customer.
2. Replicate the entire factorial to further increase statistical confidence and power and return a more rigorous analytical model for the same flight test resources of the original case study. The benefit of a replicated factorial is that, for each flight condition, experimental error can be measured since the two data points can be directly compared.
3. Add another variable and repeat the entire factorial. The case study conclusions stated "insufficient test points were collected to complete an experimental investigation of the [synthetic jets'] effect on ventral fin buffet when the F-16 did not have external wing tanks…further flight testing is required."[56] The entire central composite design could be repeated with a different aircraft configuration—no external wing tanks in this case. A

more complete analytical model of the ventral fin response could be delivered to the stakeholders with an identical investment of flight test resources.

The conclusions and recommendations of the test team are particularly vital to the report. As they are closest to the investigation, the test team has special insight into the. Based on the results, conclusions, and recommendations, the program office can evaluate the cost, schedule, and performance risk of continuing the synthetic jet development. Even if the program is discontinued, the analytical model of the F-16B ventral fin under flight conditions is valuable to any future investigations.

4. Conclusion

The application of the design of experiments to a flight test investigation is valuable by both increasing the information gained from the data collected and helping to ensure efficiency of test. The information gained from the data set is fully exploited to generate a system response model with a known statistical confidence and power. The test is efficient because no more data are collected than required to produce that system response model.

The flexibility in designed experiments allows for test teams to apply safe test practices, investigate interesting results, and continually build on the data set in a structured manner. Most safety planning addresses a build-up approach. Designed experiments can assist that effort by producing insight into the system response in the initial data collection which will educate the judgment of the test team and mitigate risk through increased confidence in the predicted system response as the data collection progresses closer to the predicted edge of the operating envelope. Also, the flexibility of the experimental designs allow the test team to analyze a complete data set, note areas of interest, and completely investigating those regions. Often, the test team can use part of the existing data set instead of launching an entirely new test program. In a designed experiment, data is collected in such a manner that the effect of an independent variable can be isolated from the system response. Therefore, future test efforts can use previous data collected in a designed experiment as part of the structure of another designed experiment.

In times of enormously expensive flight test programs, the efficiencies realized through the application of designed experiments to flight test could mean the difference between the timely delivery of a needed capability to the warfighter; an over cost, late, under-performing system; or outright cancellation of the system. Design of experiments has the capability to make flight test safer, more effective, and more efficient.

VITA

Aaron Alexander Tucker received a BS cum laude in Aerospace Engineering in May 1997 and a MS in Aerospace Engineering (Aerospace Design and Innovation) in August 1997 from the University of Southern California in Los Angeles, California. His research included turbulent boundary layer transition, delay of boundary layer separation using active piezoelectric actuators, full-scale drag testing of platooning vehicles, and aerodynamic optimization of micro-unmanned aerial vehicles. Aaron received a MS in Systems Engineering in May 2007 from the University of Missouri-Rolla in Rolla, Missouri. He is currently a doctoral student in the Department of Aerospace Engineering at Texas A&M University studying boundary layer stability and transition. His dissertation topic is designing a laminar flow control flight experiment.

Lieutenant Colonel Tucker was commissioned in the United States Air Force in May 1997. After pilot training, he served as a C-5 Galaxy Flight Examiner Aircraft Commander at Travis Air Force Base, California from May 1999 to August 2003 during which time he had the privilege of commanding combat missions into both Afghanistan and Iraq. After a year as an Air Force Intern at the Pentagon, Washington DC, he attended the United States Air Force Test Pilot School at Edwards Air Force Base, California in 2005. He served as an Assistant Operations Officer in the 418th Flight Test Squadron and a C-5 Galaxy and C-17 Globemaster III Experimental Test Pilot at Edwards Air Force Base, California. Lt Col Tucker is a member of Tau Beta Pi, the Society of Experimental Test Pilots, and the American Institute of Aeronautics and Astronautics.

---

[1] Kreisher, Otto. "Flight-Test Worries." Air Force Magazine (February 2006): 60.

[2] "Interview with Marvin Sambur, Assistant Secretary of the Air Force (Acquisition)." Program Manager 22, no. 4 (July-August 2003): 3.

[3] Veselenak, Jeff and Lori Beutelman. "The Evolving Role of Developmental Test and Evaluation at the Air Force Flight Test Center." American Institute of Aeronautics and Astronautics 2004-6856, presented at the USAF Developmental Test and Evaluation Summit, 16-18 November 2004, Woodland Hills CA.: 5.

[4] Murman, Earll, Thomas Allen, Kirkor Bozdogan, Joel Cutcher-Gershenfield, Hugh McManus, Deborah Nightingale, Eric Rebentisch, Tom Shields, Fred Stahl, Myles Walton, Joyce Warmkessel, Stanley Weiss, and Sheila Widnall. Lean Enterprise Value: Insights from MIT's Lean Aerospace Initiative. New York: Palgrave, 2002: 169.

[5] Murman: 194.

[6] Carreras, Carmen E. "Opportunities for Lean Thinking in Aircraft Flight Testing and Evaluation." Master's thesis, Massachusetts Institute of Technology, 10 May 2002: 33.

[7] Carreras: 38.

[8] Carreras: 38.

[9] Carreras: 37.

[10] Skalamera, Robert, Colonel Warren Anderson, John Snoderly, and David P. Brown. "Systems Planning, Research, Development, and Engineering Career Field." Defense AT&L 35, no. 4 (July-August 2006): 41.

[11] Murman: 204.

[12] Hutto, Gregory. "Design of Experiments: Meeting the Central Challenge of Flight Test." American Institute of Aeronautics and Astronautics 2005-7607, presented at US Air Force T&E Days, 6-8 December 2005, Nashville TN: 11.

[13] Montgomery, Douglas C. Design and Analysis of Experiments, 5th edition. New York: John Wiley & Sons, Inc., 2001: 1.

[14] Simpson, James R. and James W. Wisnowski. "Streamlining Flight Test with the Design and Analysis of Experiments." Journal of Aircraft 38 no. 6, American Institute of Aeronautics and Astronautics 0021-8669 (November-December 2001): 18.

[15] Simpson,2001: 18.

[16] Tucker, Aaron A. and Cihan H. Dagli, "Design of Experiments as a Means of Lean Value Delivery to the Flight Test Enterprise," Journal of Systems Engineering 12 no. 3, 2009.

[17] Murman: 218.

[18] Hutto: 1.

[19] Simpson,2001: 4.

[20] Nair, Vijayan N., ed. "Taguchi's Parameter Design: A Panel Discussion." Technometrics 34, no. 2 (May 1992): 132.

[21] Hutto: 1.

[22] Simpson,2001: 1-4.

[23] Simpson,2001: 2-3.

[24] Hutto: 14.

[25] Simpson, James R. and James W. Wisnowski. "A Systematic Approach to Planning Flight Tests Using Designed Experiments." American Institute of Aeronautics and Astronautics 2000-2695, presented at the 21st AIAA Aerodynamic Measurement Technology and Ground Testing Conference, 19-22 June 2000, Denver CO: 8.

[26] Hutto: 13.

[27] Simpson, 8.

[28] Hutto: 11.

[29] Hutto: 13.

[30] Hutto: 11.

[31] Nair: 146.

[32] Phadke, Madhav S. "Design Optimization Case Studies." AT&T Technical Journal 65, no. 2 (March April 1986): 59.

[33] Baudin, Michel. "When to Use Statistics, One-Piece Flow, or Mistake-Proofing to Improve Quality." Institute of Industrial Engineers, http://iienet.org accessed on 30 Apr 08.

[34] Montgomery: 303.

[35] Phadke: 63.

[36] "The Air Force Flight Test Center Test Safety Review Process." The Society of Experimental Test Pilots, accessed at http://www.flighttestsafety.org/bestpractices/afftcsrb.doc on 10 Feb 07.

[37] Montgomery: 34.

[38] Hutto: 5.

[39] Hutto: 24.

[40] Murman: 226.

[41] Simpson,2001: 2.

[42] Nair: 139.

[43] Tucker, Captain Aaron A., Captain Shawn D. Morgenstern, Captain Heather C. Giebner, Captain Dick Wong, Captain Aniello Violetti, and Lieutenant Commander Gregory Burgess. "Limited Characterization of Dual Bimorph Synthetic Jet Actuators: Aeroelastic Load Control—Phase I." American Institute of Aeronautics and Astronautics 2007-475, presented at the 45th Aerospace Sciences Meeting and Exhibit, 2007.

[44] Tucker, 2007: 4.

[45] Simpson,2001: 2-3.

[46] Montgomery: 274.

[47] Tucker, 2007.

[48] Montgomery.

[49] Wickert, Maj Douglas P. "Mitigating Risk at the Edge of the Envelope: Revising Basic Aircraft Limit Test Procedures." Proceedings of the 50th Annual Society of Experimental Test Pilots Symposium, September 2006, Anaheim California.

[50] Montgomery.

[51] Landman, D., Simpson, J., Mariani, R., Ortiz, F., and C. Britcher. "A High Performance Aircraft Wind Tunnel Test using Response Surface Methodologies." American Institute of Aeronautics and Astronautics 2005-7602, presented at the US Air Force T&E Days 2005, Nashville TN.

[52] Montgomery: 12-13.

[53] Tucker, 2007: 7.

[54] Tucker, Aaron A., Hutto, Gregory T., and Cihan H. Dagli, "Application of Design of Experiments to Flight Test: A Case Study." Journal of Aircraft 47 no. 2, American Institute of Aeronautics and Astronautics (March-April 2010).

[55] Tucker, 2010.

[56] Tucker, 2007: 13.