Systems Theoretic Process Analysis as a Practical Tool for Comprehensive Flight Test Hazard Identification

by

Noam D. Eisen

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICS AND ASTRONAUTICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

© 2024 Noam D. Eisen. This work is licensed under a CC BY-NC-ND 4.0 license.

The author hereby grants to MIT a nonexclusive, worldwide, irrevocable, royalty-free license to exercise any and all rights under copyright, including to reproduce, preserve, distribute and publicly display copies of the thesis, or release the thesis under an open-access license.

Authored by: Noam D. Eisen
Department of Aeronautics and Astronautics
May 17, 2024

Certified by: Nancy Leveson
Professor of Aeronautics and Astronautics, Thesis Supervisor

Accepted by: Jonathan P. How
R. C. Maclauren Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee, Department of Aeronautics and Astronautics

Systems Theoretic Process Analysis as a Practical Tool for Comprehensive Flight Test Hazard Identification

by

Noam D. Eisen

Submitted to the Department of Aeronautics and Astronautics
on May 17, 2024 in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN AERONAUTICS AND ASTRONAUTICS

ABSTRACT

Flight test is an endeavor inherently imbued with risk. In order to conduct flight testing safely, hazards of consequence must be identified and mitigated in advance of testing. While adequate practices are widely in place for the mitigation of hazards that have been identified, the practices generally used to reveal and identify hazards in the first place rely on brainstorming and other fragmentary methods that can leave critical gaps in safety preparedness.

Mainstream flight test risk management techniques such as Test Hazard Analysis (THA) rely on expert brainstorming for the identification of hazards, and lean heavily on experience and lessons learned from subjectively 'similar' past test programs. Frequently for a given new program, the THA report from a past program is simply duplicated in full, with edits then made to accommodate perceived differences. Such processes have left critical gaps in hazard identification coverage even where 'similar' technologies and test methods are concerned; moreover, as airborne technologies evolve– with increasingly complex systems interactions, software, and human/machine interplays– the gaps in hazard coverage are becoming ever more pronounced, leaving the legacy risk management techniques unable to support a level of safety that meets industry needs.

With each hazard in a THA documented separately, and mitigations addressed individually to each hazard, no underlying framework is available to unify hazard identification or analysis across functionalities or disciplines. Safety reviews and preflight briefings based on THA become lengthy and disjoint, as well as potentially incomplete.

Systems Theoretic Process Analysis (STPA) is a forward-looking safety analysis methodology grounded in systems theory. Based in the System-Theoretic Accident Model and Processes (STAMP) model, STPA is able to produce meaningful results even where other methodologies struggle, such as in systems involving software, human interactions, or other forms of complexity such as exist in aviation and flight test. This thesis proposes to apply STPA to the problem of hazard identification and management in flight test, specifically focusing on piloted ('manned') aircraft.

The state of the art in THA is examined, and STPA and THA are compared in frameworks, constructs, and work products in the context of flight test. STPA is applied to an example flight test campaign to illustrate its use in test hazard identification. A final section describes more broadly how STPA could be incorporated into flight test organizations now, and in a future where STPA is more widely used by design and engineering departments as well.

Thesis supervisor: Nancy Leveson
Title: Professor of Aeronautics and Astronautics

Contents

0. Abstract 3

1. Introduction 8

1.1. Motivation 8

1.2. Research Questions 13

1.3. Research Methodology 13

1.4. Contributions 14

2. Background 15

2.1. The Role of Flight Test in the Larger Safety Picture 15

2.2. Hazard-Informed Flight Test Management Processes 19

2.3. Flight Test Hazard Analysis: Review of Current Practices 23

2.4. Limitations in THA 33

2.5. Requirements for an Improved Methodology 36

2.6. Systems Theoretic Process Analysis (STPA): an Overview 38

3. STPA in Flight Test 40

3.1. STPA and THA: Merging Mental Models 41

3.2. Harnessing Abstraction to Bring Hazards into Focus 58

3.3. Testing and Violating Safety Assumptions 60

3.4. Hazard Mitigation and Analysis of Changes 62

4. Example Use of STPA to Identify Flight Test Hazards 65

4.1. Auto-GCAS Testing for General Aviation 66

4.2. STPA in Action: A Walk-Through 69

4.3. Observations 74

4.4. Discussion 75

5. Adopting STPA: Promises and Challenges 77

5.1. STPA at the Local Level: Flight Test Organizations & Processes 77

5.2. 'Concept-to-Boneyard STPA': Possibilities & Pipe Dreams 79

6. Conclusions 81

7. Acknowledgements 84

8. Appendix 86

9. Bibliography 90

List of Figures

Figure 1: 'Hazards River' decomposition and classification of hazard, adapted from [15] 9
Figure 2: Desirable precedence of hazard mitigation [16] 10
Figure 3: Risk level scale comparison [50] 17
Figure 4: The 'safety space' [44] 18
Figure 5: Risk Management Process outline [23] 20
Figure 6: AFTCI 91-202 Process Flow [56] 21
Figure 7: Cessna Citation Sovereign Test Planning Review and Approval Process [57] 22
Figure 8: Elements of THA described [60] 24
Figure 9: Example THA: Engine Temperature Exceedance Hazard [51] 24
Figure 10: Example flight profile-related hazard analysis [3] 26
Figure 11: Example project equipment-related hazard analysis [3] 26
Figure 12: Representative 2-D Risk Matrix [74] 32
Figure 13: purely social system example 43
Figure 14: system, system boundary, and environment [78] 43
Figure 15: controller elements (adapted from [76]) 44
Figure 16: generic control loop [76] 46
Figure 17: example control structure (adapted from [76]) 47
Figure 18: generic control structure (adapted from [76]) 48
Figure 19: The four steps of STPA [81] 50
Figure 20: 'Day-of-flight' control structure, flight test with chase and ground station 53
Figure 21: annotated UCA description example [76] 55
Figure 22: example UCA table 57
Figure 23: 'On the aircraft' control structure, 2 crew 59
Figure 24: Design/redesign and analysis/re-analysis cycle [90] 62
Figure 25: Re-analysis process following modification [90] 63
Figure 26: Re-analysis guidance following modification (modified from [90]) 64
Figure 27: Auto-GCAS flight test system diagram [94] 66
Figure 28: Cozy Mk-IV test platform [96] 68
Figure 29: Auto-GCAS flight test control structure 71
Figure 30: Listing of controllers' control actions 72
Figure 31: UCA table structure excerpt 73

1. Introduction

> "[T]o keep at the problem [of flight] long enough to really learn anything positively [one] must not take dangerous risks. Carelessness and overconfidence are usually more dangerous than *deliberately accepted risks*." (emphasis added)
>
> – *Wilbur Wright in a letter to his father, September 1900.* [1]

1.1. Motivation

In April 2011, four crewmembers lost their lives in a fatal crash of the prototype Gulfstream G650 business jet they were testing. "Contributing to the accident was Gulfstream's failure to… [ensure that] potential hazards had been fully identified."[2]

Flight test is an endeavor inherently imbued with risk.[3], [4], [5], [6] Whether experimental, developmental, operational, or otherwise, flight test fundamentally involves probing the edges of the known to discover, validate, or quantify the unknown.[3] "Far more so than with any other system, airborne test incidents have the ability to rapidly escalate in seriousness, leading to the potential loss of aircraft and crew."[7] "Not infrequently, the cost for obtaining the data that reduces uncertainty is realized in catastrophe."[8]

Many hazards can emerge in undertaking test activities, and in order to conduct flight test safely, hazards of consequence must be mitigated prior to exposure such that the residual risks and severities of their occurrences are reduced to acceptable levels. Before any hazard can be purposefully mitigated, however, its very existence must first be identified. Those hazards that are not successfully identified in advance (the 'unidentified risks,' or 'unknown unknowns' as they are colloquially described [9], [10], [11], [12], [13]) can produce catastrophic outcomes when they emerge unexpectedly during test. "[T]he complexity of today's modern systems and the effect of unknown unknowns on flight test safety" is a chief concern of the flight test discipline.[14]

This thesis aims to improve flight test safety by advancing the state of test hazard identification, even in novel or complex systems under test.

A conceptual decomposition of hazards and their management is illustrated in Figure 1, graphically tracing the flow and categorization of hazards as treated in a nominal flight test program. At left in the figure is the entire complement of 'initial hazards' present at the outset of a given flight test program; included are both 'identified hazards' and 'unidentified hazards.' Identified hazards are those hazards that have been recognized as existing and thus can be treated, whereas unidentified hazards are the 'unknown unknowns' that are unwittingly assumed, and which, having not been recognized, cannot be prepared for. After analysis, some of the identified hazards will be deemed 'unacceptable hazards,' too dangerous or otherwise impactful to tolerate; others will be deemed 'acceptable hazards' and willingly assumed in operation. Some of the unacceptable hazards may be eliminated (yellow arrow), thus removing those hazards from the system; others may be reduced or controlled (green arrow) and thus shrunken to less impactful 'mitigated hazards.' The combination of mitigated hazards and acceptable

hazards is collectively termed ‘assumed hazards,’ and describes all hazards knowingly assumed in the program. Assumed hazards together with unidentified hazards comprise the total complement of ‘residual hazards’ remaining present in the program, as shown at right in the figure. Not shown in this illustration are new hazards introduced into the program along the way, wittingly or unwittingly, either in the course of managing the identified hazards or simply in the course of modifying the test program as the program progresses. The latter might include changes to program goals; data requirements; maneuvers, test points, or techniques to be used; the test aircraft or other assets; personnel; location; safety or other equipment; and other aspects of conduct, equipment, staffing, methodology, continuity, scheduling, and execution.

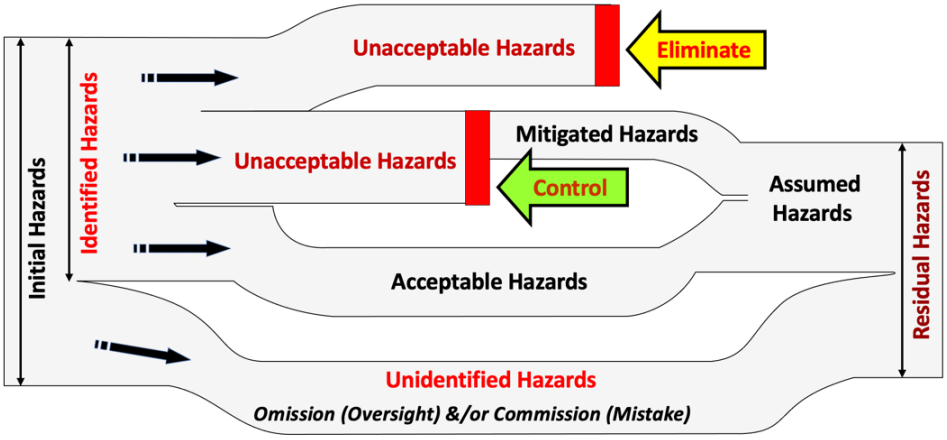


Figure 1: ‘Hazards River’ decomposition and classification of hazard, adapted from [15]

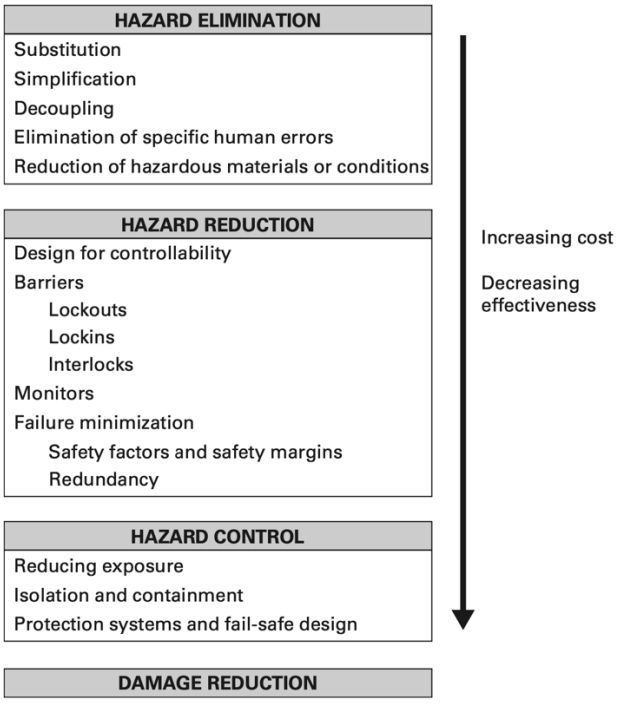


Figure 2: Desirable precedence of hazard mitigation [16]

Unfortunately, "[t]he history of flight testing from the days of mythology and antiquity to the present day is littered with accidents many of which could have been avoided if effective safety management practices had been deployed."[7] While sound processes and methodologies are widely in place for controlling flight test hazards once they have been identified (Figure 2, for example, outlines a best practice prioritization of mitigation approaches), the step of identifying hazards in the first place still relies on incomplete solutions that can leave critical gaps in preparedness.

To compensate for the lack of a comprehensive hazard identification methodology, the flight test community relies principally on individual experience and brainstorming based on the insights of test veterans for the identification of hazards. These experts draw on scenarios and incidents they have seen in the past to imagine what hazards may affect a given new program or a particular system under test (SUT). The United States Air Force "test-safety planning method relies on senior engineers and operators around a table drawing from their experience to identify the most plausible chains of events[1] that could lead to accidents... Often the final safety plan document is built by copying the plans from previous similar tests" and revising them as needed [17].[2]

Civilian flight test organizations largely mirror these approaches, using "personnel who are involved in similar projects… [to help] identify areas possibly overlooked" [5]. The curricula of accredited Test Pilot schools teach these same backward-looking, history-based strategies. So reliant is the aviation community on

[1] It is worth noting this 'chain of events' framework is a limiting view of system safety, as will be shown.
[2] NASA, in conjunction with other organizations, created an online test hazards database to support this type of approach.[18]

brainstorming [19], [20] by "persons who have [past] first-hand experience with similar aircraft designs" [21] for the step of hazard identification that the approaches commonly used to identify flight test hazards have been described as "akin to storytelling" [17] and collectively termed "tools of experience" [8].

The pool of experienced professionals to draw upon for hazard identification is limited; at NASA, "[t]he departure of many senior and experienced personnel [in the late 1990's] was identified as being a major contributor to poor risk assessment in several high visibility programs."[21] Similar challenges are faced in certain test and certification offices of the FAA as personnel with "experience with similar projects" retire [22], causing a "loss of corporate memory" [23] without which these approaches can become nearly worthless.

The history-based, brainstorming-driven approaches currently in use have already proven to be susceptible to failures even when legacy technology is at play. [8], [24], [2], [25], [26], [27] Brainstorming relies on imagination, but "the complexity of aircraft design, the proliferation of software based systems, etc.—all of these things create hazards that we haven't even begun to imagine" [28]. Systems interactions can quickly produce emergent behaviors that are not at all obvious on their surface. As technology advances, it may "take quite some time before a pool of experts is even created" [17]. This exposes the flight test profession to the risk of unnecessary casualties and recked equipment in the meantime. And as airborne systems under test grow increasingly complex and novel– with "complex software, system interactions, and human/machine interactions" at play [28]– system behaviors can change drastically by even small modifications, making expertise difficult to capture, and the legacy methodologies for hazard identification ever less able to deliver the level of safety that the profession demands.[12], [17], [24], [29], [30]

Attempts have been made to formulate comprehensive frameworks for flight test hazard and risk management [8], [31], [32], [33], but few if any meaningfully tackle the hazard identification phase. Instead, they focus on decision-making and management of already-identified risks. Furthermore, in addition to being generally incomplete or onerous solutions in practice, the frameworks proposed have been "too academic, too complicated, and too theoretical to be easily and practically applied to flight test. This may partly explain why both military and civilian flight test has largely failed to adopt any of the newer risk management frameworks."[8]

Some theoretical constructs, such as the idea of the 'accident chain' [34] and Reason's famous 'Swiss cheese' model [35], [36], [37], have penetrated the vernacular of aviation to the extent that they now implicitly inform many people's thinking (in both useful and damaging ways), but they do not provide methodologies for the comprehensive prospective identification or management of actual hazards. Others, such as the FAA's 'PAVE' and 'IMSAFE' acronyms, as well as its list of 'hazardous attitudes' [38], [39], attempt to tabulate common classes of hazards that may inform brainstorming for hazard identification in certain common types of operations. These, however, are limited in scope, and do not scale meaningfully to complex operations with novel systems.

Because the effectiveness of any attempt at risk management is limited by "how well the real or perceived risks can be identified and discriminated" [21], it is essential

that the identification step be both comprehensive and detailed, regardless of the technology, personnel, or operational environment in question.

The U.S. Federal Aviation Administration (FAA) mandates risk management for certification flight testing via FAA Order 4040.26 "Aircraft Certification Service Flight Test Risk Management."[5] Currently in its third revision, the document is updated periodically as new methods and information become available. The flight test community is aware that traditional THA is increasingly inadequate as flight systems are becoming more complex, and the FAA may look to incorporate improved methods into a future revision of that Order.

Systems-Theoretic Accident Model and Processes (STAMP) is an accident causality model founded in systems theory.[40] By representing systems as 'control structures,' STAMP considers the processes occurring within a system, and the controls and feedbacks amongst them, to elucidate behaviors and hazards that can arise. Systems Theoretic Process Analysis (STPA) is a safety analysis methodology based on STAMP that is able to produce meaningful results even where other methodologies struggle, such as in systems involving software, human interactions, or other forms of complexity such as exist in aviation and flight test.

While select groups within certain large flight test organizations have adopted STAMP-based safety management techniques (though many have only applied these to flight test engineering system safety assessments, rather than to flight test hazard identification), adoption appears more advanced in military settings, and civil organizations are largely still using legacy approaches. While design organizations perform meticulous systems safety analyses, flight test organizations are not equipped or intended to go into such detail, but rather to piggyback on top of the analysis of the design organizations.

The flexibility of STPA to work coherently at different levels of abstraction provides the opportunity for flight test organizations to achieve comprehensive[3], forward-looking test hazard identification without taking on the full burden of a complete systems safety analysis at the engineering level. It also offers potential for coherent, unified mitigation activities and shorter, more concise preflight safety briefings.

This thesis aims to show why and how these additional tools may be used to good effect in the flight test context.

[3] "Comprehensive" is used in the sense of 'inclusive, wide-ranging' throughout this thesis, not in the sense of 'complete'.

1.2. Research Questions

It would appear STPA shows promise to solve key problems presently experienced in flight test. This thesis will explore whether and how this may be true. In addition to questions of theoretical compatibility, there exists the potential for both practical and systemic hurdles making implementation difficult or impractical. Exploration and understanding of these issues can help position and inform adoption, and elucidate what level of performance might be reasonably expected. Four central questions stand out in relation to the use of STPA for flight test hazard identification and management in lieu of THA, viz:

i. Is STPA compatible for use in the flight test context?
ii. What organizational restructuring or process changes are necessary to the flight test organization in order to reap the benefits of a systems theoretic approach to flight test hazard analysis in this context?
iii. What could an end-to-end application of STPA in aircraft development look like, beginning at ConOps and continuing through certification test and into line pilot training, and what hurdles would need to be solved for this to become a streamlined process?
iv. Given that aircraft design organizations have not yet widely adopted systems theoretic hazard analysis methods, can flight test organizations use STPA to more safely test new and emerging aircraft without substantially increased workload?

1.3. Research Methodology

Flight test as a discipline has been strongly practice-centric since its inception [41], and the topic of flight test safety especially remains much more in the realm of institutional knowledge and regulatory/administrative documentation than in academic publications. This thesis aims to marry advances derived from theory in academia with paradigms borne of experience in practice, to inform and hopefully improve each from the other. The material in this thesis integrates literature from both academia (where such exists) and other research organizations, as well as regulatory and topical publications, industry knowledge, formal and informal contacts in the field, methodical application in a case study, and pragmatic synthesis. This thesis works along four main avenues:

i. Review of existing processes and expectations in flight test organizations, and their limitations.
ii. Comparison of STPA and THA processes and expected products.
iii. Preliminary analysis of the flight test context to determine appropriate levels of abstraction for STPA control structures to be used for test hazard identification.
iv. Construction of a proposed example control structure that could be usefully employed as an illustration/starting point for future flight test hazard analysis using STPA.
v. Expository walk-through of the use of STPA in an example flight test campaign.
vi. Synthesis of possible organizational or process changes needed to adopt STPA where THA is currently used.

1.4. Contributions

This thesis contributes to the state of flight test safety by analyzing use of a systems theoretic approach for test hazard identification. Elements of the contribution include:

i. A cohesive academic review of flight test hazard and associated processes, challenges, context, and opportunities. Previously, aside from a few papers circulated mainly in professional circles, the topic has largely been piecemeal and confined to flight test organizational knowledge, conveyed via on-the-job training, or implied in guidance documents. (See Sections 2.1-2.5.)
ii. An analysis of the similarities and differences between THA and STPA, including their work products and integration into existing safety review process. (See Sections 3.1 and 3.3.)
iii. Exploration of the use of abstraction to allow analysis of flight test at many different useful levels. (See Section 3.2.)
iv. Illustration of the application of STPA to a representative flight test scenario. (See Section 4.)
v. Exploration of the integration of STPA into flight test in lieu of THA, and possibilities for improving hazard identification and management more broadly throughout the aircraft conception, development, and fielding lifecycle. (See Section 5.)

2. Background

> We dare not count on "luck" or "heroic skill"
> ...to assure flight safety
>
> – *Capt. Thomas Imrich,*
> *SETP PNW Section Meeting, 2019.* [42]

2.1. The Role of Flight Test in the Larger Safety Picture

Aviation safety is a multifaceted undertaking. Contributions to safety span the range of human endeavors and come at many different levels of detail. From concept sketch through decommissioning, the quest for safety touches all aspects of aerospace endeavors, and intertwines in every phase of an aircraft's lifecycle. At every step, potential dangers must be identified and mitigated, and lessons learned must be conveyed and captured appropriately so as to ultimately comprise a complete package of design, technology, processes, culture, procedures, training, and more that collectively embody safe and effective flight operations.

Frustratingly, safety is ultimately a *retrospective* measure. It is impossible to say for sure whether a system or operation is safe or unsafe until it has been actualized in practice– and even then, safety may only be disproven, never proven, as new dangers can crop up at any stage.[4] Many efforts are thus in place to cultivate and enhance safety.

A key element in the overall aerospace safety picture is the endeavor to eliminate hazard and validate operational safety prior to entry into service, so that safety may be 'baked-in' from the beginning, rather than "passing the problem on to the end user"[43] to be 'learned the hard way' through operational accidents and loss of life. This endeavor spans many steps along the way throughout design and development.

ICAO Safety Management Manual identifies three ways in which hazards may be identified: 'reactive,' 'proactive,' and 'predictive.'[44] Reactive identification involves investigation past accidents, incidents, and occurrences to learn how hazards may be prevented in the future. Proactive identification involves probing present conditions to uncover and prevent present hazards. Predictive identification involves analyzing future systems, processes, and environments to foresee and preempt hazards that may become instantiated.

In order to assure safety in operation, one must be able to identify and create conditions conducive to safety even *prospectively*. This is where hazards become a useful concept, denoting prospective, identifiable, mutable potential sources of accidents. By implementing processes, techniques, and mental models that facilitate the

[4] Essentially, in order to know the safety of a particular product in a particular use, we must first give it ample opportunity to show itself as unsafe; absent accidents or incidents after 'enough' time, we may deem it safe in an empirical sense. Colloquially, we can describe safety as 'the absence of accidents,' or more empirically as 'lots of operational success with few enough bad outcomes (or none).' But even once an aircraft and operation have 'proven' themselves safe by experience, they always have the potential to slip into un-safety (abruptly or gradually) at any time due to changes in conditions, degradation of skills or culture, aging of equipment, or other factors. This makes safety difficult to measure for use in design, even though it remains the final objective.

identification, control, and elimination of hazards prospectively, it becomes possible to introduce new equipment and operations without incurring unacceptable levels of operational risk.

In practice, some safety efforts are overt and deliberate, while others are implicit, habituated, or subconscious. Some are formal, while others are embodied in cultural or organizational norms. Some touchpoints of aviation safety are obvious, and others less so.

Safety is focal from the very beginning of the development process, where conops, configuration, conceptual design, and systems architecture begin to sketch an aerospace product. Engineering and detail design further incorporate safety considerations both implicitly through process and focus, and explicitly via hazard analysis methodologies and requirements tracking. Once a design takes shape, campaigns of bench testing, ground testing, and flight testing typically aim to probe, demonstrate, and validate behaviors, performance, and safety. Flight test is a key process in the introduction of such new products and operations. Certification or acceptance flight testing demonstrates compliance (or non-compliance) with regulatory or other and requirements, and gates entry into service.

By the time an aircraft is ready for developmental testing and later certification/approval testing, an extensive safety analysis package will have been prepared and vetted. This package is typically available to the flight test organization to familiarize and gain trust in the aircraft; to determine key test points needed to prove the safety of the design; and to discover and evaluate hazards that may be revealed in testing. The test organization may partly validate this package 'on the ground' by review, but sooner or later it must be demonstrated in the air by actual flight.

Many techniques are used in the design and safety engineering phase, including those that are bottom-up, top-down, probabilistic, functional, and special methods. These include Functional Hazard Analysis (FHA)[5], Fault Tree Analysis (FTA), Failure Mode and Effects Analysis (FMEA) / Failure Mode, Effect, and Criticality Analysis (FMECA), Common Cause Analysis, Design appraisal, Installation appraisal, Zonal Safety Analysis, Particular Risk Analysis, Common Mode Analysis, Aircraft Safety Assessment, Operational & Support Hazard Analysis, Health Hazard Analysis, System Hazard Analysis, and Safety Assessment Reports.[45], [46], [47] An increasing number of organizations have begun to investigate the use of Systems Theoretic Process Analysis (STPA) for design as well.

Together with design, test, and certification, parallel efforts of procedure design, crew training, operational planning, organizational culture, and review and reporting processes informed by the former help to shape safe operations.

Once a product goes into operation, various tracking, reporting, and sustainment/improvement efforts continue to sustain or improve levels of safety. These include not only Safety Management Systems, safety audits, and inspections by certification and oversight organizations such as the FAA, but also maintenance, Airworthiness Directives, and other broad-reaching monitoring and improvement programs.

While systems safety engineers complete extensive safety analyses of a system during design (and prior to the start of testing), flight test departments may desire not

[5] Also known as "Function Hazard Assessment"

only to manage test-related hazards, but also to verify and validate the results of these engineering safety analyses in the full broader context. In flight test, a new, modified, or repaired system undergoes an ultimate hurdle of validation: operation in the real world– probed, scrutinized, and pushed to its limits– to characterize and validate its behaviors. "[F]light test may be the first time that the entire system is looked at together. It is therefore essential that the flight test team look at the provided system risks [from engineering] and ensure that they make sense and there are no obvious misses."[45]

In addition to measuring performance and other characteristics, a core responsibility of flight test is to validate the safety of a new aircraft or other system under test, and reveal any negative tendencies before they may materialize in operational environments. This itself incurs risk. "Engineering design deficiencies are not uncommon in project equipment; therefore, a hazard analysis and risk assessment is required [to] systematically determine possible hazards and minimize surprises."[48]

In flight test, where safety guardrails normally embodied in operational environments may be absent or deliberately violated, safety management is a core focal point of deliberate activity. In "the realm of operational aviation[, the] aircraft configuration is fixed, the aircraft performance is known, the training regime for the crew is defined, and the route network well practiced. That leaves weather as the variable. But for flight test, the configuration and system performance are the variable under assessment."[49]

With more unknown and less yet proven, the risk involved in flight testing can be inherently high. Beyond probing the unknown, flight test also often purposely explores worst-case and off-nominal design conditions, thereby cutting into safety margins that may have been designed into the system under nominal operating conditions. Therefore, flight test risk is generally expected to be higher than the risk in normal operations (which have had the benefit of de-risking through test as well as all the safety margins in design).

Risk tends to be evaluated ordinally rather than absolutely, and flight test risk is often rated on a shifted scale, as illustrated in Figure 3, where "Low risk level in Flight Test is significantly more risky than low risk in standard aviation operations."[50]

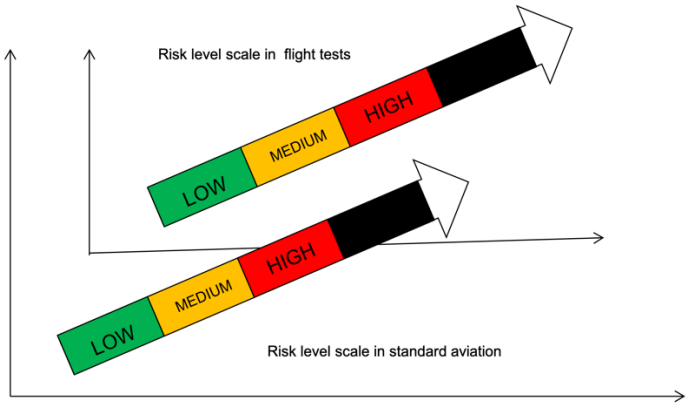


Figure 3: Risk level scale comparison [50]

Because risk is seen as stemming from the unknown (or yet unproven), the object of flight test is to "reveal vehicle characteristics in a controlled manner so that discovery of aircraft shortcomings does not lead to catastrophe."[51] In order that test risk isn't

unbearably high, flight testers use special additional mitigations to keep themselves safe while exploring novelty and hazard that would be unacceptable in operational contexts. Catastrophe can spell the end of a program.[6] "A well-designed test program will move as many hazards and outcomes as possible from the unknown into the known,"[43] and do so safely.

In doing so, flight test takes on risk so that operational organizations don't end up saddled with it. It is expected that certain unexpected hazards may surface during flight testing, and regulations specifically require these to be properly examined and documented, so that their risk is not passed on to the end user. For instance, 14 CFR 23.1309(b) (amdt. 23-63) specifies that "Minor, major, hazardous, or catastrophic failure condition(s), which occur during Type Inspection Authorization or FAA flight-certification testing, must have root cause analysis and corrective action." Analogous requirements are stated elsewhere for other categories of aircraft.

The promotion of flight test safety, of course, is balanced against budgetary constraints. If flight test spends too much time and money mitigating risk, they can drive a program or company to bankruptcy. Figure 4 illustrates a conceptual 'safety space,' in which safety must be successfully managed to avoid catastrophe on the one side, and bankruptcy on the other.

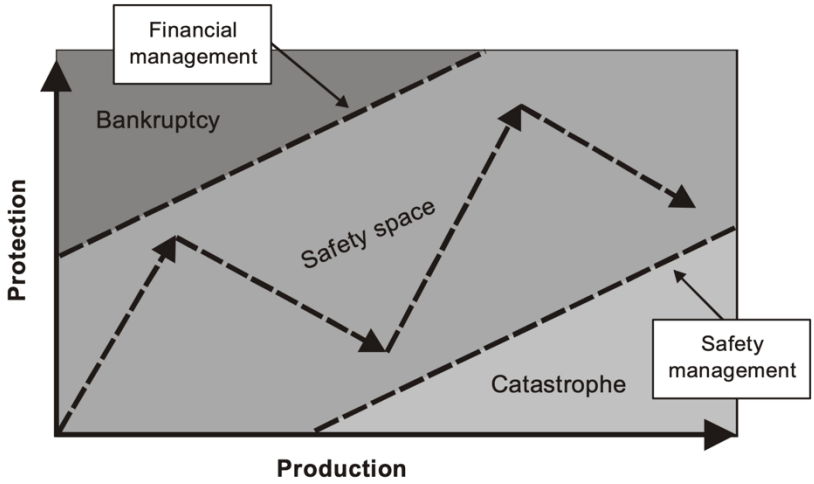


Figure 4: The 'safety space' [44]

Flight test– and especially development and acceptance flight testing–

"…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 [programmatic] 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." [53]

[6] "If you think safety is expensive, try an accident!"[52]

The reduction in risk that occurs in the course of a test program reflects precisely the risk that is taken on by the program and reduced (or 'bought down') in the course of it.

Flight test can be useful at many different stages of an aircraft's lifecycle. From experimental flight test, to development flight test to debug and perfect a new design, on to certification or acceptance flight test, operational flight test, return to service flight test, and more, various levels of risk may be acceptable, and various safety management processes may be desirable. In all, an emphasis prevails on preventing casualties and losses, and on passing correct, useful, and actionable safety information back to engineering and design for improvement and forward down the line to those who will operate the aircraft next, in order to reduce their risk.

Experimental flight testing is sometimes used to reduce technical risks, by testing key aspects of design early in the program, and in controlled manners that mitigate some of the hazard involved. This can expose hazards that might not be tolerable later in the program. Later in the aircraft lifecycle, operational testing, post-modification testing, and post-maintenance testing may each assume less risk than in the early stages, since the aircraft will have been better validated and explored, but still assume higher levels of risk than in regular operations. Between the two lies developmental flight test, at an intermediate level of flight test risk.

In order to navigate the 'safety space' successfully, flight test organizations must be able to keenly determine which risks exist, and to dispatch them efficiently according to their value and gravity.

2.2. Hazard-Informed Flight Test Management Processes

Unlike in the early days of flight test, where a daredevil 'kick the tires and light the fires' approach was common , flight test risk management nowadays is a methodical, carefully managed core competency of flight test organizations worldwide.[7] Aiming to balance learning against risk, the flight test discipline endeavors to take on only those "deliberately accepted risks"[1] that appear to be worth their prospective rewards in data and insight.[54] "[M]anagement of safety in a systematic and proactive way enables… [organizations to address] potential hazards and associated risks before aviation accidents [can] occur."[55] Once the hazards and risks are understood, "a balance must be struck between the "worth" of [the test] objectives and the "risks" involved in achieving them"[3].

To this end, deliberate risk management initiatives are in place in most established test organizations, and are often also mandated by regulatory authorities. Figure 5 depicts a general risk management process. In this process, risks are identified; then analyzed and often classified or prioritized; plans are then formulated for handling the various risks in question; and the process is tracked and controlled as it progresses.

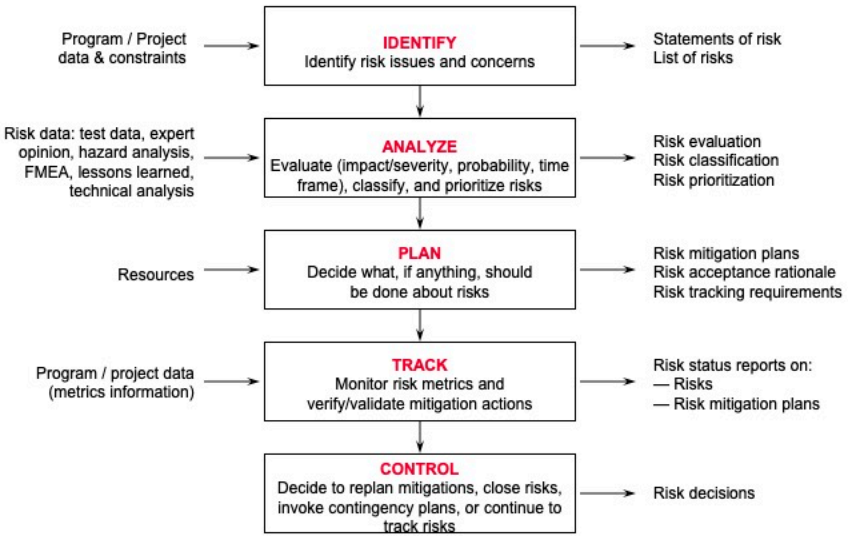


Figure 5: Risk Management Process outline [23]

Flight test risk management processes are typically informed and substantiated through a formal hazard analysis. "[T]he object of the hazard analysis is to first identify hazards that could occur, identify the cause of the hazard, determine the probability of that hazard occurring[7], assess the risk should the hazard be encountered, and establish precautionary measures to eliminate or reduce the hazard."[3]

A completed hazard analysis serves as a core safety artifact of the test program, referenced across safety management and risk-informed decision-making in program management and execution. It becomes "the core risk management tool that informs SRBs [Safety Review Boards], FFRRs [First Flight Readiness Reviews], etc. [Test Readiness Reviews (TRRs)]."[8]

Figure 6 shows the risk management process outlined in AFTCI 91-202 to guide Air Force Test Center (AFTC) organizations, which relies on a complete hazard analysis for review and decision-making. Figure 7 shows the hazard-informed process used by Cessna to manage all testing of the Model 680 'Citation Sovereign' business jet.[8]

[7] In practice it is notoriously difficult to determine probabilities of hazards occurring that are not due to statistically characterized failures of standard components.

[8] Here they use the term 'Test Hazard Assessment'

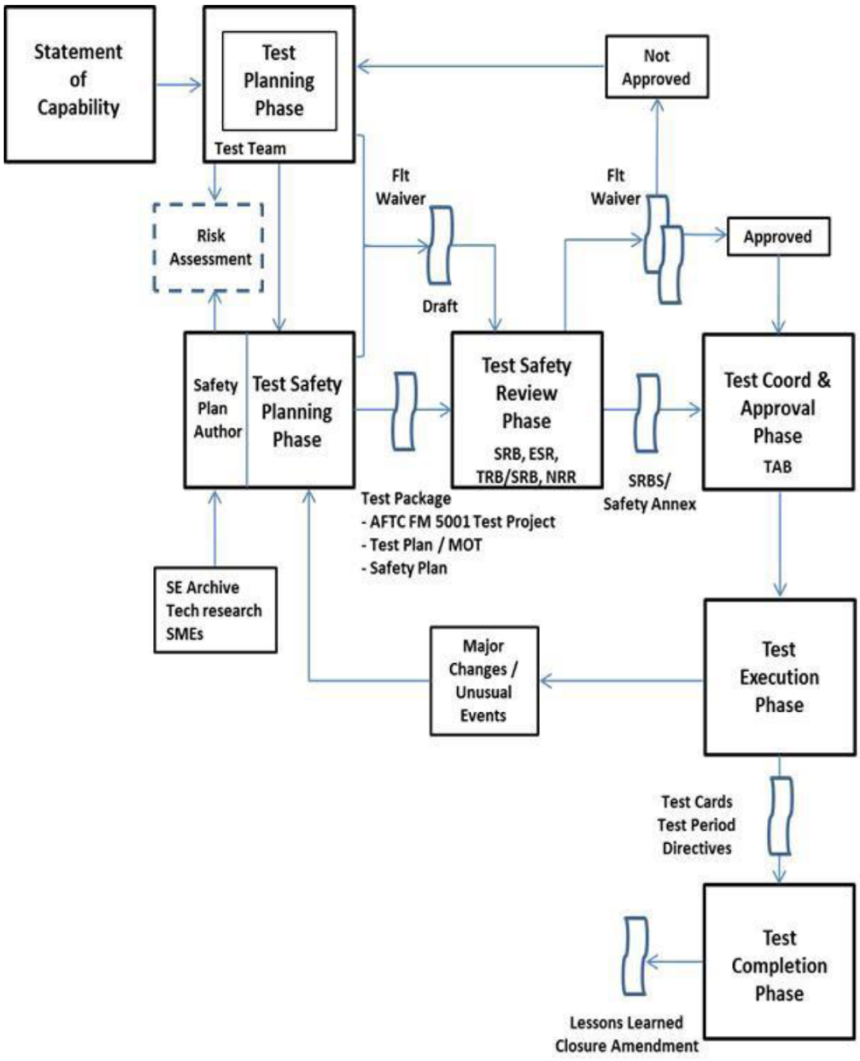


Figure 6: AFTCI 91-202 Process Flow[9] [56]

[9] Definitions of acronyms and abbreviations used in the figure [56]:

Flt : flight	SE : Safety Office
SRB : Safety Review Board	SRBS : Safety Review Board Summary
TRB : Technical Review Board	ESR : Electronic Safety Review
NRR : Negligible Risk Review	SMEs : Subject Matter Experts
AFTC FM : Air Force Test Center Form	MOT : Method of Test
TAB : Test Approval Brief	

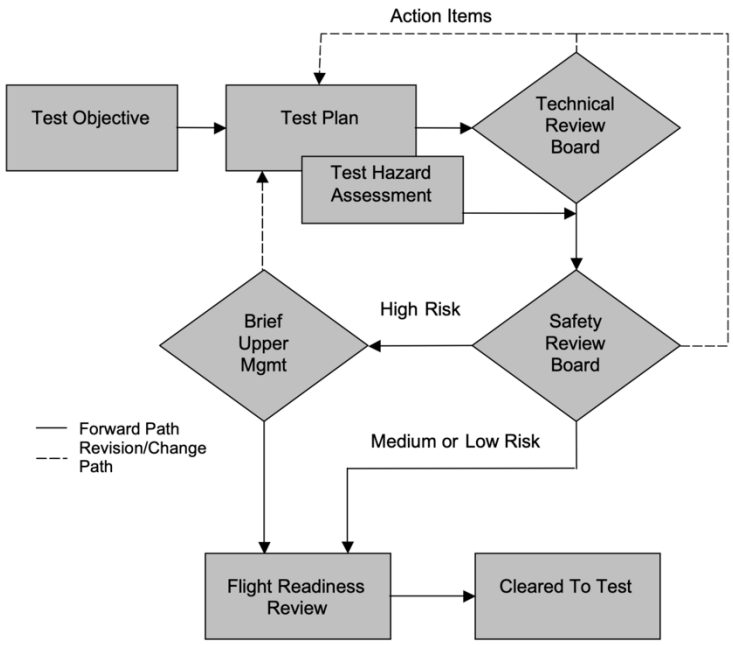


Figure 7: Cessna Citation Sovereign Test Planning Review and Approval Process [57]

The Australian Civil Aviation Safety Authority describes flight test risk management generally as a process in which:

- *hazards are identified;*
- *an assessment is made of the risks involved;*
- *mitigating procedures are established to reduce or eliminate the risks; and*
- *a conscious decision is made, at the appropriate level of authority, to accept residual risk. [58]*

Boeing Test and Evaluation UK suggests more formally "[t]he identification and subsequent management of Risks follows a simple step by step process" of:

- *Risk Identification;*
- *Risk Assessment;*
- *Risk Control and Decision Making;*
- *Implement controls; and*
- *Supervise/Revise.* [7]

FAA Order 8040.4C provides guidance outlining such a Safety Risk Management Process [59], and FAA Order 4040.26 institutionalizes the use of such a flight test safety risk management processes in certification flight testing [5], [60]. Air Force Test Center Instruction (AFTCI) 91-202 likewise outlines analogous elements in its Test Safety Review Policy [56], as does MIL-STD-882E for Department of Defense (DoD) Standard Practice: System Safety[10].[61]

Preparatory risk management has traditionally been undertaken via the technique of Test Hazard Analysis (THA), which draws heavily from reactive identification, and is supplemented by predictive identification in the form of brainstorming. When new or novel systems are at play, flight test hazard identification attempts to focus more

[10] See Task 303.

intensely around predictive identification, where the capability of THA may be further diminished.

It is worth noting that the foregoing describes preparatory hazard analysis undertaken before test, which is the subject of this thesis. Such will not be the only hazard mitigation at play during flight test, though it may be the most methodical and specific such activity. As flight test works to probe the safety of new equipment and operations, while mitigating against hazards that may result, it is a goal of flight test that in addition to those hazards identified by analysis, even those hazards not identified in advance be revealed and handled safely prior to operational use. Thus, flight test organizations employ reactive, proactive, and predictive forms of hazard identification. Techniques such as buildup sequences, data and trend monitoring, test continuance criteria, and general mitigating measures implement additional measures of hazard identification and mitigation, as do preflight briefings and test pilot training and decision-making.

THA is also a core element in flight test mission preflight briefings, covering the identified hazards that are expected to be relevant to each particular flight card.[7]

THA is further described in Section 2.3.

2.3. Flight Test Hazard Analysis: Review of Current Practices

To identify (or 'discover') relevant risks, the hazards associated with the systems to be tested and the testing to be carried out must be understood. "The object of risk discovery is to accurately determine 'What can go wrong?' [i.e., what hazards could occur]. … it is the most critical part of the risk management process and is the basis of the mitigation strategies"[21]. Once the hazards have been identified, then their potential impacts may be evaluated, and plans may be made to mitigate their occurrence or their adverse outcomes, and thereby prevent or reduce any resultant losses.

Test Hazard Analysis (THA)

Test Hazard Analysis (THA)[11] is the core hazard analysis process presently employed by flight test organizations worldwide. THA identifies hazards associated with test, analyzes these for cause and effects, and identifies precautionary measures that may be used to mitigate or eliminate them, as well as corrective actions that may be taken to mitigate negative outcomes should a hazard occur. Risk is assessed for each hazard based on severity and likelihood estimates (usually qualitative), typically both before mitigation ('unmitigated') and after mitigation ('residual'), and the risk levels are categorized and often color-coded[12]. Overall test plan risk is then categorized based on the individual risks. [7], [51] "Test teams use system safety techniques, prior experience, legacy system research, lessons learned and overall engineering judgment to identify test unique hazards and assess risk by evaluating the credible outcome (mishap severity) of each hazard together with the associated probability of occurrence."[56] Figure 8 describes elements in a THA.

[11] Sometimes also called Test Safety Hazard Analysis (TSHA) [2] or "Trials Risk Assessment" [7].

[12] Risk categorization typically follows a 'risk matrix' type approach based on estimated hazard likelihood and consequence, such as [7], [39], [51], [61], [62].

a. Hazard – the front end of the test safety process is identifying those hazards that are unique to a particular flight test activity.
b. Cause(s) – in order to reduce the probability of suffering the hazard, the potential causes of that specific hazard must first be identified.
c. Effect(s) – the potential effect of the hazard need to be understood in order to assess the risk level.
d. Mitigation(s) or Minimizing Procedure(s) – this is the part of the process wherein procedures are established to either eliminate the cause or reduce its probability of occurring, or to reduce the effect of the hazard should it occur.
e. Emergency Procedure(s) – procedures planned in the event that the mitigations do not completely succeed.
f. Residual Risk – once all of the safety planning is in place, this section estimates the level of risk that remains, which can then be used to establish the appropriate level of approval for that particular flight test activity.

Figure 8: Elements of THA described [60]

Each hazard identified in a THA is typically documented with its causes, effects, precautionary measures, corrective actions, and risk category (pre-mitigation, post-mitigation, or both), and documented hazards are collected into binders and sorted by test program, sortie, test maneuver/technique, or test card. Figure 9 shows an example hazard document for a "Engine Temperature Exceedance" hazard.

Hazard – Engine Temperature Exceedence

Cause:
1. Sustained high power operation.
2. Poor ventilation or heat exchanger performance.
3. Engine exhaust gas ingestion into cooling inlets.
4. Engine exhaust gas backflow into compartment.

Effect:
1. Degraded system operation, leading to possible system failure.
2. Potential for fire.

Precautionary Measures
1. Ambient temperatures during tests will have margin below maximum design operating temperatures.
2. Propulsion Flight Test Engineer(s) (FTE) to perform real-time telemetry monitoring of engine performance and engine surface and bay environment temperatures.
3. Progressive envelope expansion.
4. Temperature corrections used for predictions of exceedences.
5. Fire detectors in engine bays and APU fire zone.

Corrective Action(s):
1. Reduce power demand
2. Terminate maneuver and return to previously cleared flight condition.
3. If fire, execute ENIGNE FIRE Emergency Procedure as defined in NATOPS

Residual Hazard Level/Risk Category:
III/C = Category B Test Hazard

Figure 9: Example THA: Engine Temperature Exceedance Hazard [51]

In the interest of efficiency, THA typically aims to address only those hazards that are "directly associated with the testing (test-specific). 'Generic' hazards associated with normal operation of the aircraft or test equipment should not be included."[7] Hazards

thus directly associated with testing (i.e. the "test unique hazards"[63]) include those "uniquely introduced or aggravated"[6] by the "flight test activity [or] specific flight tests"[55], but excludes "general hazards (also referred to as routine hazards)"[63] which are present in flight operations regardless of whether or not testing is undertaken and are not elevated by the testing. The US Air Force Research Laboratory clarifies test unique hazards specifically to include those hazards that are:

- *Not present in normal operation of equipment/system/vehicle*
- *Associated with initial testing of [a] new system*
- *Exacerbated by test [or]*
- *Introduced by testing being conducted*. ***[63]***

As an example, "a specific flight test may introduce specific risks associated with that particular flight test. For example, a minimum control speed test introduces a risk of loss of directional control leading to loss of the aircraft. This risk is only valid for flights with that activity,"[55], whereas "a bird strike hazard would normally be considered nontest unique unless there is something in the test plan that elevates the risk of bird strike above that occurring during normal flight operations."[64] In other words,

> *A hazard associated with the normal operation of the aircraft, vehicle, SUT [System Under Test], or facility is not a test unique hazard. A hazard ordinarily encountered in a typical activity is also not a test unique hazard (e.g., sunburn while working outside). But some test activities may elevate the risk associated with normal operational hazards. For example, midair collision with non-participating aircraft and bird strikes are not generally considered test unique hazards. However should the very nature of the test increase the exposure to these hazards above that of normal operations, they should be addressed as test unique hazards. Hazards associated with the initial testing of a new system should also be addressed as test unique hazards since normal operations for this system have not been established. [56]*

US Navy guidance indicates "[w]hen defining potential hazards, consideration should be given to the specific test item, the test maneuvers and flight conditions planned, and the environment in which the test will be conducted," including workload due to testing [7]. Hazards associated with the specialized use of test equipment are also included.

Figure 10 illustrates THA analysis of a representative test hazard due to a test maneuver or flight profile (specifically in this case, lateral drift during vertical landing). Figure 11 illustrates analysis of a representative test hazard related to project test equipment (asymmetric dumping from a flight test ballast tank system).

Hazard	Cause	Precautionary Measures	Probability of Occurrence (POC)	Risk Assessment (RA)
Lateral drift on VL causing excessive sideloads on outrigger landing gear	• crosswind/turbulence • jet exhaust impingement on wing due to bank angle when close to deck • insufficient lateral control authority due to airplane dynamics • excessive lateral weight asymmetries • poor pilot scan	• buildup in crosswind, lateral asymmetry • no excessive lateral inputs when close to deck • LSO will waveoff any unsafe approach • FCLP will be accomplished for pilot proficiency prior to detachment	Low	Mod

Legend:
Low POC: The hazard is unlikely to occur.
Moderate POC: The hazard may occur.
High POC: The hazard is likely to occur.
Low RA: Minor damage to aircraft is likely to occur.
Moderate RA: Moderate damage to aircraft is likely to occur.
High RA: Loss of aircraft and aircrew is likely to occur.

Figure 10: Example flight profile-related hazard analysis [3]

Hazard	Cause	Precautionary Measures	Probability of Occurrence (POC)	Risk Assessment (RA)
Asymmetric dumping of water tanks creating asymmetric moment beyond airborne limits resulting in aircraft impact with water	For asymmetric dump: • faulty dump valve • faulty wiring • "popped" circuit breaker for one tank and not the other (they are on separate circuit breakers) For not being able to secure asymmetric dump: • after detecting asymmetric dump, power is lost to the "good" tank which would leave the valve in its position at power loss (open) • after detecting asymmetric dump, valve fails (sticks open) on the "good" tank	• Valves are approximately 2 years old and have been reworked for this project. • All electrical wiring in tanks have been replaced with new wiring • LSO will visually monitor water tank dump after each launch and confirm "both tanks dumping" via radio. If asymmetric dump is detected, LSO will declare "secure dump, secure dump" via radio. If asymmetric dump detected, attempt to secure dump immediately and BINGO to shore. If asymmetric dump continues, jettison tanks. • Jettison circuits will be checked daily • Fresh water vice salt water will be used in the tanks to aid in preventing corrosion.	Low	High

Legend:
Low POC: The hazard is unlikely to occur.
Moderate POC: The hazard may occur.
High POC: The hazard is likely to occur.
Low RA: Minor damage to aircraft is likely to occur.
Moderate RA: Moderate damage to aircraft is likely to occur.
High RA: Loss of aircraft and aircrew is likely to occur.

Figure 11: Example project equipment-related hazard analysis [3]

The hazard identification step of THA aims to capture all test unique hazards involved in an intended test. Hazards are identified via brainstorming. Personnel are chosen to participate in the brainstorming based on their prior experience, their perspective, and the familiarity and lessons learned that they can be expected to bring to the table. Thus, the THA is "informed by the experience and best practices of senior test personnel"[17] in the hopes that as many relevant hazards as possible can be identified.

In the interest of identifying as broad a range of hazards as possible, the hazard identification step typically opens with a brainstorming session "in which every imaginable test hazard is proposed"[6] by participants "from various backgrounds and

experience levels."[65] "This exploratory step must include some ludicrous "what-if's" and compound failures[13] that might initially breech the bounds of plausibility[14]."[6] Both the experiences of a seasoned engineer who has seen many unexpected hazards emerge in testing programs, and the wide-eyed view of an inexperienced team member who has not been biased by group think should be present."[65] "A pilot with first-flight experience is invaluable."[66] Together, these capture both seasoned experience and a level of creativity in hazard identification that may not be present otherwise.

In at least one instance where inexperienced members were part of the test team, a workshop was organized to review previous experiences of the company, and highlighting "what had worked well but also lessons learned from previous tests and how this could be avoided during the upcoming flight tests."[67] Many practice runs with previously-proven revisions of the aircraft were run, as well as simulations, to develop experience in the new team members.

Depending on their capabilities and staffing, test organizations assemble (as described by SAAB AB) "a group of experienced personnel from most involved flight test disciplines, including maintenance. . . .to ensure that… all consideration [is] taken to flight safety."[7] This may include (as described by Boeing Test and Evaluation UK) "[t]est specialists including Test Engineers, Test Aircrew, Safety Engineers and other subject matter experts help to identify the hazards associated with the operating of a test asset during specific test manoeuvres/activities."[7]

In the hopes of inspiring good coverage of hazard identification, many organizations further review the results of the brainstorming process with third parties and senior management personnel. As described by Boeing Test and Evaluation UK, "[t]he THA are subject to peer and independent review at an appropriate level to the risk of the T&E [Test & Evaluation] activity."[7] Some organizations specify particular personnel who must take part in brainstorming, or must fill a reviewer role, such as by staffing the Safety Review Board.[15] The SRB or equivalent committee acts to provide a

---

[13] Note that this 'compound failures' view of complex hazards is given far clearer focus via systems theory, as will be seen later.

[14] Hazards that are not test unique will later be discarded, but in order not to stifle thorough brainstorming, no judgement or filter should be imposed during the initial process of imagining what could go wrong.

[15] NASA Dryden included in an SRB "members of each of the disciplines involved (aerodynamics, stability and control, performance, etc.) as well as subsystem experts, instrumentation experts, and a pilot."[66]
The Canadian Forces Aerospace Engineering Test Establishment specifies the following personnel as mandatory SRB attendees: "a) FTA (Chair); b) Senior Test Pilot (STP); c) Senior Test Engineer (STE); d) Senior Design Engineer (SDE); e) Unit Safety Officer (USO); f) PO; and g) At least one member of the flight test crew. . . .If available and applicable, the SRB should also be attended by: a) Test Plan Approval Authority; b) Project Pilot (PP); c) Test Director; d) PE; d) Experimental Aircraft Maintenance Engineering Officer (XAMEO) and aircraft technician representatives, if project safety issues impact aircraft maintenance, technical issues affect test safety or test procedures create new hazards to aircraft maintenance or servicing activities. If testing is to be conducted at a deployed location, a maintenance representative from the host unit should attend the SRB."[7]
Textron Aviation specifies "Four primary participants
• Flightworthiness Authority: Senior Vice-President Engineering
• Engineering Authority: Project Engineer
• Aircraft Maintenance Authority: Experimental Engineering Manager
• Flight Test Authority: Director, Engineering and Defense Flight Test"[54]

last check of the completeness of hazard coverage, the analysis and treatment of those hazards identified, and the acceptance of any residual risk.

In order to stimulate comprehensiveness in hazard identification, flight test organizations also employ a variety of prompting and templating strategies to bolster and supplement their brainstorming. Templating employs copy-pasting and then editing of prior THAs, as well as review of databases of existing THAs, to give a 'head start' on brainstorming, incorporating those deemed to apply, striking those that do not, and modifying others to suit the specific testing in question. Prompting utilizes checklists and other collections of categories, ideas, hazard types, and sources of reference as a way of "subtly steering the hazard identification process along the many dimensions of the operation and possible kinds of hazards"[19]. One professional summarizes three core steps to "identifying the list of hazards that go into a THA," plus a final reflection step:

1. *The bottoms up review of the test plan*
    1. *Focus on the plan, the system under test, the environment, and other unique factors that may present hazards in the test.*
    2. *Some hazards are the result of aircraft (im)maturity, and some are the result of exposure, i.e., danger inherent in a maneuver. (The former hazards may not present as much risk as the latter after development testing is complete).*
    3. *Other Resources*
2. *Your experience*
    1. *The experience of your teammates and/or organization (similar tests on other projects/programs)*
    2. *Industry Guidance (FAA Order 4040.26, US Navy/US Air Force Equivalent)*
    3. *Other industry specialists*
3. *THA databases (NASA/Flight Test Safety)*
    1. *Brainstorming*

Step back and think: what else can happen that is unique to this test? [28]

He further describes the following "non-exhaustive list of things to consider when developing a THA for something new (or just new to you):"

1. *Start with what is known & collect past hazards from like systems*
    a. *Utilize the NASA database (https://ftsdb.grc.nasa.gov/)*
    b. *Find and review old test plans*
    c. *Do your own research (is anything ever "new?")*
2. *Talk to experts*
    a. *In the field*
    b. *not in field but familiar with the technology*
3. *Work with system owners, system safety, operators and those doing simulation and lab test*
    a. *Component failures*
    b. *Operations issues – location, atmospherics, test team, ops tempo, etc.*
    c. *Systems Interactions*
    d. *Human Factors*
    e. *Wade through the alphabet soup of failure scenarios*

   f. *Review new technologies and how they perform in the aerospace environment*
      i. *Vibration, temperature, etc.*
   g. *Look very closely at safety enhancing systems – especially automatic ones!*
      i. *Do they impede our capabilities when activated?*
      ii. *Do we understand the failure modes of the added complexity?*
      iii. *Do they fail safe or unsafe?*
      iv. *Will the safety systems add operations complacency?*
4. *Review your flight test techniques*
   a. *Are they new or different from classical*
   b. *Are they appropriate to the aircraft*
5. *Consider New operational considerations*
   a. *VTOL takeoff in with wind*
6. *New control types or methodologies*
   a. *Different inceptor types*
   b. *Stick and throttle vs. touch screens*
   c. *Rotor vs. other powered lift*
7. *Brainstorm*
   a. *Prognostic fishbone diagrams*
   b. *Be curious – think like the system*
8. *What else can we be missing?*
9. Let the hazards and risk evolve with the program
10. Be honest with yourself – urge the team to think honestly and objectively about their risks
11. Re-focus! There are general risks out there that apply broadly, your application has to make *it more likely or it doesn't qualify*
    a. *Example: Bird strike - not applicable to general flying*
    b. *However, if you have a test that is low and fast, then it may* *[45]*

The US Navy draws from "test team discussions, conducting fault tree analysis, reviewing historical data, and reviewing hazard analysis and flight test lessons learned databases maintained by US Navy, US Air Force, and Society of Experimental Test Pilots"[7], [48], among other avenues, when identifying relevant hazards. They suggest "[interfacing] with other engineering competencies as appropriate for identification and mitigation of any engineering hazards."[48]

The US Air Force Test Center suggests consulting:

- *Archived test and safety plans, to include lessons learned and THAs, across the enterprise for consideration of similar tests.*
- *Personnel or test teams with experience in similar test activities or testing.*
- *Technical libraries, internet, etc. to research technical aspects.*
- *System safety hazard analyses of the test article and test facility.*
- *Applicable safety reviews from other organizations such as the Program Office, Nonnuclear Munitions Safety Board, Directed Energy Safety Board, 711 Human Performance Wing Independent Review Board or the contractor.*
- *Aircraft modification documents*. *[56]*

Textron Aviation looks to "Test plans from previous programs," "Team members and other EDFT [Engineering Development Flight Test] experts," "Engineering,"

"CPSST [Chief Pilot for Safety, Standardization and Training]," "External sources [including] SETP / FTSW / NASA," and "Standard THA Library [an internal company database of THA's accumulated from experience]."[54]

The Turkish Air Force draws information from:

1) *Similar tests conducted before,*
2) *Lessons learned from experience,*
3) *Results from ground testing,*
4) *Data from Engineering analysis, (CFD [Computational Fluid Dynamics], FEA [Finite Element Analysis], etc.),*
5) *Modelling and Simulations, etc.* [7]

At the Royal Netherlands Aerospace Centre (formerly known as the National Aerospace Laboratory, NLR), "[h]azards are collected from different sources (internal and external hazard databases, expert sessions, lessons learned, Best Practices from SETP [Society of Experimental Test Pilots] and SFTE [Society of Flight Test Engineers]) and combined into scenarios."[6]

Various organizations internationally, including some whose testing focuses more on return-to-service or Functional Check Flight (FCF) type flying, rely more or less on test pilot judgement and experience rather than formal Test Hazard Analysis and review processes.

Additional Methodologies Sometimes Used

Occasionally organizations have supplemented or replaced THA with modified or alternative processes. In efforts to facilitate more structured, graphical contemplation of hazards, and to better expose and communicate a complete safety case, groups have experimented with the Bowtie method used in oil and gas industries [7], and a 'visual THA' process derived from it [12]. Both graphically emphasize the elements of 'prevention,' 'recognition,' and 'recovery' from hazards. However, Bowtie implies a restrictive linear causality model and was found to be cumbersome when attempting to chain causes and effects into a complete description of the system.[12]

A NATO compendium of flight test risk management best practices ([7]) identified no specific tools developed for managing the flight test safety process, but rather only "generic" tools like Bow Tie and Hazard Analysis. Other tools, however, have been used occasionally.

Prognostic Fishbone Diagrams offer something of a "combined FHA and FTA but are taken more from the flight testers point of view."[45] They aim to help with "finding 2nd order issues with interactions" that are not readily found by standard THA. It is one more tool available to increase the number of scenarios, "that occur to you."[45]

Comparison of Initial STPA Experiences with THA

Some flight test organizations have experimented with STPA to various extents. Their experiences and comparisons to THA provide insight into the benefits, drawbacks, and obstacles that may present in transitioning from THA to STPA at an organizational level.

The US Air Force 412th Test Wing was one of four wings that explored the use of STPA in lieu of traditional THA.[68] Projects selected for analysis by STPA were mostly not new capabilities or complex systems, but rather programs representative of

established work (software block upgrades, etc.), however a novel hypersonic application was also examined. They found STPA produced no less hazard coverage than traditional THA, but at the expense of additional investment of time: "Overall, [traditional THA] produced similar results in less time, but they speculate whether this was due to familiarity with the test and plan to continue experimenting with [STPA]."[69] As of presentation in 2018, they found "STPA is not a form, fit, function replacement for the current 412th [Test Wing] safety process," but that aspects of traceability and the graphical system representation were valuable assets.[68]

They noted STPA does not provide the 'corrective actions' nor the risk assessments that are part of a THA; both of these must be accomplished via existing or other methods. They also noted their application of STPA was hindered and time-consuming without access to appropriate Subject Matter Experts (SMEs).

They found the traceability and insight provided by STPA facilitated superb accident mitigations. And the graphical system depiction at the core of STPA was found to be extremely helpful in understanding, communicating, and collaborating around system behaviors and hazards. They found they could perform STPA without needing to have a 'completed' test plan in hand. And they found STPA produced system design mitigations in addition to test hazard mitigations. The test center commander "approved the center to gradually implement STPA (where appropriate) across its wings/complexes based on [their] combined findings."[69]

The 412th Test Wing suggested that although they "[d]id not identify [a] large gap in traditional test safety planning methods" on the example cases, "STPA could be very powerful and useful, IF performed early enough to influence system design and test requirements" (emphasis removed), and "[m]ay be useful for projects that are not well understood."[68]

STPA has also been applied to testing of at least two unmanned aerial systems (UASs)[70], [71], and both identified underlying system design hazards. It appears that although both were undertaken as flight test hazard analysis activities, the majority of effort and results ended up contributing to design rather than testing endeavors.

Boeing used Systems Theoretic Process Analysis (STPA) to analyze hazards in integrating Automatic Test Maneuver equipment into certification flight testing. They observed:

> *Conducting a through safety review is almost taken for granted to the flight test community. However the structure of the STPA process drives a level of rigor, and in turn confidence, that is unachievable with normal THA style safety reviews. . . . The use of STPA was a key enabler for gaining approvals, as it allowed the team to speak with confidence and prove thorough coverage [of (sic)] all aspects of the safety of the system while also simultaneously having a high degree of confidence in being able to provide increased value to the program."[72]*

Because much of the team was new to using STPA, two facilitators helped to guide and process, and this was found to be "critical to the success of this analysis."[73] At several points the analysis team realized their solutions relied "on discipline and training [of the crews]. Instead, through STPA, the requirements drove simple engineering solutions to reduce or eliminate that reliance."[73] In this case, STPA was used to examine the safety of integrating the automatic test maneuver equipment into flight test, which can be thought of as a step before where THA might normally be used. Still, the analysis

identified 27 new requirements, including 5 covered by preexisting system limitations (no action required), 8 that drove updates to their software, and 14 that motivated amendments directly to test plan procedures or checklist items.[73]

It is likely that several additional organizations or programs unknown to the author at present have also used STPA in flight test, whether to complement, augment, or replace THA. Generally, such efforts appear to be in an experimentally-driven exploratory phase wherein practical applications are undertaken on the promise of prospective gains, and lessons-learned captured for future implementation– this as opposed to an analytical effort to methodically fill specific gaps in current practice using the strengths of tools that have been developed based on advances in theory. One group of researchers suggests that "[t]he frameworks that will enable advanced test and evaluation of tomorrow's complex systems are in their infancy and are just becoming visible outside of academia."[49]

Hazard-Based Risk Assessment

Many flight test risk management processes utilize hazard-based risk assessments. THA incorporates both hazard-specific and overall risk assessments, but any risk assessment methodology may be used. Most commonly, a Risk Matrix framework [61] is used, based on hazard probability and consequence. As illustrated in Figure 12, the risk of an identified hazard is rated as a function of predicted probability of occurrence and expected severity of outcome. The more severe the expected outcome (e.g., 'catastrophic') and the more likely its predicted rate of occurrence (e.g., 'frequent'), the more severe the risk. Those that are less severe or less likely are deemed less risky.

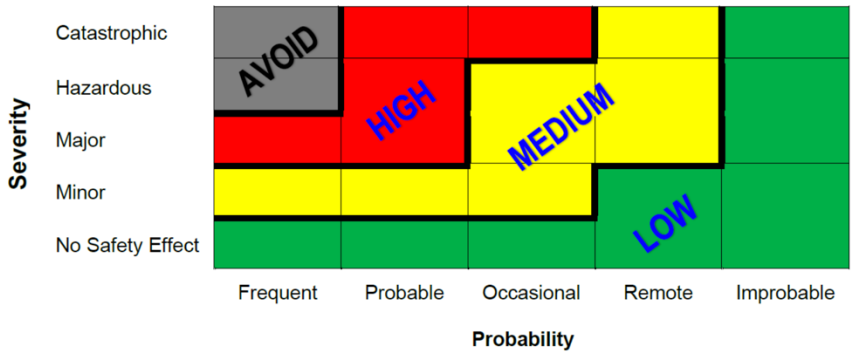


Figure 12: Representative 2-D Risk Matrix [74]

While the risk matrix of Figure 12 conveys actionable judgement on the risk as well (i.e., "avoid"), not all risk matrices do; often the risk matrix merely rates the risk and a separate table of criteria is used to determine acceptability or actionability. ICAO Safety Management Manual, Chapter 2, Appendix 2 offers severity and likelihood ranking schemes, a standard non-actionable 2-D risk matrix (with ratings ranging from 'negligible' to 'extreme'), and separate risk acceptability guidelines.[44]

Shortcomings of the risk matrix have been pointed out both in industry and in academia (e.g. [62], [74], [75]). Importantly, the probability estimation aspect is often highly subjective from the outset (especially where humans, software, or untested novel hardware are concerned), and additional drift is often introduced at the organization level (e.g., 'the schedule is already slipping, make it less red'). Jaconetti points out:

A trap that testers sometimes fall into is the probabilistic nature of risk. Risk in experimental flight testing is rarely a statistical problem. There are methods to help predict risk of a system, but using new component reliability data is generally of little use because of the limited time in service. It is one thing to use statistics to try and trick a regulator into agreeing with your risk assessment, it's an entirely different thing to try and trick yourself. Further, the tricks aren't always a result of a conscious manipulation. Inherent biases can lead you down a path to convincing yourself that the risk is lower (or higher) than it really is. \ Unless the component is something that has very robust in-service history, the quantitative failure rates are of little value. [45]

And Luther et al. note:

While the limitations of a two dimensional risk assessment tool are impactful upon complicated systems… an extension to complex systems is undermined" by the very nature of complexity, where "[use] of probability in the model cannot be satisfied since probability is the ratio of a set of outcomes to all possible outcomes, a quantitative value that is unattainable for a complex system. . . .Kay and King (2020) rail against physics-envy in the social sciences when they note that the conflation of probability and uncertainty is a cultural product of the 20th century. They advocate for a return to differentiating between probability, the mathematical concept requiring knowledge of the denominator, and uncertainty. Uncertainty describes variable outcomes, for which the likelihood is unknown. The value in the concept is not propagating a false sense of knowledge. [49]

Various improvements and adaptations have been proposed, and some of them are used in practice.[62], [74], [75]

2.4. Limitations in THA

> "What we really learned from the Apollo fire, in the words of [former astronaut] Frank Borman, was the failure of imagination," said William H. Gerstenmaier, NASA's associate administrator for space operations. "We couldn't imagine a simple test on the pad being that catastrophic. "The failure of imagination." What a powerful phrase! What it means is that the complexity of aircraft design, the proliferation of software based systems, etc.—all of these things create hazards that we haven't even begun to imagine.
>
> – *Flight Test Safety newsletter* [28]

A core challenge of THA is its difficulty in comprehensively identifying test hazards.[54] This is aggravated where novel or complex systems are concerned. THA relies on open-ended brainstorming, which inherently introduces shortcomings. The ability to ideate hazards depends on the ability to imagine them, making certain types of hazards especially difficult to identify. Because of this, THA can result in incomplete hazard coverage and unanticipated losses.

Hazards that are particularly difficult to identify by THA include those without precedent, those which have not previously occurred or have not been previously experienced by the participants present during brainstorming, those emergent from

complex systems, those involving humans or software, and those which may occur even when all elements of a system perform as designed.

Aids to ideation, such as mental prompt lists, test hazard databases, and copy-pasting of prior analyses offer only partial relief, and introduce their own shortcomings. For example, "[w]hile a list of hazards such as that provided by FAA Order 4040[.26] can be a useful stimulation for ideas after the test team has exhausted its own imagination, its use as the initial brain-storming activity will constrain the identification of hazards and may result in potential hazards not being identified."[55]

Completeness and coverage of hazards identified by THA is dependent on the composition of the brainstorming team. The background, experience, culture, mindset, and interpersonal chemistry of the team members can all affect the results. While a cohesive team of experienced flight testers may produce as complete a list of test hazards as brainstorming may allow under the circumstances, gaps may still remain, especially where novel[16] or complex[17] systems are involved. Such team members and cohesion are not always available, and gaps in hazard identification have historically led to losses. Depending who shows up at each brainstorming meeting, entirely different sets of test hazards may be identified.

Even where non-novel or non-complex systems are involved, there is not always an appropriately experienced team of flight testers to pull from. For example, in the 1970's, "…very few new aircraft [had] been designed and flown in the past fifteen years, and therefore even industry… had little recent experience. Many aerospace companies were essentially without a real flight-test organization during the lean years, and some did not have company test pilots. Thus, with the renewal of aerospace activity, many companies had to put a flight-test team together from scratch. That can spell trouble."[66]

Similarly, even with an established and experienced team, novel or complex systems can stifle the ability to identify hazards comprehensively via THA, as in these cases relevant experience may be lacking or so sensitive to difference as to be immaterial. THA relies on experience and historical learning, yet novel systems may offer no prior experience to learn from. "Safety approaches that utilise prior performance are sensitive to repetition of the same loss scenario, but they are insensitive to new loss scenarios. This is not to invalidate the use of prior knowledge in a statistical model within a safety program. Rather it is to recognise the limitations upon a statistical model and that it is not applicable to complex systems."[49] Experienced teams may be able to identify test hazards with some success in environments of incremental changes to legacy, non-complex systems. However, the same is not true when novel systems are introduced; the behaviors of complex systems are sensitive to even small changes, and may therefore exhibit novelty unexpectedly.

[16] A more nuanced view of 'novelty' in this context recognizes that those technologies, configurations, situations, etc. that are sufficiently different from a flight tester's prior experience will appear to them to be novel, and thus suffer the effects of unfamiliarity, even if others have previous experience with them.
[17] Further discussion of complexity and its relationship to novelty follows shortly…

Complex systems exhibit several properties that render THA less capable of identifying hazards scrupulously.

- Emergence: a complex system may produce behaviors that are not readily predictable by decompositional analysis of its constituent elements.
- Non-linear behavior: interactions between system elements can produce salient behaviors far more complex than the proximal actions that initiate them.
- Sensitivity to changes: even small changes to the design of a complex system can produce large differences in behavior, often experienced as novel.
- Non-failure hazards: due to interactions between system elements, complex systems may exhibit hazardous behaviors even when no failure occurs and all elements function as designed.

Because of these characteristics, hazard identification in complex systems can require more structured and methodical prospective identification than THA provides.

"The impact of complexity on flight test is to preclude system knowledge, escalating the work and decreasing the safety."[49] Some complex systems may change configuration over time, as described by the Cynefin model, and therefore when testing them "it becomes an open question as to what version of a system an assessment was made upon."[49] Often, however, the configuration may be static, but with so many state variables present that observed behavior may change drastically for small changes in treatment. A 'System of Systems' approach is sometimes used, but such decompositional analysis can fail to reveal emergent properties.

Complexity, like novelty, can be a function of the tools and experience available by which to analyze and understand a system. To an organization with sophisticated holistic analysis tools, certain systems that appear complex to other organizations may merely appear complicated. "[A]n organisation may be unwilling or unable to invest the resources to learn a complicated system, that system will appear complex to them."[49] Other systems, however, may be so complex that no tool yet exists by which it may appear merely complicated. THA does not provide a high level of sophistication in hazard identification.

THA has been unable to keep up with some of the developments in aviation, and accidents have resulted. "[M]any [of today's aircraft] have new/novel features and complex software that will challenge historical thinking,"[45] yet "[t]est hazard analysis (THAs)… has remained unchanged for decades despite the increasing complexity of the systems undergoing test."[8]

No particular graphical representation or mental model underpins THA, but rather each participant is relied upon to understand the technology and its planned operation and environment, and from this to ideate hazards and plan mitigations. The lack of an underlying framework in THA to organize around hinders coherent, cooperative hazard identification across disciplines, sessions, participants, and meeting sessions. This is true for hazard identification, as well as for analysis and mitigation planning around hazards once they have been identified. Lack of an underlying framework hinders coherent, comprehensive hazard mitigation and management, as without a framework it is difficult to analyze the effects of any given mitigation on the rest of the system, and individual mitigations for each identified hazard may be overlapping, repetitious, or even harmful.

Lacking a unifying framework, THA can also fail to provide all the information assurance necessary for good flight test safety decision-making, especially to Flight

Readiness Review Boards and other such organizational gating functions. "[E]ven Best Practice THAs represent an incomplete communication of the whole safety argument."[12]

Preflight briefings from THA include hazards manually selected from the total binder based on relevance to the test techniques to be used and the equipment onboard, the hazards are typically briefed sequentially. The lengthy, disjointed, scattershot hazard-by-hazard briefings that ensue can leave crews exhausted and distracted from key takeaways by an inundation of details and repetition. "[O]ur airplanes are getting more and more complex. Our THA's are getting more and more detailed, more and more intricate, and we're starting to get to the point where we're getting stuck in this wall of text."[12]

While guidance is in place to require preflight hazard analysis and mitigation, the tools to do so can fall short as described.

In order to bolster safety despite these shortcomings, additional measures in standardization are often used to eliminate additional variability. Many flight test organizations "declare that the use of common facilities and tools is an enabler to flight test safety as it reduces the need for learning and introduces a common workplace irrespective of project."[7] This suggests yet another area of reliance on prior experience as in THA.

Together, this leaves the following core gaps in present practice:

1. Incomplete coverage in hazard identification.
2. Team composition-dependent results.
3. Reliance on availability of relevant prior experience.
4. Methodology not suited to analyzing complexity.
5. Does not provide a unifying framework.
6. Lengthy, disjointed preflight briefings.
7. Difficulty in analyzing mitigations.

2.5. Requirements for an Improved Methodology

While hazard identification will always rely on engineering subject matter expertise, it should not be required to rely so wholly on individuals' prior flight test experiences. This is especially true when novel technologies make past testing experience only superficially applicable. Furthermore, whether borne of experience, subject matter expertise, or otherwise, it should be possible to organize the hazards that are identified into a coherent framework, so that each hazard that is identified can help to reveal other related hazards. An underlying framework can also help to mitigate against the effects of there being "inconsistent expert knowledge at any given test-safety review board" [17] by enabling whatever group may be assembled at a given meeting to effectively address all hazards without loss of context.

Organizing hazards into such a framework will also allow for mitigations to be designed in a comprehensive manner across multiple hazards, and for the effects of mitigations (including design or procedural changes) to be properly analyzed and vetted with regard to other hazards and to the system as a whole. This helps to avoid the case

where an amalgam of mitigations "renders the system opaque" and enables new mistakes to be made [35].

An improved methodology for test hazard identification should provide comprehensive coverage, and facilitate those other activities associated with analysis, preparation, mitigation, safety, and decision-making. "To inform go/no-go and other decisions in flight test, we need to know where the holes are before the test."[8] And "[r]eliance upon statistical databases of prior occurrences needs to be avoided on account of the unique (new & singular) nature of complex hazards. Rather, top- down analytic tools… are required to capture emergent functions within an analysis. With complex systems featuring emergence, it will be necessary to adopt a top-down approach in at least one analysis, to capture emergent functions."[49]

A structured process should be used for identifying, documenting, and analyzing hazards even in novel systems without prior experience, and with top-down capability to analyze complex systems.

Safety review and decision-making processes such as Flight Readiness Reviews should be provided with useful, informative, concise information.

A unifying mental model should be included to organize around, permit cohesive work across different meetings with different experts, and if possible facilitate productive contributions by non-experts.

To the extent practical, an improved methodology should facilitate existing flight test safety organizational, decision-making, and training approaches, and not require extensive reworking of processes or culture. Established thought patterns of 'identify, mitigate, recognize, recover,' and the elements of the THA (hazard, cause(s), effect(s), mitigations(s), recovery, risk assessment) should be maintained.

Organized documentation should be readily produced to brief from. This should include a way of facilitating identification of relevant items to brief, and allowing concise, thorough, on-point, and operationally valuable briefings with a minimum of repetition.

Ideally, an improved methodology should also permit ready re-analysis following any changes, learnings, or reconfigurations, especially by introducing hazard mitigations. This should especially facilitate efficient re-analysis when test equipment, staffing, or locations are changed to accomplish different tests.

An improved methodology should integrate readily with existing and legacy analyses (FMEA, PRA, etc.) in a streamlined way.

Summarizing, an ideal improved practice would provide the following:

1. Complete coverage in hazard identification.
2. Results independent of team composition across meetings.
3. Process not reliant on prior relevant experience.
4. Methodology suited to analyzing complexity.
5. Underpinned by a unifying framework.
6. Ability to produce concise, efficient preflight briefings.
7. Ability to analyze mitigations.
8. Supports existing organizational processes.
9. Ability to re-analyze efficiently after changes.
10. Integrates with existing and legacy analyses.

It is hypothesized that STPA may provide a way to accomplish these goals.

2.6. Systems Theoretic Process Analysis (STPA): an Overview

Systems Theoretic Process Analysis (STPA) [76] is a hazard analysis technique rooted in the Systems-Theoretic Accident Model and Processes (STAMP) model of accident causality.[40] It operates within the broader framework of systems theory, capturing a holistic understanding of systems that includes the interactions within them and the emergent properties they produce. By facilitating methodically structured analysis of system-level behaviors, STPA enables penetrating insights across system types and levels of abstraction.

STAMP offers a holistic perspective on the system. Even when all elements of the system operate as designed, it is still possible for hazardous or otherwise detrimental behaviors to emerge. Rather than describing accidents as undesirable chains of events, uncontained energy, lapses in layers of mitigation, phase changes, or other concepts that are linear, abstruse, or hindsight-centric and difficult to derive predictions from [8], [34], [35], [77], STAMP instead recognizes that accidents emerge from the collection of processes occurring within a system, and the controls and feedbacks amongst them. Systems theory treats the system "as a whole, not [merely] as the sum of its parts… 'the whole is more than the sum of its parts.'"[76] This allows for an integrated view that easily incorporates not only hardware, but also equally software, human factors, and any other contributing discipline. Applicable at various levels of abstraction, STAMP is a powerful framework for understanding emergent properties even in complex systems.

STPA is a forward-looking safety analysis methodology based on STAMP, and as such is able to produce meaningful results even where other methodologies struggle, such as in systems involving software, human interactions, or other forms of complexity that exist in aviation and flight test. By considering the interconnections and dependencies within a system, STPA can capture conditions that could lead to hazards, whether due to design, operation, environment, failures, errors, or other aspects.

Once the boundaries of a system are delineated, and the losses (negative outcomes) of concern determined, STPA provides methodical tools for modeling the system, elucidating its hazards (potentially unsafe behaviors), and envisioning their causes and means of mitigation. STPA and the systems theoretic framework offer more than one window into the holistic picture, yielding multiple opportunities to identify and analyze even the kinds of hazards that can tend to evade the more linear traditional analysis methods. Because the STAMP framework considers the system as a whole, with all its complexity and interactions, analysis from STPA is not restricted to any one line of inquiry, and so can reveal broader-reaching and more subtle emergent behaviors. Common-mode failures, for instance can sometimes have knock-on effects not identified in a top-down analysis, but which STPA may be able to find thanks to its holistic integrative approach.

STPA is described in further detail in Section 3.1, using the framework of flight test.

Of note, while STPA is forward-looking, or *prospective*, a complementary *retrospective* methodology, Causal Analysis based on System Theory (CAST) [78], also

exists. CAST can be used wherever unwanted behaviors have been observed, such as for accident investigation, in order to obtain comprehensive, actionable systems theoretic insight.[78] Given the events of an accident or incident, CAST helps to elucidate the system interactions at play in causing the event, as well as other hazardous interactions and properties of the system that may be uncovered in the process of investigation. Together, the STAMP-based tools of STPA and CAST provide comprehensive prospective and retrospective system theoretic hazard analysis capabilities.

An appealing aspect of STPA in the flight test context is "the methodical way [it] identifies hazards using the control theory framework, rather than trusting the (possibly disorganized) brainstorming ability of the test team."[79]

3. STPA in Flight Test

> "Planning is bringing the future into the present so that you can do something about it now."
>
> – *Alan Lakein* (as quoted in [43])

STPA offers strengths that can meet key needs not fulfilled by THA, but it is also substantially different from THA, both in conception and in implementation. It is therefore not a 'form, fit, function' drop-in replacement for THA. This section exposes and elucidates key differences between STPA and THA to better understand the implications and opportunities of using STPA in lieu of THA. Emphasis is placed on the elements provided by THA that are used to facilitate organizational processes around test, such that STPA may be used in lieu of THA without change of broader processes. The section then further explores opportunities to harness the framework behind STPA for several additional aspects of flight test safety.

STPA and THA operate in different theoretical frameworks. Regardless of framework, however, "[h]azard identification is a prerequisite to the safety risk management process… A clear understanding of hazards and their related consequences is essential to the implementation of sound safety risk management."[44] STPA, and the STAMP framework more broadly, offer a mental model and a methodology to comprehensively identify and analyze hazards, even in novel or complex systems.

The STAMP framework employs certain language with specific meanings, sometimes using terminology common to flight test in different senses (i.e., conveying different meanings) than their colloquial use in the profession. This can cause confusion when adopting STPA as a new tool for test hazard identification and analysis. Therefore, STPA is explained and clarified in the context of flight test as to consciously avoid negative transfer due to primacy.

Some key concepts and terms related to STPA appear to have already caused misunderstanding in the context of flight test.[8], [49] These are clarified in order to facilitate smoothest possible access to these tools by flight testers.

Of note, while STPA is not explicitly a tool for surfacing or documenting assumptions, the methodology does naturally help to expose and account for them to some degree. Because flight test sometimes violates or tests assumptions (explicit or implicit) made in design, using STPA for test hazard identification and management can assist in revealing some of these to permit deliberate rather than unwitting exposure.

When using STPA in the context of existing risk management practices, it will be necessary to supplement the analysis it produces to satisfy the needs of those processes.

STPA, like other safety management tools, is not a stand-in for strong safety culture, which is required in any case. "An analysis of Flight Test Accidents often reveals a breakdown in the Safety Culture under programme cost and time pressures especially where management fail to live the values [of safety]."[7] STPA, however, unlike other methods, facilitates analysis of the safety culture and processes themselves if desired, in addition to analysis of test hazards.

By offering a unified mental model and operating framework, STPA may be able to facilitate improved preflight briefings. Since many test campaigns involve similar

infrastructure, templatized STPA control structures may offer opportunity for accelerated documentation and analysis without the potentially misleading suggestibility of copied THAs.

3.1. STPA and THA: Merging Mental Models

The product of THA is a compendium of hazard documents, each detailing the following collection of elements:

- Identified hazard description
- Cause(s) of the identified hazard
- Effect(s) of the identified hazard
- Mitigations / minimizing procedures / precautionary measures against occurrence of the hazard
- Recovery / emergency procedures / corrective actions to prevent or mitigate the severity of effects in case of occurrence of the hazard
- Risk assessment of the hazard

Flight test risk and safety management processes presently employed in both civilian and military organizations are designed around this collection of elements. Various authors have detailed shortcomings and potential improvements to some elements (especially to the hazard identification and risk assessment elements), but the collection as a whole appears to work relatively well. Therefore, STPA for flight test will be outlined in the context of this same collection of elements, showing where STPA produces the requisite analytical content.

STPA differs from THA in the structure it provides for comprehensively identifying hazards, and the richness it offers in understanding causality. STPA identifies hazards via a methodical, structured process informed by the full context at hand, rather than by open-ended brainstorming. Hazards identified by STPA are revealed in context together with their causes and effects, providing the first three elements of the THA product. STPA does not inherently provide the elements of mitigation opportunities or corrective actions for recovery, however both may be more readily revealed by the systems view offered in STPA, even if standard mitigation strategies may still be used; and proposed mitigations and corrective actions may be evaluated for effectiveness as well as any knock-on effects under the same STAMP causality model. The risk assessment element from THA is also not provided by STPA, though the thoroughness and richness of STPA results may better inform any risk assessment methodology that is used to produce this element (in the event a risk element remains desired despite the theoretical difficulties of risk assessment in this context).

Some of the terminology used in STPA carries specific meaning that is integral to the operation of the method. The meanings of some the terms in the context of STPA are different from their meanings when used in THA. Whether across disciplines, cultures, "organisations, and indeed national boundaries, [terminology] varies for aspects of safety management for testing and aircraft operation. . . .It is important to constantly verify that the test team understand the terms being commonly used in relation to safety management and risk mitigation. This ensures that the team have a consistent application of risk mitigation and safety management."[7] Therefore, important terms will be explained in context as their relevance to STPA is introduced.

The constructs of STPA (and its underlying STAMP framework) are different from those of THA, and more precisely defined than typical flight test framings. These will be described and compared to show their implications on the use of STPA instead of THA.

System & System Boundary:

STPA carefully defines a system to analyze, and then sets about analyzing it methodically. In THA, a somewhat nebulous collection of systems and components may be mentally contemplated, pulling in or leaving out elements according to their intuitively perceived relevance. In STPA, by contrast, the collection of elements to be analyzed are carefully selected according to the purpose of the analysis, and a clear boundary drawn around them.

The word system is used by STPA in a systems theoretic sense, which is different from some colloquial usage. Pilots often think of a system as describing the technological components that collectively offer a particular functionality, such as an aircraft fuel system, electrical system, hydraulic system, or even an entire weapon system. This meaning is also used frequently in THA. In the context of STPA, however, the word 'system' can be considered to mean a 'system under analysis,' and can refer to any collection of elements selected to be analyzed together, and includes the interactions and relationships amongst them. "A ***system*** is a set of components that act together as a whole to achieve some common goal, objective, or end." (emphasis added)[76] System components may be technological components (hardware, software, electronics, etc.) as implied in colloquial use of the term 'component,' but they may just as readily be non-technical components such as people and organizations. For clarity, this thesis will refer to the pieces within a system more generally as 'elements.' Of note, "[a] system may contain subsystems and may also be part of a larger system,"[76] which is to say, the system being examined may be part of a larger whole, or composed of smaller segments.

To analyze what would traditionally be called an aircraft "fuel system," one could indeed define the system as comprising the fuel tanks, plumbing, valves, cockpit controls, fuel management systems, etc., just as a pilot might conceive of that system. However, aviation systems are typically 'sociotechnical'– that is, they comprise both social (human) and technological elements. To analyze the "fuel system" using STPA, it might be beneficial to define the system under analysis to include also the flight crew, the fuel truck operator, the fuel truck itself, etc., in addition to the physical components mentioned above. These elements would not necessarily be considered overtly by THA. With them, however, the analysis is able to address many more safety-critical aspects of the overall system in use.

Alternatively, some systems that may benefit from STPA analysis are purely social; for example, for STPA analysis of flight test continuation go/no-go decision-making, the system might be defined to include just the set of people directly involved in that decision-making, such as the test pilot, flight test engineer, program manager, and safety chief, and the documented hazard analysis results. Such a system is illustrated conceptually in Figure 13. These might be less readily addressed by THA. Or perhaps continuation criteria are based on telemetered data, in which case the telemetry system might also be included in the system under analysis, making it yet again a sociotechnical system.

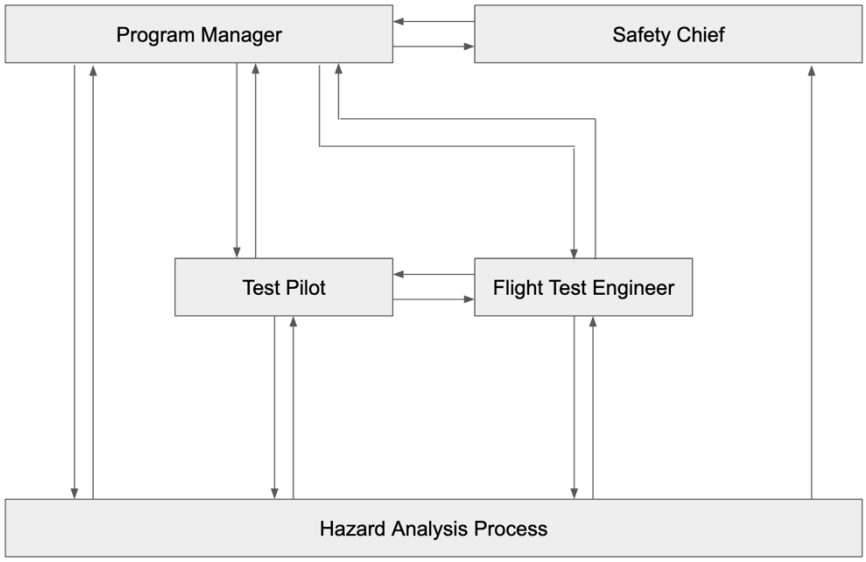


Figure 13: purely social system example

Delineating between those elements included within a system under analysis and those that are not is a **system boundary**. While anything inside the system boundary is of course our system, anything outside can be referred to as the **environment** or 'operating environment' in which the system operates. Once a system boundary has been defined, then the system may be modeled and analyzed. Without a delineated system boundary, it would be impossible to know how far afield the analysis should extend.[18] Thus STPA becomes more deliberately scoped than THA. With a system thus encapsulated, in addition to examining its internal workings, it is also possible to contemplate its external responses to external stimulation, as illustrated in Figure 14.

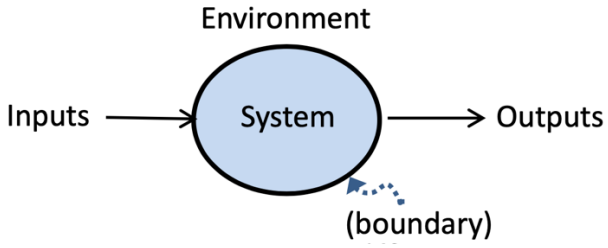


Figure 14: system, system boundary, and environment [78]

Control and Feedback Loops:

The STAMP model of system behavior centers on the concept of control: elements within a system exert control over other elements, and these in turn may alter their behaviors accordingly. The various controllers choose what controls to exert over time by observing behaviors within the system via the feedbacks they produce. The interaction of control and feedback amongst the various system elements produces the total behavior of the system. This is not dissimilar to THA, however it is far more

[18] The famous Carl Sagan quote comes to mind, "if you wish to make an apple pie from scratch, you must first invent the universe."[80]

structured and formalized, and thus not so sensitive to the mental models of individual participants as in THA. The formalized concepts are straightforward:

Each **controller** contains a process model and a control algorithm, which inform its beliefs and direct its actions, respectively. Controllers can be technological, such as a thermostat or a piece of software– or they may be sociological, such as a human or an organization. Controllers exert control via 'control actions' (further detailed below), and assimilate information via feedback (also further describe below). Figure 15 illustrates a generic controller, showing its internal process model and control algorithm, as well as the control actions it issues, and the feedback it receives. Note that this same conception works whether the controller is technological, sociological, or otherwise.

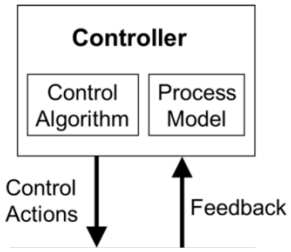


Figure 15: controller elements (adapted from *[76]*)

Of note, the term control is used differently in STPA than in risk management. In risk management, the term control is used to denote measures taken to reduce risk (in the sense of 'risk controls' and 'controlling the hazards') or to increase safety (e.g., 'safety controls'), and the term 'control' can seem problematic as "it is difficult to control that [risk] which is not known"[8]. In STPA, by contrast, the term **control** is used in the sense of 'command and control'– that is, of one entity controlling (instructing or commanding) another. Auto-throttle, for example, is a controller that controls the engines by issuing throttle commands in accordance with its programmed modes and sensed flight conditions; the flight crew are controllers who control the auto-throttle modes and engagement/disengagement by pushing buttons and turning knobs in accordance with the desired mission profile of the flight.

The particular commands issued by a controller are called **control actions**. Engaging and disengaging the auto-throttle are each a control action which might be issued by the pilot. Advancing or retarding the throttle might each be control actions issued by the auto-throttle logic. Continuing or discontinuing envelope expansion to the next test point are each control actions that might be issued by the decision-making committee mentioned above. Control actions are described in terms of their content or meaning rather than their method of delivery. For example, a pilot guides and directs the physical course of an aircraft; it is less important whether they do this by applying pressures to a yoke, issuing commands to an autopilot, or any other means.

Controllers issue control actions in accordance with their understanding of the state of the world around them, and their model for how best to act in the circumstances.

A controller's understanding of the state of the world is called its **process model**. This internal representation of beliefs about the process being controlled may be anything from totally open-loop assumptions to full and timely information. For example, an auto-throttle's process model may include present airspeed and mode settings, which may be

updated periodically or filtered over time from measurements, and stored as state variables. A human pilot's process model may include airplane dynamics, flight parameters, mission objectives, and air traffic control commands, which may be assimilated intuitively or anticipated based on past actions, and stored as kneeboard notes, memories, or situation awareness. Controls and estimation engineers might conceive of process models as analogous to state variables and their update algorithms; psychologists might relate them to mental models and sensory synthesis.

Controllers update their process models via **feedback**. Feedback is information received by a controller from other elements in the system. A flight crew may receive mode status feedback from an auto-throttle system via status indicators on the panel. The auto-throttle system may receive airspeed feedback from the airplane via pitot-static measurements. A human pilot may observe flight instruments and panel indications, a view out the window, control feel, and auditory cues.

A controller's model for how best to act is called its **control algorithm**, and may take the form of discrete logic, mathematical models in the time or frequency domains, intuition, rulebooks, or other representations. The auto-throttle, for example, may have a control algorithm that advances the throttles to full power when 'takeoff/go-around' mode is engaged. The human pilot may have a control algorithm that commands 'takeoff/go-around' to the auto-throttle when initiating a missed approach.

A system element accepting control actions from another system element is called a **controlled process**. An element can be both a controller and a controlled process. For example, in the context of an IFR flight, the pilot is clearly a controller, issuing various control actions to the airplane; but the pilot also accepts instructions (control actions) from air traffic control, and thus is also a controlled process. Many controlled processes provide feedback to their controllers to facilitate closed loop control.

Much of the above is understood implicitly when conducting THA, and is sometimes described explicitly for individual pieces of a system under that method. However, defining and analyzing controls and feedbacks explicitly as in STPA provides structure and thoroughness that is lacking in THA. The full implications of system element behaviors may be completely traced when they are assembled and analyzed together. When control and feedback occur in closed circuits ('control loops,' detailed further below), then holistic system behaviors emerge.

A **control loop** is when a controller and a controlled process are linked together such that control actions from the controller affect the controlled process, which in turn provides feedback to the controller, thus closing a loop. Figure 16 illustrates a control loop with a generic controller and generic controlled process (each labeled); the control loop flows counterclockwise, with the controller issuing control actions (downward arrow at left) to the controlled process, the controlled process evolves accordingly, the controlled process then sends feedback (upward arrow at right) to the controlled process, and the controlled process updates its process model and issues new control actions per its control algorithm accordingly, thus completing the loop. The loop illustrated contains only two system elements, however control loops may occur over any number of elements. Controllers may themselves be controlled processes, and vice versa. Control loops can occur across multiple elements, and over varied control and feedback paths, including over multiple paths simultaneously.

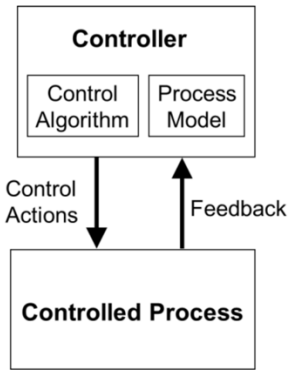


Figure 16: generic control loop [76]

Control loops with few elements or those sequentially arranged are readily contemplated when conducting THA, especially where they have been intentionally engineered. STPA is also able to capture and analyze more complex, multi-element, and unintentional control loops that are difficult to capture by THA.

Control Structures:

In order to consider complete systems, STPA models their elements together, drawing the connections and relationships of control and feedbacks between them. Represented as a schematic, such a model is called a **control structure**[19], and it depicts the various elements of a system and the links between them. Such a representation goes beyond the disparate design documents which typically inform THA, to actually represent the system holistically and completely all at once. For example, Figure 17 shows a representative control structure for a system including Air Traffic Control, a Flight Crew, Aircraft Automation, and a Physical Aircraft.

[19] Previously sometimes called "Functional Control Diagram" (FCD) or "Functional Control Structure"

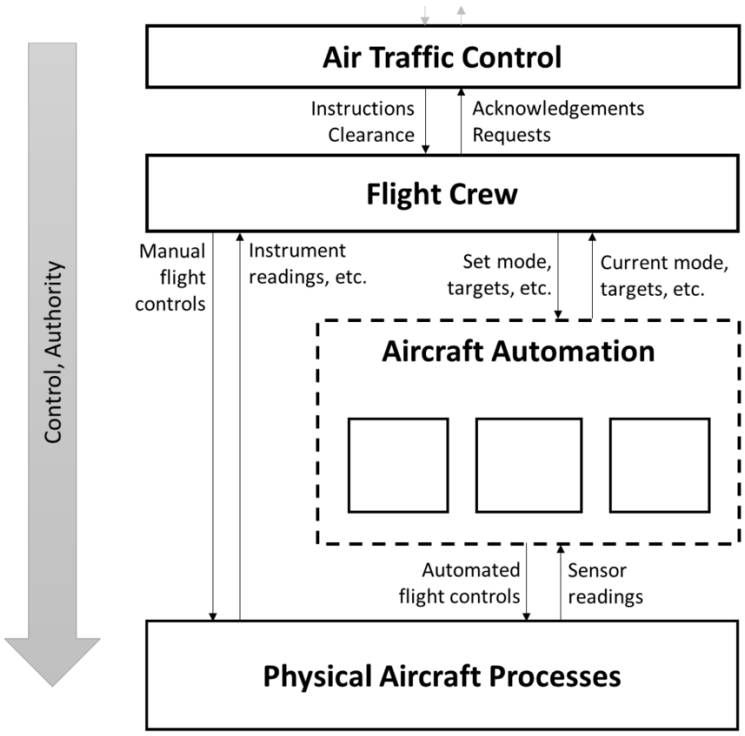


Figure 17: example control structure (adapted from [76])

The elements in a control structure are laid out in a hierarchical format, with top-level controllers typically at the top of the page, lower-level controllers below, and onward with lower and lower levels of controllers down the page, as illustrated in Figure 17. Controlled processes that are not controllers are typically placed at the bottom. Control and feedback pathways are drawn as arrows to indicate the sources and recipients of information flow. Control pathways are drawn as arrows connecting the controller issuing control actions to the recipient of such control actions; these arrows typically point downward due to the hierarchical arrangement. Feedback pathways are similarly drawn as arrows connecting the sources of feedback to the recipients of their feedback; these arrows typically point upward due to the hierarchical arrangement. Figure 18 depicts an illustrative control structure. At the center is the control loop of Figure 16, and around it are additional elements ("Part A," "Part B," "Additional Parts of the System," and "Even More Parts of the System."). As evidenced by the arrows denoting control and feedback flows, all of these elements are controllers, and some are also controlled processes. The control loop amongst Part A, 'Even More Parts of the System,' and Part B is interesting in that three elements are involved in closing the loop; this is a small example of the more complex interrelationships inherent in some systems. When modeling a real system, each arrow should be labeled with the types of control actions and feedbacks it denotes, for example 'aircraft guidance,' or '6 DOF state information,' etc.

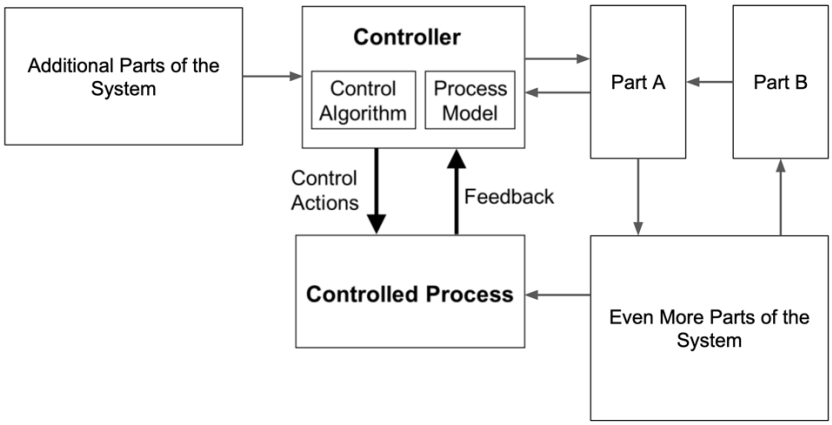


Figure 18: generic control structure (adapted from [76])

System elements may be modeled at various levels of **abstraction** to suit the needs of the analysis. For instance, it may not be necessary to model each component of a circuit as its own element, but rather to model the collection of components together– representing the entire circuit collectively as a single more abstract element. Conversely, while sometimes it can be insightful to model an organization collectively as a single abstract element (for instance when analyzing the interplay between several organizations), it might in other circumstances be more productive to model each division within an organization (for example when analyzing internal organizational issues), or even to model each role within a team separately (such as when analyzing a division). These illustrate decreasing levels of abstraction.

Analysis may proceed at different levels of abstraction in parallel, yielding results in different levels of detail. Further thoughts on abstraction in flight test system modeling are in Section 3.2.

Even in the event the full methodology of STPA is not used, the control structure diagrams it produces have been found to be useful in informing test hazard identification and facilitating communication and collaboration among test team members.[68]

<u>Abstraction:</u>

When modeling a system as a control structure, a lot of freedom is afforded in the level of detail chosen to represent various elements. Taking a beehive as an example, the system might validly be modeled detailing each individual honeybee– or even each neuron within each bee– as well as the hive box, its frames, and the plinth it sits on (significant detail, hardly abstracted from basic physiology). Alternatively, the same system may be modeled abstractly as a collection of bees and a physical hive (low in detail, very abstract). It might also be modeled at an intermediate level of abstraction as a collection of worker bees, a collection of drone bees, a queen, a collection of larvae, and a physical hive.

Abstraction is a key to balancing salience, workload, and detail when using STPA. In a theoretical sense, abstraction describes the degree to which the behaviors of elements of a system model are lumped together or considered individually. In a practical sense, it determines how ‘deep in the weeds’ an analysis goes, and how detailed or general its results.

When a group of sub-elements (such as individual bees) are collected into one abstract element that represents all of them together at a higher level of abstraction (such as the abstract collective element 'worker bees'), the collective behavior of all the sub-elements is then treated together. Why or how their collective behavior emerges may or may not be relevant for any particular analysis, but the effects of the collective behaviors can still be analyzed. Collecting them together reduces the workload and the search space involved in analyzing them.

Abstracting a system can be something of an art, though given the general decompositional approach adopted in engineering fields, certain choices may seem fairly obvious to engineers. When abstracting the beehive at the intermediate level, for example, it seems natural to group bees who share social roles together into their own system elements. Similarly, to group hive elements together. One could, of course, do otherwise, but ultimately such choices may make analysis easier or more difficult.

Mental models employed by participants in THA may naturally adjust abstraction levels to explore and reveal potential hazards. In STPA such a process may be undertaken deliberately, and shared across participants via graphical control structure diagrams to facilitate intentional communication and collaboration.

Losses, Hazards, and Safety Constraints:

Where THA may implicitly assume 'any damage, injury, destruction, or death' are to be avoided and prevented, STPA is instead explicit and structured in its definition of losses. **Losses** describe particular negative outcomes that may result from system behavior, and which can hopefully be prevented. Each loss describes a particular negative outcome, such as (for the examples stated above): 'damage to the aircraft,' 'bodily injury,' or so forth; and does not describe how or why such outcome may occur.

Hazards in turn describe situations that may lead to losses. In STPA, "[a] *hazard* is a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss."[76] Hazards may be described at various levels of abstraction or detail, and it can be helpful when performing analysis to work downward from a high level into the details. At the top level, a basic set of hazards may be described that may lead to each relevant loss. For example, under unfavorable conditions the loss 'damage to the aircraft' might occur due to the following top-level hazards (and possibly others):

- Aircraft does not maintain safe separation from obstacles or terrain.
- Aircraft does not maintain safe separation from other aircraft.
- Aircraft exceeds operating envelope.
- Aircraft subjected to excessive temperatures.
- Aircraft exposed to destructive materials.

A hazard describes an overall system state, rather than the details of any particular components, and is addressed to aspects and behaviors that are controllable.

Explicit loss and hazard definitions in STPA permit traceability that is not possible in THA. Each top-level hazard may be traced to the losses it may cause, and later in the process, each loss scenario may be traced to its relevant top-level hazards. Lower-level hazards may also be delineated and traced to their respective higher-level hazards, and on upwards to top-level hazards and relevant losses.

To implement traceability, each loss and hazard is labeled with a unique identifier (often, simply "L1," "L2," etc., and "H1," "H2," etc.), and each hazard tagged with its relevant higher-level hazards or originating losses, as shown elsewhere in this thesis.

Four-Step STPA Process:

Unlike THA, where analysis is essentially freeform, the STPA process is composed of four concrete steps with increasing levels of detail. The process is depicted as a flow chart in the upper part of Figure 19. Each step is further illustrated in the lower part of the figure as indicated by its respective green arrow, and further described after a brief overview:

1. Define the purpose of the analysis. This highest-level step includes delineating the system and the system boundary; defining the losses of relevance; and determining the top-level hazards that might lead to those losses.
2. Model the control structure. This involves modeling the elements of the system at an appropriate level of abstraction, arranging the elements hierarchically, and describing the control and feedback pathways that connect them.
3. Identify unsafe control actions (UCAs). In this step, each possible control action is analyzed for contexts in which it may become unsafe, and the UCAs identified are documented. This creates a structured framework in which to identify detailed causal loss scenarios.
4. Identify loss scenarios. This step involves identifying and capturing the causes, system behaviors, and contexts that lead to UCAs occurring and developing into losses.

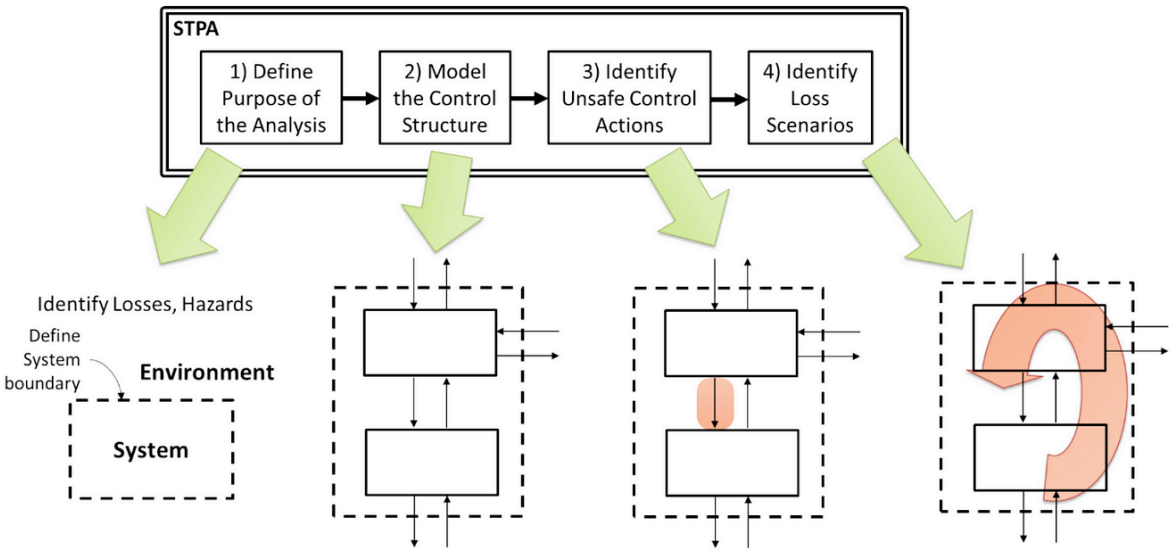


Figure 19: The four steps of STPA [81]

Step 1: Define the Purpose of the Analysis

The first step in STPA is to define the purpose of the analysis; that is, to specify the system to be analyzed, the losses that are of interest, and the top-level hazards that can lead to those losses. These are typically implicit in THA, but are described overtly in STPA. This offers structure and traceability (and thereby intellectual accountability) in the analysis that THA does not provide.

The purpose of STPA analysis in the context of flight test will be to reveal hazards that may jeopardize the safety of flight. The system should typically be defined to include those elements involved in the operation or functions under analysis, and that can be reasonably segmented off from the surrounding environment by via a clear system boundary. Losses of interest typically include those that unacceptably impair safety, but may also involve losses related to security, financial, or other negative impacts. Safety-related losses in aviation will typically include:

- L1: Injury or loss of life of passengers or crew.
- L2: Damage or loss of aircraft.
- L3: Injury or loss of life of persons other than passengers and crew.
- L4: Damage or destruction of property or infrastructure.

Losses to be analyzed in flight test can productively include process losses and reputational losses that can impede present and future productivity, for example:

- L5: Incorrect, corrupt, or missing test result data.
- L6: Loss of trust by controlling/regulating organizations.

Each loss is labeled (e.g., "L1," "L2," etc. as above) to facilitate traceability throughout analysis all the way back to top-level concerns.

With the system and losses of interest defined, a set of top-level hazards is produced; these are high-level system states or behaviors that may lead to the losses, and serve to frame safety constraints on the system. Hazards are likewise labeled to facilitate onward traceability in the analysis, and each such top-level hazard will be tagged with its associated losses for traceability. Top-level hazards in flight test may include:

- H1: Aircraft gets too close to terrain or obstacles (L1, L2, L3, L4, L5, L6)
- H2: Aircraft gets too close to other aircraft (L1, L2, L3, L4, L5, L6)
- H3: Aircraft exceeds safe operating envelope (L1, L2)
- H4: Flight violates regulated airspace (L1, L2, L6)
- H5: Flight crew loses sense of being in command of the aircraft (L6)

From each hazard can be derived a corresponding safety constraint. For example, if it is hazardous for an aircraft to get too close to terrain or obstacles, then a corresponding safety constraint states that 'the aircraft must maintain a safe distance from terrain and obstacles.' Each constraint must be suitably enforced by the system in order to prevent the occurrence of losses. By simple inversion, each hazard informs a constraint. Each constraint is one-to-one traceable to its respectively labeled hazard. For example, the top-level hazards above induce the following top-level constraints:

- C1: Aircraft shall maintain safe separation from terrain and obstacles
- C2: Aircraft shall maintain safe separation from other aircraft
- C3: Aircraft shall remain within its safe operating envelope
- C4: Flight shall remain clear of regulated airspace
- C5: Flight crew's ability to command the aircraft shall not be interrupted

Step 2: Model the Control Structure

The second step of STPA creates a control structure modeling the system. Flight testing takes various forms, each involving different combinations and configurations of elements. In order to analyze the system in STPA, it must first be modeled as a control structure. For each system under test and various methods of testing, the system and system boundary may be somewhat different, and in order to analyze a system at various levels, choices of abstraction level will also be made, so many different control structures are possible. While THA may make use of various diagrams and schematics during analysis, in STPA the control structure diagram is a central reference element, and may be refined and updated over time as analysis proceeds and changes are made.

Figure 20 shows a representative control structure for day-of-flight analysis in representative civilian certification flight testing. It is abstracted to a level showing participants and whole aircraft and includes those elements actively participating in flight operations. The control structure was created with informal consultation from test pilots involved in civil certification flight test safety. In it, a Test Aircraft is flown by a Test Crew, accompanied by a Chase Aircraft and its Chase Crew, in contact with a Control Room, supervised by a Test Director, and under the guidance of Air Traffic Control (ATC). Test data is also telemetered from the Test Aircraft to the Control Room.

Each of the human elements in this example (perhaps aside from the test director) may comprise several individuals, however for the purposes of the analysis served by this model perhaps it is not necessary to dive into such detail; as such, that complexity is abstracted away in this control structure, and only its higher-level behaviors and interactions will be analyzed. The detailed view, should it be determined to be useful, can always be expanded later, either in-place in the same control structure, or separately as its own analysis. Similarly, both aircraft comprise numerous parts, controllers, and controlled processes; these too are abstracted away for the purpose of this analysis.

Hierarchically, this example places ATC and the Flight Director as co-equal top-level controllers; the control room is placed below them, still in command of the test and chase crews; and the test and chase aircraft are at the bottom, controlled processes taking command only from their respective crews.

As shown in the figure, the Test Crew solely controls the Test Aircraft, from which it also receives feedback, and around them are the other elements. ATC and the Control Room each control both the Test Crew and the Chase Crew, and the Test Director controls these both as well as the Control Room. This means each crew receives control from three different controllers. Unless a clear hierarchy and division of command are emplaced, this could cause undesired behaviors.

Each controller in this example receives feedback from its respective controlled processes, creating small control loops. Larger, more complex control loops exist as well, such as possibly (following arrows around the diagram) from ATC to the Test Crew to the Test Aircraft to the Control Room to the Test Director to the Chase Crew and back to ATC. Within this loop are also many interrelated smaller or adjacent loops, and already it becomes obvious that what may have appeared a simple collection of people and equipment going out to fly a mission can in the wrong conditions produce surprising emergent behaviors. While THA treats these from experience and intuition, STPA treats them methodically.

The control structure itself may prompt modifications to practice that can solve such problems, or other procedural or design solutions may be implemented. Already just having a control structure to refer to can be hugely helpful. In addition to informing analysis and mental models, and facilitating unified group teamwork, control structures can also be useful when making a safety case to a review board. "[Control structures] are a great way to communicate function and potential risks to a safety board or approval authority. It ensure[s] everyone has the same mental model of the system under test."[82] This may be true whether THA or STPA is used for the remainder of the analysis.

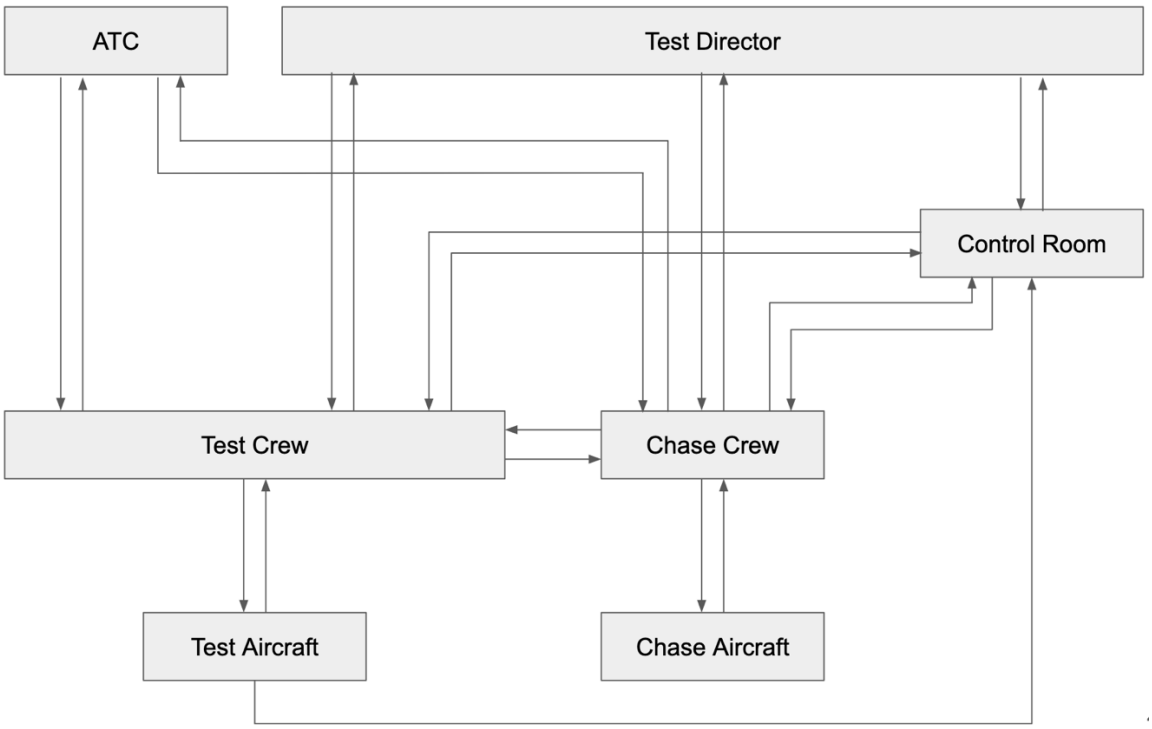


Figure 20: 'Day-of-flight' control structure, flight test with chase and ground station

The example in the figure is, of course, just one of many possible configurations and system/system boundary definitions, modeled at one particular level of abstraction. Different system elements, system configurations, etc. could all be modeled, including at higher or lower levels of abstraction as discussed further in Section 3.2. The above example control structure represents one generic civilian certification test configuration. US Air Force 412th Test Wing, for example, might instead include Range Control in lieu of ATC, coordinating with a Test Director, and a Test Conductor in a Control Room coordinating with the crew. Each team, each program, and possibly each flight may have its own unique configuration, and it is important to capture and model these correctly.

Once a control structure is modeled, the remainder of the analysis may proceed; however, the control structure initially drawn should be considered an initial draft, as errors and edits may be discovered as analysis progresses.

Step 3: Identify Unsafe Control Actions

The third step of STPA introduces another layer of structure to the analysis. In contrast to the open-ended brainstorming of THA, the third step of STPA constructs an organized framework within which hazard identification will be undertaken methodically in the fourth step. The third step requires knowledge of the system and its operating environment. It is a bottom-up search, and produces a structured set of conditions that may cause hazard. For this, each controller in the system is examined, and all of its possible control actions assessed to determine under what conditions they may become unsafe. These are referred to as 'unsafe control actions.'

An **unsafe control action (UCA)** is a control action a controller may issue that, in some context and under worst-case conditions, may cause a hazard to be realized. For example, an auto-throttle issuing a throttle-retard command when 'takeoff/go-around' mode is engaged and the aircraft is near the ground could result in the aircraft impacting the ground. Here, the 'throttle-retard' control action can be hazardous if issued in the ''takeoff/go-around' mode is engaged and aircraft is near the ground' context. Similarly, a pilot disengaging the auto-throttle when on approach could cause the aircraft to impact short of the runway (for example if the pilot expects the auto-throttle to be holding V_{REF} approach speed). Each of these examples describes an instance when *providing* a particular control action is hazardous under a particular set of circumstances.

UCA's may also exist wherein *not providing* a control action, or *providing it in the wrong way*, can prove dangerous, as will be described below. It is also the case that the same control action may engender multiple UCA's each with different contexts. For example, an auto-throttle issuing a throttle-retard command to one engine when a multi-engine aircraft is taking off could result in the aircraft impacting obstacles beside the runway; this is the same control action (throttle-retard) illustrated above, but becoming unsafe under an additional circumstance different from that illustrated above. Thus, a UCA is composed of several simultaneous elements, and multiple UCA's may exist based on each control action that a controller may supply.

A control action engenders a UCA when, under some set of conditions that may occur, one of four general patterns of action may compromise safety [76]:

1. *Not providing* the control action (when it should be provided) leads to a hazard.
2. *Providing* the control action (incorrectly or when it should not be provided) leads to a hazard.
3. Providing the control action *too early, too late, or out of order* leads to a hazard.
4. Maintaining the control action for *too long or too short*[20] leads to a hazard.

These together fully describe the presence or absence, timing, and duration of a binary pulse, allowing a mathematically complete analysis of continuous-time control action behavior. Only the first three patterns pertain when considering discrete-time control actions (e.g., the passing of a message or any other one-at-a-time[21] act). For continuously variable (i.e., analog) control actions, behaviors such as 'too little or too much,' also exist, and these can be treated as sub-categories of pattern 2. Pattern 2 in fact encompasses many possible sub-categories, including for example:

a. Contexts in which the control action may never be safe

[20] i.e., continuing to provide it for too long, or ceasing to provide it too soon.
[21] i.e., 'atomic' in the computer science sense of the term.

b. *Contexts in which the control action has an incorrect parameter (e.g. setting an incorrect emergency frequency on a radio)*
c. *Contexts in which an insufficient, excessive, or repetitive control action may be unsafe…*
d. *Contexts in which the direction of the control action may be unsafe (e.g. providing turn left instead of turn right commands)*
e. *Contexts in which the control action has already been provided (e.g. repetitive, oscillatory, intermittent control actions)*
f. *Contexts in which the control action is provided too quickly or too slowly (e.g. ramp rate, frequency, etc.)*
g. *Etc.* [76][22]

Five elements specify any given UCA, as illustrated in Figure 21. (Note: the figure uses an example UCA related to a Brake System Control Unit (BSCU), and labels this example UCA as 'UCA-2' for identification among the full list of UCA's.)
<Source> identifies the controller that can provide the control action.
<Type> describes the pattern of the UCA (i.e. 'provides,' 'does not provide,' etc.).
<Control Action> specifies the control action in question.
<Context> lays out the context that makes the UCA unsafe.
<Link to Hazards> labels the hazards that threaten to occur, allowing traceability.

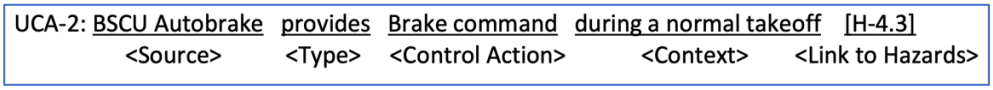


Figure 21: annotated UCA description example [76]

> *Any relevant context can be referenced in a UCA, including environmental conditions, controlled process states, states of the controller, previous actions by the controller (e.g. repetitive actions), states of other controllers, previous actions by others, simultaneous or conflicting actions, parameters or properties of the control action (e.g. a particular braking rate being programmed), or any other relevant conditions. Using words like "when", "while", or "during" in UCA construction is often helpful in developing the context. [76]*

A control action may engender UCA's in more than one (and often in all four) of the four patterns described; and may also beget multiple UCA's of each type, under different contexts. There is no limit to how many UCA's may exist for a given control action.

Each UCA describes one archetype for how its associated hazard(s) may be realized. A UCA is not analogous to a 'cause' or 'effect' as those terms are used in THA; rather a UCA is an abstract form that captures an aspect of a system's potential adverse behavior. The search for UCAs is a preparatory step to identifying hazards, and is not analogous to any part of the standard THA process; perhaps a most similar (but still conceptually totally different) analogy in THA would be an outline of prompts to brainstorming. However, in STPA the UCAs 'prompt' discovery of hazards that fit highly structured and complete characterizations, rather than areas of thought as in THA prompts.

[22] *Leading word 'Consider' removed from each, and 'c' capitalized*

This third step of STPA aims to find all UCAs within a system. To do this, each controller in the system is examined to identify and list all control actions it is capable or responsible for producing. For each control action in the list, all four potential UCA patterns ('not provide,' 'provide' (including its sub-categories), 'too early/late/out of order,' 'too long/short') are examined, to determine under what conditions each may become unsafe. Typically, one can identify several unsafe conditions under each pattern for each control action, though occasionally a control action may present for which a particular pattern only has one unsafe condition or even none. In flight test, the test matrix can provide useful contexts to consider. Each such condition under which a pattern of control action can be unsafe is documented as its own UCA, specifying all of its elements: source, type, control action, context, and relevant top-level hazards.

For example, taking an airplane wheel-brake system, and examining the 'applying brakes' control action, one can find that 'not providing' braking action could be hazardous in the context of being on the ground, and the top-level hazard might be collision. This forms a complete UCA:

Braking System Does Not Provide Applying Brakes When aircraft on ground [H1]

Context is important here[23]; in the context of being airborne there may be no hazard associated with not providing braking action.

Just as top-level hazards may be inverted to derive top-level safety constraints, so too can each UCA be inverted to derive a lower-level safety constraint. For example, inverting the UCA above, one obtains the safety constraint: 'Braking system must not fail to apply the brakes when the aircraft is on the ground.'

Illustrating a continuously variable control action, the 'applying brakes' action may be hazardous when the brakes are applied too gently on the ground, with collision again being an associated top-level hazard. This too forms a complete UCA:

Braking System Provides Applying Brakes (too gently) When aircraft on ground [H1]

Figure 22 shows example UCAs collected in a table for ease of reference. Each source's control actions occupy adjacent rows, and each pattern of UCA is captured in a column. Within each cell, all the UCAs arising from a particular UCA pattern for a particular control action are listed. Each UCA lists source, type, control action, context, and relevant hazards.

Control Action	**Not Provide**	**Provide**	**Too soon, too late, out of order**	**Stopped too soon, applied too long**
Manual flying ("MF")	(1) Crew does not manually fly Airplane while Autopilot is not armed and flight guidance is required. [H1, H2, H3] (2) Crew does not manually fly	(1) Crew manually flies Airplane while Autopilot is armed. [H1, H2, H3, H6] (2) Crew manually flies Airplane counter to required	(1) Crew begins manually flying before Autopilot is disarmed. [H1, H2, H3] (2) Crew delays manually flying after Autopilot is disarmed. [H1, H2, H3]	(1) Crew continues flying Airplane after Autopilot is armed. [H1, H2, H3, H6] (2) Crew stops flying the Airplane before Autopilot is

[23] Realistically there should be richer context in this UCA, since what really matters is when the aircraft is on the ground and the crew or an auto-braking system is commanding braking action but for purposes of illustration the example is kept simple.

	Airplane when Airplane is in a flight condition where Autopilot will not recover the Airplane. [H1, H2, H3]	guidance. [H1, H2, H3]		armed. [H1, H2, H3]
Select AHRS sources ("SS")	(1) Crew does not select appropriate AHRS sources when inappropriate sources are selected. [H1, H2, H3]	(1) Crew selects inappropriate AHRS sources. [H1, H2, H3]	(1) Crew delays selecting appropriate AHRS sources when sensor units begin providing incorrect data. [H1, H2, H3, H6]	n/a: discrete control action

Figure 22: example UCA table

Step 4: Identify Loss Scenarios

The final step of STPA identifies how or why each UCA may occur. This step delves deeply into the system and requires a higher level of subject matter expertise.

UCAs may of course occur due to failures within the system, but some may also occur even while every part of the system operates as designed. To identify the many ways any given UCA may occur, the control loops within the system are examined to determine any behaviors that could lead to the UCA occurring and resulting in a loss.

Each UCA may have one or more associated **loss scenario**(s), and this is where the constructs of STPA again become analogous to elements of THA. Each loss scenario elaborates a full set of contributions that may *cause* the UCA to occur, as well as the *effects* by which it may lead to a loss. The loss scenarios naturally include details of cause and effect in rich context, since they are derived from the workings of the system under analysis. Each UCA may have any number of loss scenarios, each describing one way the UCA may come about and lead to loss. Each loss scenario reveals a hazard in context together with its causes and effects; these may be teased apart to fit the format of THA or taken collectively as they are produced.

Background in legacy accident models and hazard analysis methodologies may make it temping to view loss scenarios as 'chains of events' that lead to accidents; however, this is a restrictive view. Loss scenarios are not predicated on sequential or linear chain-of-events thinking, but rather are based in systems theory. Therefore, loss scenarios are able to capture the kinds of simultaneous or interactive behaviors that have made complex systems historically difficult to analyze. A loss scenario facilitates all the contributors to a loss being captured and described in unison without temporal or sequential restrictions.

Each loss scenario is tagged with its relevant UCA, thereby maintaining traceability back through the many steps all the way to top-level hazards and losses.

Finding loss scenarios for each UCA involves taking that UCA for granted and asking 'how could this happen.' Typically there will be many plausible ways, and again here unlike the freeform brainstorming of THA, there are a variety of more and less

structured approaches that may be used. Generally, both the forward path (i.e., control pathway) and the backward path (i.e., feedback pathway) should be considered, as well as the process model, the control model, and other inputs (controls or feedbacks) that the controller may receive.

A detailed methodology is under development [83] that breaks control loops into constituent sections of control, actuation, sensing, and feedback to further sub-categorize loss scenarios within each UCA. Under this framework, four classes of loss scenario exist:

1. The controller receives correct feedback, but issues the unsafe control action.
2. The controller receives faulty feedback, and issues the unsafe control action.
3. The controller issues a safe control action, but the controlled process receives an unsafe control action.
4. The controlled process receives a safe control action, but produces an unsafe result.

Together, these completely capture the combinations of inputs and outputs possible for an unsafe outcome in a control loop. Further development based on this structure provides sets of simplified archetypal questions, the answers to which produce complete loss scenarios. "The "Why" answers [to these questions] may come from SMEs [subject matter experts], not the STPA practitioner."[83]

As a worst case, however, assuming freeform brainstorming is used for loss scenario discovery, this brainstorming still occurs within the structured environment of individual UCAs, and therefore stands to be at least as thorough as THA if not more.

3.2. Harnessing Abstraction to Bring Hazards into Focus

Abstraction is a powerful tool for understanding systems at different levels, and for exposing or suppressing detail and workload. Generally, it can be useful to analyze a system at a level of abstraction coincident with the level of behavior of interest. In order to choose an appropriate level of abstraction, one must first specify the purpose of the analysis.

Returning to the beehive example in Section 3.1 to illustrate abstraction, if one desires to analyze bee-to-bee communication dances, it may indeed be useful to model each neuron (though even here this seems perhaps excessively detailed); if one seeks to analyze hive temperature regulation or defense response to large predators it may be sufficient to model at the highest level of abstraction (a collection of bees and a physical hive); and if one seeks to analyze population regulation, then an intermediate level of abstraction (groups of bees and parts of hive) may be most useful.

Similarly in flight test, if one is concerned with mid-air collision hazards, it may not be necessary to model each aircraft down to its components; whereas if one cares about cabin temperature exceedances such a level of detail may indeed be desirable.

Performing analysis at a high level of abstraction can sometimes make focal certain system-wide behaviors that might be difficult to see when analyzing at a more detailed level, and vice versa. Thus, the choice of abstraction levels may be used proactively to expose particular areas. Together with system boundary selection, such choices can be used to deliberately focus analysis as desired and control workload. Sometimes it may be beneficial to carry out parallel analysis at different levels of detail.

The more abstractly a system is represented, the more widely its system boundary may be drawn while still maintaining a reasonable amount of salient content. At the highest level of abstraction, representing each organization as a collective abstract element, it may be reasonable to model an entire government or agency– a wide system boundary. Detailing each employee of even a post office would suggest drawing a much narrower system boundary around only the post office. While it is possible to model an entire government at an individual level, doing so could be onerous. Similarly in flight test a balance can be struck between abstraction level and system boundary inclusiveness to yield useful analysis balancing salience, workload, and detail.

Certain archetypal levels of abstraction may serve as useful examples or templates for flight test modeling and analysis; any model should always be adapted and matched to the task at hand. Figure 20 in Section 3.1 depicts a representative day-of-test system model. The level of abstraction is moderate, with each aircraft, each crew, and each supporting function modeled as its own element; and the system boundary (implicit, including all elements depicted and none not depicted) is fairly tightly scoped to those participants involved on the day of flight.

Analysis for flight test might likewise care about on-aircraft behaviors. For this one can 'zoom in' on the day-of-flight example to focus only on the test aircraft and its crew, but in finer detail than before (simultaneously narrowing the system boundary and decreasing the level of abstraction). Still, it may be onerous and unproductive to model every rivet and nut, but it may be desirable to model salient aircraft systems as their own elements. Figure 23 shows an example such model, with each crew member, the flight control system, the hydraulic system, the flap system, and their flight deck controls each modeled as its own element.

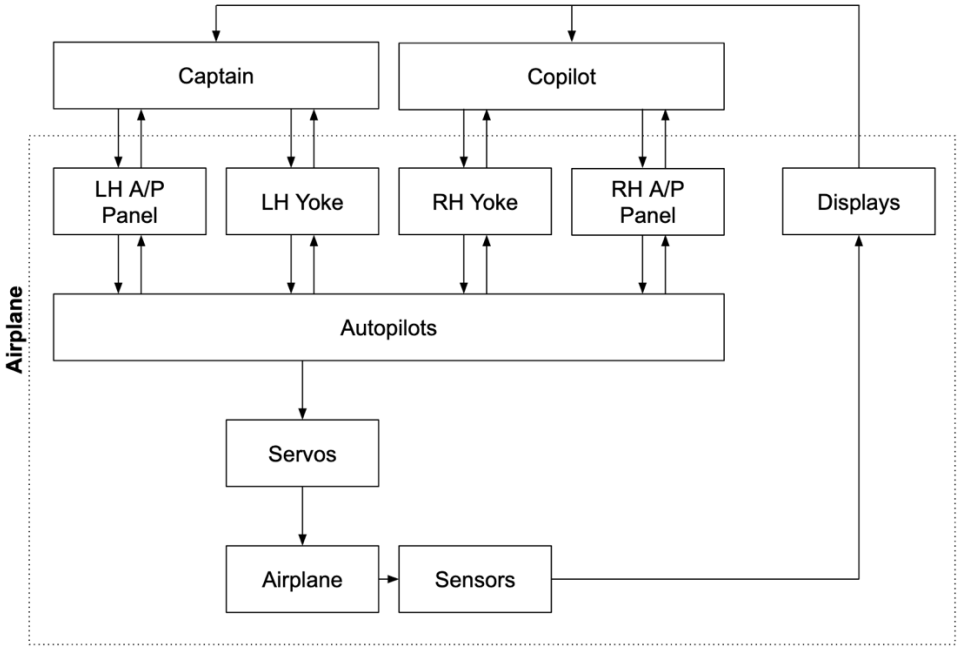


Figure 23: 'On the aircraft' control structure, 2 crew

It is, of course, possible to continue 'zooming in' to analyze at lower and lower levels of abstraction, theoretically limited only by particle physics. At the lower levels, however, it may be unproductive to model sub-elements of a human, and modeling of technical elements begins to describe individual components. Even somewhat above this level, analysis may depart from flight test-relevance and enter the domain of engineering and design systems safety. STPA can be exceedingly fruitful here too (see, for example, [84], [85], [86]), however for purposes of flight test it is worth noting this is not necessary.

Analyses passed along by other organizations, whether using legacy methodologies or modern ones, may be readily folded in to STPA to produce informative results (up to the limit of the correctness of such other analyses of course). A boundary may be drawn between STPA and other forms of analysis with a useful and productive interface between the two. Using STPA as the highest-level framework, it is easy to fold in lower-level analyses at whatever level their results yield. For example, an off-the-shelf actuator may have been certified using FMEA; the actuator may be modeled in STPA as an element with the same behaviors and outcomes expressed in the FMEA analysis, rather re-analyzing a potentially black-box system using STPA.

Stepping back and 'zooming out', it may be desirable to analyze organizational safety review and test process behaviors. Broadening the system boundary and increasing the level of abstraction from the day-of-flight example, it becomes possible to model of an organization responsible for flight test continuation criteria and real-time data monitoring, and to analyze its processes and hazards.

Similarly, abstract control structures may be useful in analyzing organizational processes and roles involved in configuration management, build-up planning, or other high-level functions of flight test safety. Gulfstream's "existing SOPs [at the time of the G650 crash discussed in Section 1.1] were ineffective in "...establishing control gates for key decision points, implementing processes for validating engineering methods, and clearly defining roles and responsibilities for on-site test team members."[87] Such functions lend themselves readily to analysis and redesign using STPA.

3.3. Testing and Violating Safety Assumptions

It is clear that "flight test organizations put systems through extreme and unusual scenarios, and as such… have ways of looking at use of systems a bit different from the designers."[88] The safety case of any aircraft or product designed for operational service will be based on particular assumptions of usage, operation, crewing, redundancy characteristics, and more. However, while some of these assumptions may hold true under flight test, others may be violated– deliberately or unwittingly– whether by test technique, test point selection, emergency and abnormals testing, certification demonstration requirement, or otherwise.

Some safety assumptions are explicit, for example flight envelope limits, procedure usage in response to particular scenarios, and certain human operator reactions[24]. These may be easily identified and fairly readily enforced. Violations of

[24] E.g. see FAA Advisory Circular 25.1309-1A "System Design and Analysis", paragraph 11(a) [89]

these assumptions for testing may be made deliberately, and with appropriate mitigation and acceptance of residual risk.

Others may be implicit, such as periodic power cycling or proper execution of certain procedures. These can be difficult or impossible to identify by legacy methods, and therefore often revealed unexpectedly during test, but possible and sometimes even easy to capture by STPA. Only once they are identified can they be intentionally enforced, or purposely violated for test. Otherwise they may present unidentified hazards and unwitting risk.

Yet others occupy a 'gray area' between implicit and explicit, perhaps not often examined, and based on historical non-violation, such as correct usage of certain controls or interpretation of displayed text, or certain elements of ConOps. By providing a holistic framework, STPA facilitates intelligent analysis of these as well.

In the course of testing and demonstrating a product, it may be necessary to perform certain UCAs and violate certain design safety constraints. Flight envelope or other limits may be pushed beyond nominal for certification (e.g., V_{NE} testing); some may be tested to failure criteria; and others may be unknown within some bounds or miscalculated in engineering.

Elements of systems architecture and redundancy may need to be disabled, or may be yet-untested and therefore unavailable. Crew training and mental models may not be appropriately updated to the situations at hand if analysis results are improperly conveyed or assumptions made about capability or primacy. Procedures may need to be tested for resiliency to incorrect implementations. Where complex systems are involved, even small variations can produce large changes in behavior

Where UCAs will be performed, it is critical to recognize this will be the case, and that proper analysis and appropriate mitigations are considered. Legacy methods have built up 'cheat sheets' and prompts to help try to identify such cases[25], but these again are patchy and historically-oriented by necessity. Using methodical systems analysis tools like STPA it becomes possible to holistically recognize and analyze such situations.

Stepping back to the organizational and procedural level, assumptions of what analysis has been done by whom and when, and their communication of results, are also of relevance. These, too, can be analyzed using STPA. By using STPA at a level of abstraction that brings to focus those parts of the organization responsible for supporting buildup planning and go/no-go decisions, organizational processes can be analyzed for their safety contribution. Such analysis enables design of appropriate data-driven organizational processes to mitigate hazards associated with defining inappropriate continuance criteria or overstepping of them in operation. Real-time data monitoring with appropriate process control can boost safety in test.[7]

In analyzing systems for flight test using STPA, errors in design or design assumptions may sometimes be revealed. Similarly, STPA may reveal legacy test point design or test data selection to be off-base. Whereas in non-complex systems it can be practical to explore a bounded set of test points that collectively reasonably characterize a system, with complex systems this is often prohibitive. Instead, however, analysis of the

[25] For example, one prompt asks: "If any safety device or interlock will be bypassed or overridden in these tests, what additional hazards are involved and what steps will be taken to reduce these risks?"[48]

system by STPA reveals key points in control of the system, and thus it may be possible to "design a test point matrix to assure the controls on system performance."[49]

3.4. Hazard Mitigation and Analysis of Changes

When mitigating identified hazards in preparation for flight test, any changes made to alleviate one hazard can introduce, alter, or eliminate other hazards. Especially where complex systems are involved, it can be desirable to re-analyze the system as changed. In THA this is done on an 'as-needed' basis by intuition, brainstorming, and experience. With STPA it becomes possible to do so more methodically.

One challenge in re-analyzing changes to a large system, as with hazard mitigations in flight test, is in the scope. A naïve approach requires re-analyzing the entire system when any change is made– a potentially laborious process, especially when multiple changes may be made at different times. Therefore, it is desirable to minimize the re-analysis required for any given change without loss of completeness. "The ability to determine what parts of the analysis need to be revisited after a change is made can help assist with resource planning and engineering change management procedures."[90] Doing so is particularly challenging in complex systems, where emergent behaviors are not simply attributed to direct causes.

Reducing the scope required in re-analysis may lower the barrier to execution enough to facilitate multiple cycles of methodical design and re-analysis, as illustrated in Figure 24. "Focusing efforts only on selected portions of the analysis will reduce the amount of effort and time to perform STPA in the overall design/analysis iteration."[90]

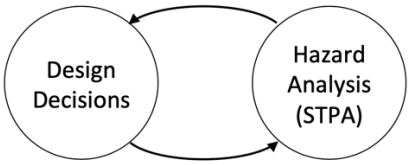


Figure 24: Design/redesign and analysis/re-analysis cycle [90]

Using STPA it is possible to employ tagging and traceability to comprehensively direct re-analysis to only those UCAs and loss scenarios effected. Such an approach is described by Sgueglia [90], as summarized in Figure 25. When modifications (i.e., design decisions) are made, their effects on the control structure are identified, and the control structure is updated. Then the high-level hazards reexamined, the impacts on STPA UCAs and scenarios are analyzed, and the relevant items selected for re-analysis. Such re-analysis then informs any follow-on design decisions, and the cycle may be repeated.

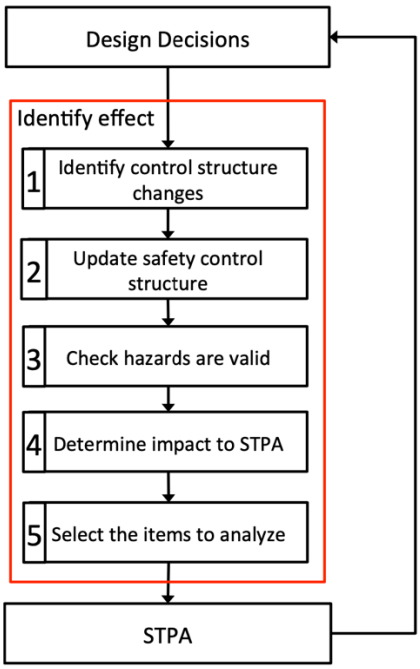


Figure 25: Re-analysis process following modification [90]

If a modification changes any of the elements of a UCA (source, type, control action, context, or hazards; see Figure 21), then that UCA and its associated loss scenarios must be reconsidered; likewise the introduction of new UCAs by such modification must be considered. Figure 26 summarizes re-analysis actions suggested following removal, modification ('change'), or addition of various system element types.

> *Changes to [a] control action may result in new [UCAs]. For example, if [a] signal is changed from discrete to continuous, the 'provided too long' or 'stopped to[o] early' unsafe control action types should be examined to verify if they could lead to a hazard. The context of [a UCA] could be impacted by changes to [its] controller's process model, inputs, or environmental conditions. Context added that was not initially considered might lead to new unsafe control actions. Generally, any changes to [a UCA] will require [re-analysis of its] existing causal scenarios and [to] identify any new ones… Design changes made to the control side should be investigated for how a safe control action being provided but not followed could occur. Similarly, changes made to the feedback side of the control loop can limit the analysis to causal scenarios related to the other unsafe control actions. [90]*

By comprehensively tagging what system elements are implicated in each loss scenario, it should be possible to search for scenarios impacted by changes to or removal of these system elements. When new elements are added, UCAs and loss scenarios generation must be performed; additional analysis is also required to update any existing loss scenarios that may be affected.

> *[An existing] scenario may [remain] valid [following a modification] but with a new causal factor, such as [with the addition of a second sensor to the system], one related to a common cause failure of the two sensors. Causal scenarios that were not related to the design change may need to be verified for the new causal factor. An example*

of this could be when a second controller is added. Previously there was not a causal factor related to conflicting commands from multiple controllers to the controlled process but now this becomes a potential cause of an unsafe control action. [90]

Control Structure Element	Modification	Impact to STPA to consider	
		UCAs	Loss Scenarios
Controller	Remove	Results can be eliminated	Exclusive causal scenarios eliminated
	Add	Must be performed	Must be performed
Control Algorithm	Change	Use previous result	Check previous and newly related scenarios
Process Model	Change	Check unsafe control actions	Check previous and newly related scenarios
Control Input	Change	Check unsafe control actions	Check previous and newly related scenarios
	Remove	Check unsafe control actions	Check previous and newly related scenarios
	Add	Check unsafe control actions	Must be performed
Control Action	Change	Must be performed	Must be performed
	Remove	Results can be eliminated	Exclusive causal scenarios eliminated
	Add	Must be performed	Must be performed
Actuator	Change	Use previous result	Check scenarios for safe but not followed
	Remove		Exclusive causal scenarios/ factor eliminated
	Add		Check scenarios for safe but not followed
Sensor (Feedback), Controlled process (Inputs, Outputs), Disturbances, Other controllers	Change		Check previous and newly related scenarios
	Remove		Exclusive causal scenarios eliminated
	Add		Must be performed

Figure 26: Re-analysis guidance following modification (modified from [90])

Further work is required to properly illustrate this process for manual usage.

4. Example Use of STPA to Identify Flight Test Hazards

"However vast the darkness, we must supply our own light."

– *Stanley Kubrick, Playboy* interview*,* Sept. 1968*,* as quoted in [91] (used poetically, out of context)

In order to demonstrate the use of STPA for flight test hazard identification, STPA analysis was carried out on an actual flight test campaign– a development flight test initiative to inform the development of certification criteria for Automatic Ground Collision Avoidance capability in civil aircraft.

The Automatic Ground Collision Avoidance System (Auto-GCAS, or simply AGCAS) is an auxiliary safety system presently operating on some fighter jets. It is intended to preclude losses caused by Controlled Flight into Terrain (CFIT). To do this, it monitors for imminent projected ground collision, warns the pilot upon crossing set thresholds, and, in case of no pilot response, applies corrective action to the flight controls to avoid flight into terrain. The military has been operating the system on fighter airplanes since 2014, and has been pleased with its record of reducing what had been the most common cause of death in F-16's.[92]

In General Aviation, CFIT has accounted for 17% of all accident fatalities.[93] Following military success with AGCAS, studies are now underway to determine how to certify AGCAS systems onto general aviation aircraft, where the approval criteria, regulatory process, and aircraft characteristics are all different.

Aircraft performance, systems architecture, equipment, and implementation differences between civil and military fighter aircraft introduce meaningful differences in system capability and behavior. To support development and inform civil certification requirements for AGCAS capability in civil aircraft, a development flight test campaign seeks to characterize system behaviors and advance maturity.[94], [95], [96]

Based on personal experience using THA for AGCAS on military programs, some have concluded that "THA fails to identify relevant hazards" in such a setting.[97] Therefore, no THA was performed on this test campaign; instead, the project test pilot's prior experience testing military Auto-GCAS systems was relied upon to assure test safety. STPA, by contrast, was able to expose meaningful hazards in this campaign.

This section demonstrates the application of STPA on this flight test campaign for the purpose of identifying flight test hazards. The application is an illustrative example, showing a template for how such analysis may be carried out and the types of work product produced from it. For demonstration purposes it is scoped to consider entire flights and the entire test campaign. Representative parts of the analysis were taken to completion in order to illustrate what is possible and how it may be accomplished, however a full analysis was not completed. The outputs may be viewed as one 'recipe' for applying STPA to flight test hazard identification, and should not be considered a completed analysis of Auto-GCAS. Rather, they may be taken as an illustrative example of how an organization could apply STPA to Auto-GCAS to identify flight test hazards.

4.1. Auto-GCAS Testing for General Aviation

The high-level theory of operation of Auto-GCAS is as follows: flight trajectory is tracked and projected forward in time. The forward-projected trajectory is evaluated for intersection with terrain or obstacles, which are stored as a database. Where an intersection is found, the time to intersection is calculated, and when such time is evaluated to be less than a 'caution time' threshold, Auto-GCAS alerts the flight crew to the impending collision; when the time to intersection is evaluated to be too close to the minimum time that would be required to avoid collision given the maximum-capability dynamics of the aircraft, then Auto-GCAS wrests command of the aircraft (warning the pilot it is doing so) and executes a 'fly-up' maneuver to guide the aircraft to recovery safe from terrain and obstacles.

Figure 27 shows a conceptual schematic of the envisioned aircraft control. In the interest of easing adoption in the General Aviation (GA) fleet, a civilian implementation of Auto-GCAS is envisioned to utilize commercially-available off-the-shelf (COTS) products to the extent practical, such as existing certified navigation sources, autopilots, Terrain Awareness and Warning Systems (TAWS), and cockpit controls and displays.

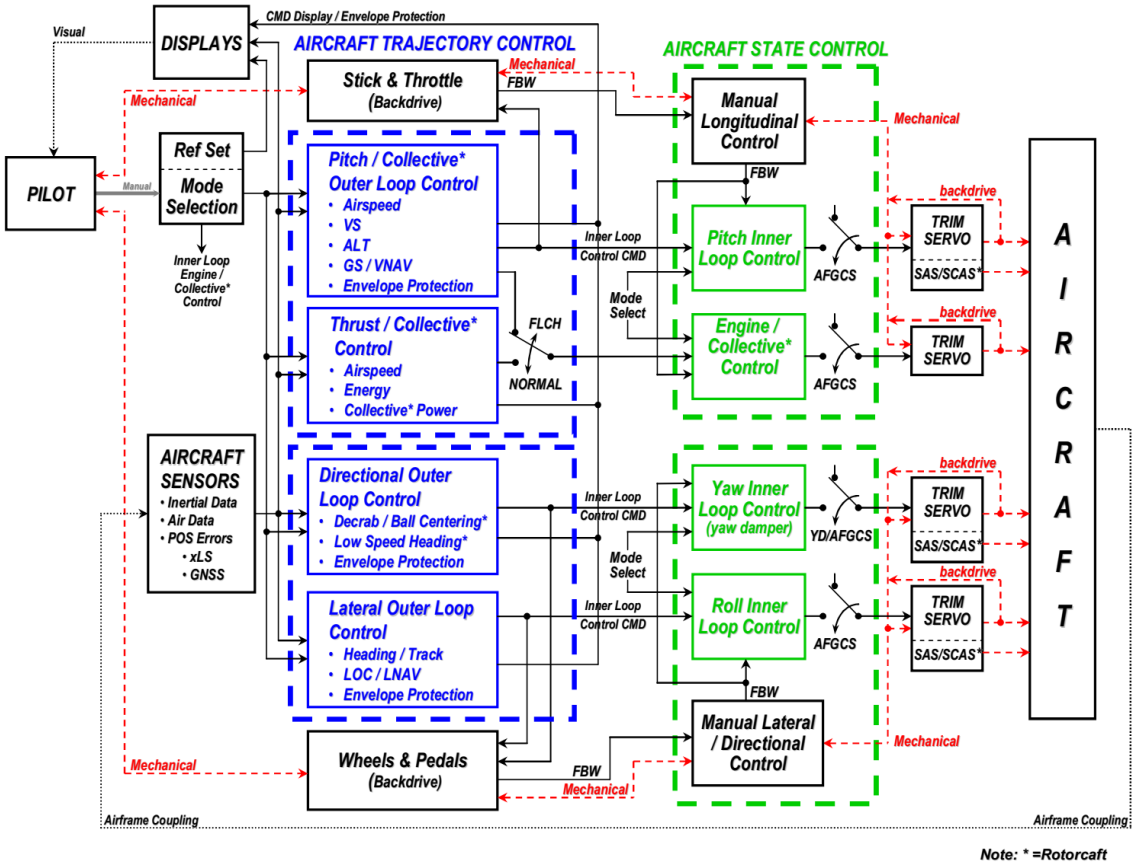


Figure 27: Auto-GCAS flight test system diagram [94]

The objectives of the test campaign are to assist in determining a civilian certification framework for AGCAS and to demonstrate whether a low-authority

autopilot can provide nuisance-free protection. In the course of testing, revisions and tuning adjustments are made to the AGCAS system.

The flight test campaign uses the following equipment to implement the full Auto-GCAS capabilities:

- Garmin G3X Flight Display Unit (built-in navigation and autopilot capabilities)
- Garmin flight control servos (commanded by G3X)
- Intel NUC13 Mini PC running Linux and hosting the AGCAS software
- Bespoke auto-throttle system (commanded by AGCAS software)

A tablet-based flight test visual display is additionally available in the cockpit, and can also produce aural alerts.

A cockpit switch controls power to the autopilot servos and auto-throttle system. Another cockpit switch controls power to the Mini PC running AGCAS.

The AGCAS software receives flight data (position, orientation, airspeed, etc.) over RS-232 from the G3X, and transmits commands to the autopilot by emulating a Garmin GMC 307 autopilot control panel, also over RS-232. The AGCAS software runs its own terrain database, as well as all AGCAS functionalities. (The built-in Garmin TAWS functionality runs independently, but is advisory only, and will be ignored for the purposes of this analysis.) When the G3X and AGCAS software are each booted up, they engage in a communications 'handshake' and then all systems are ready for use.

The AGCAS software is able to control the autopilot in the following ways: select airspeed climb ('FLC') mode, select heading ('HDG') mode, set airspeed bug to V_Y (85 KIAS), move heading bug, move altitude bug, and engage ('arm'). These together are sufficient to command the fly-up maneuvers. At the termination of the fly-up maneuver the AGCAS stops commanding updates to autopilot settings, leaving the aircraft in a straight-ahead autopilot climb. As presently implemented, all commands and the movement of bugs to desired values are open-loop and by dead-reckoning; however, real-time autopilot parameters are received by AGCAS, so closed loop control could be implemented easily in the future.

Upon initiating a fly-up maneuver, the AGCAS system sets the autopilot vertical mode to 'airspeed climb' and its lateral mode to 'heading'; and sets the airspeed bug to best rate of climb speed V_Y, the heading bug to the desired final heading, and the altitude bug to the present bugged altitude plus a large margin; and then engages ('arms') the autopilot.[26] It also engages the auto-throttle. Configured as described, the autopilot guides the airplane to the selected heading by turning in the direction in which the heading bug was moved until reaching and maintaining the bugged heading, and pitches to and maintains the bugged airspeed until reaching the bugged altitude, at which time it transitions to altitude-hold mode to maintain the bugged altitude. The auto-throttle advances the throttle to full power while observing limits to prevent propeller overspeed.

An 'autopilot disconnect' switch on the control stick disengages the autopilot when pressed; it also disengages the auto-throttle. Manual control of all remaining autopilot modes and settings is via a menu page in the G3X. Manual control of auto-throttle is not provisioned aside from disengagement as described; only AGCAS is able to command auto-throttle engagement.

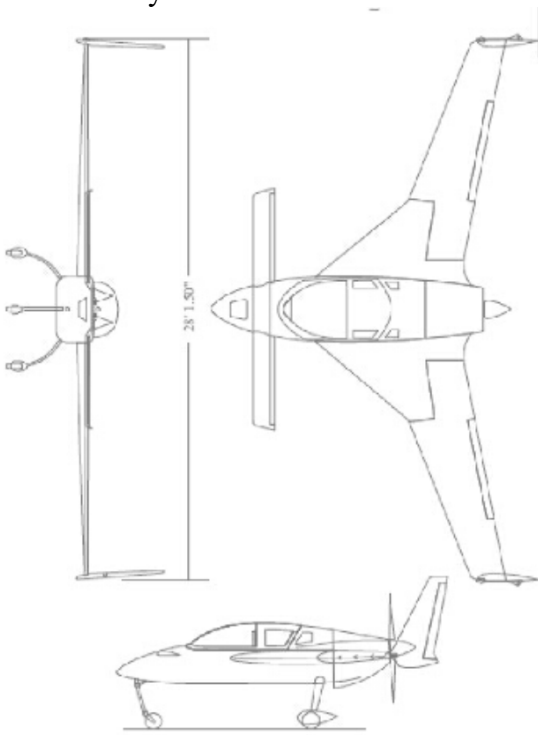
[26] In order to set each bug, the AGCAS system increments their values. The heading bug is synchronized to present flight heading before incrementing its value.

Auditory alerts are sent by the autopilot and auto-throttle when each is disengaged, and the autopilot additionally provides visual feedback via the G3X display. The AGCAS software sends the following visual and aural alerts:

- Fly-up pending (below 'caution time' threshold)
- Fly-up initiated (when fly-up is initiated)
- Fly-up complete (when fly-up maneuvering has completed)

No alert silencing functionality is presently implemented.

Testing is carried out using an Experimental Amateur Built (E-AB) Cozy Mk-IV homebuilt aircraft. It is a four-seat, fiberglass composite construction, single-engine pusher canard configuration airplane of moderate performance. Figure 28 shows a 3-view drawing of the airplane and key dimensions.



DIMENSIONS

Wing Span / Area	28.1 ft (8.6m) / 88.3 sq. ft (8.2 sq. m)
Canard Span / Area	12.1 ft (3.7m) / 13.1 sq. ft.(1.2 sq. m)
Total Wing Area	101.4 sq. ft. (9.4 sq. m)
Length	17.0 ft. (5.2 m)
Height	7.9 ft. (2.4 m)
Cockpit Width	
Front	42.0 in (1.07 m)
Rear	38.0 in (0.97 m)
Cockpit Height	
Front	38.0 in (0.97 m)
Rear	37.0 in (0.94 m)
Cockpit Length	
Front	70 in (1.78 m)
Rear	54 in (1.37 m)

Figure 28: Cozy Mk-IV test platform [96]

A single pilot manages all aircraft control and systems. A Flight Test Engineer (FTE) may optionally occupy a second seat to provide additional monitoring of safety parameters, ensure data collection, brief flight cards, and provide 'quick look' data analysis.

Flights are flown in Military Operations Area airspace in southern California. Air Traffic Control contact is not required.

The planned test matrix is large, with many test points stemming from guidance materials found to be relevant to the certification pathway under consideration. Included are several categories of investigation, including normal and abnormal conditions. The following summarizes some of the interesting aspects to consider in the analysis.

1. Normal flying (no intentional interference with terrain)
    a. Enroute & terminal area phases of flight
    b. Autopilot disconnects
2. Flying toward terrain/obstacles
    a. Normal function with AGCAS activation
    b. Escape maneuver overrides (normal & abnormal conditions)
    c. Icing shapes

4.2. STPA in Action: A Walk-Through

Walking through the four steps of STPA as applied to Auto-GCAS produces the following analysis:

Auto-GCAS STPA Step 1:

The first step in STPA is to define the purpose of the analysis. Here the objective is to identify test hazards in the flight testing of Auto-GCAS as described, taking each entire flight and the entire campaign as within scope. The system can be defined as the aircraft, the autopilot, the auto-throttle, the AGCAS logic, the associated terrain database, the navigation/attitude and heading reference system (AHRS), the pilot, and the operator who updates the terrain database. The FTE will be ignored for simplicity, though they could be added in later. ATC is unneeded and placed outside the system boundary. And the Database Manager who revises and disseminates the terrain database is likewise placed outside the system boundary for simplicity.

To keep the present analysis light, only two core losses are contemplated:

- L1: Injury or loss of life
- L2: Damage to or destruction of aircraft

Three top-level hazards are considered that could lead to these losses (as tagged):

- H1: Aircraft gets too close to terrain or obstacles (L1, L2)
- H2: Aircraft gets too close to other aircraft (L1, L2)
- H3: Aircraft exceeds safe operating envelope (L1, L2)

These hazards engender three corresponding top-level safety constraints:

- C1: Aircraft shall maintain safe separation from terrain and obstacles.
- C2: Aircraft shall maintain safe separation from other aircraft.
- C3: Aircraft shall remain within its safe operating envelope.

The safety constraints must be successfully enforced throughout flight testing to assure a safe outcome.

Auto-GCAS STPA Step 2:

For purposes of revealing flight test hazards, AGCAS flight testing is modeled at a level of abstraction that exposes top-level system interactions while encapsulating implementation/engineering detail. Many like functionalities are lumped into abstracted elements, and interacting functionalities modeled as separate elements even where implemented together.

Figure 29 shows a resulting control structure. The Pilot can control the Airplane directly by issuing 'guide & control' type control actions; can control Autoflight by issuing 'arm/set targets' type control actions; and can control AHRS/Nav by issuing 'select sources' type control actions. The Pilot receives feedback from the Airplane in the form of 'aircraft motion, status, etc.' information; from Autoflight in the form of 'status & alarm' information; from AGCAS Logic in the form of its own 'status & alarm' information; and from AHRS/Nav in the form of 'location, orientation, trajectory' information.

Autoflight can control the Airplane by issuing 'guide and control' type control actions; it also may send feedback to AGCAS Logic in the form of 'status/errors' messages. Autoflight receives many forms of feedback as depicted.

The control structure depicts the control and feedback pathways of each of the elements, together with descriptive labels. ATC and the Database Manager are shown in gray for context, but excluded from the present analysis. For simplicity, the FTE is not depicted or included in the present analysis.

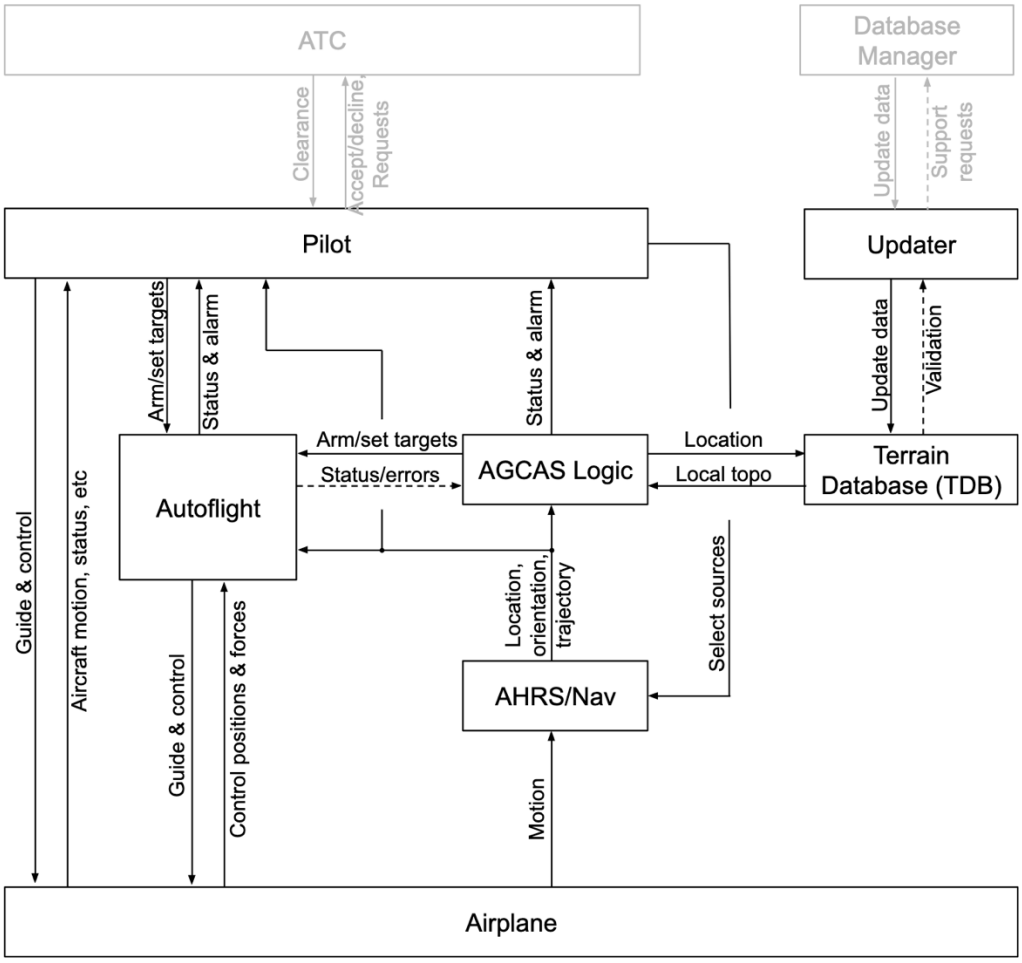


Figure 29: Auto-GCAS flight test control structure

Auto-GCAS STPA Step 3:

As depicted in the control structure of Figure 29, controller elements for analysis include:

- Pilot: a test pilot operating the aircraft.
- Operator: a person who updates the terrain database. This may be the Pilot or any other person.
- Autoflight: the G3X autopilot functionality, together with the Garmin servo units, as well as the auto-throttle system.
- AGCAS Logic: all AGCAS functionalities, including trajectory prediction, terrain evaluation, time-to-intersection calculation; and escape trajectory generation, evaluation, selection, and execution, including control of the autoflight components.

Additionally, the following controlled processes are considered:

- Terrain Database (TDB): a database of geographic terrain and obstacle elevations.
- AHRS/Nav: all navigation and flight condition measurement and estimation, including location, velocities, orientations, rates, and air data and its derivatives.

- Airplane: the Cozy Mk-IV including its flight controls, powerplant, avionics, and all systems; excluding only those explicitly included in the foregoing.

A listing of control actions associated with each controller is shown in Figure 30.

Controller	**Control Actions**
Pilot	Power the autoflight
	Un-power the autoflight
	Power the NUC13 (‘Power AGCAS’)
	Un-power the NUC13 (‘Unpower AGCAS’)
	Arm autopilot
	Disarm autopilot
	Disarm autoflight (incl. auto-throttle)
	Set autoflight modes
	Set autoflight bugs/setpoints
	Manual flying of airplane
Updater	Load terrain database to memory
	Delete terrain database from memory
Autoflight	Fly airplane
	Adjust throttle
	Alert crew
AGCAS Logic	‘Push’ autopilot arm button
	'Push' autopilot "FLC" (airspeed climb) mode button
	'Push' autopilot "HDG" (heading) mode button
	Move autopilot airspeed bug nose up
	Sync autopilot heading bug
	Move autopilot heading bug
	Move autopilot altitude bug
	Arm auto-throttle
	Request local terrain from TDB

Figure 30: Listing of controllers’ control actions

To examine each control action in each of the four potential UCA patterns (‘not provide,’ ‘provide’ (including its sub-categories), ‘too early/late/out of order,’ ‘too long/short’), a table was constructed listing each control action on a row, and each UCA pattern in a column; an excerpt is shown in Figure 31. Thus, each cell in the table represents one control action under one UCA pattern, and filling in these cells provides a methodical way to explore the space of UCAs.

Control Action	'Not Provide'	'Provide'	'Too early / too late / out of order'	'Too long / too short'
Pilot:				
Power the autoflight				
Un-power the autoflight				
Power AGCAS				
Unpower AGCAS				
Arm autopilot				
Disarm autoflight				
Set autoflight modes				
Set autoflight bugs/setpoints				
Manual flying of airplane				

Figure 31: UCA table structure excerpt

With the framework of such a table in place, UCAs were identified and documented in their respective cells. Test conditions from the test matrix were contemplated while identifying UCAs, and conditions not listed in the test matrix but associated with a documented UCA were monitored to be noted as potential additions to the test matrix. A partial listing of UCAs identified is in the Appendix. Each UCA was tagged with its associated top-level hazards for traceability.

With a total of 24 possible control actions between the four controller elements, and four possible UCA patterns of each, even merely a preliminary analysis with limited subject matter expertise identified 143 UCAs in total. Of those identified, some stem from flight test related contexts, some from test-specific equipment, and others from everyday contexts and normal equipment.

Auto-GCAS STPA Step 4:

With UCAs documented, loss scenarios were identified. Due to time constraints, the structured loss scenario identification method of [83] was not used; the results of the present analysis might be extended by such a further detailed process.

Loss scenarios naturally spanned both test unique hazards and system design hazards. In development flight test, however, where systems may not yet have reached maturity, "many system engineering hazards can be mitigated to reduce test hazards where they wouldn't be acceptable as a finished product for the public."[98] Selected loss scenarios for two representative UCAs are listed in the Appendix.

Some of the loss scenarios point to procedural hazards. One loss scenario under the UCA 'Crew does not power Autoflight system when fly-up will be needed. [H1]', for

example, suggests, “After recovery from an abnormal condition by unpowering Autoflight, the crew fails to power Autoflight again before setting up for the next AGCAS activation test point.” And a loss scenario under the UCA ‘Crew does not power AGCAS computer with sufficient time to initialize before fly-up is needed. [H1]’ suggests, “Crew initiates checklist in which AGCAS power switch is to be turned on, immediately before setting up a test point involving AGCAS activation, but does not allow sufficient time for initialization before proceeding with the test point.”

Other loss scenarios reveal potential human-machine interaction design hazards. For example, one scenario under the UCA ‘Crew sets Autopilot mode(s) while Autopilot is armed and AGCAS Logic is activated. [H1, H2, H3]’ suggests, “immediately before AGCAS fly-up activation and flying on autopilot, the pilot decides to change autopilot modes to avoid terrain; however, by the time they reach the autopilot settings menu the AGCAS has activated, so when they do adjust the modes they overwrite part of the AGCAS fly-up modes and cause flight into terrain.”

Yet other loss scenarios reveal configuration and tuning hazards. One loss scenario under the UCA ‘AGCAS Logic does not ‘push’ the autopilot ‘arm’ button when the airplane enters Evasive Condition [requiring immediate evasive action to avoid impact with terrain] and the autopilot is not armed’ suggests, “the AGCAS Logic does not detect the Evasive Condition because the dynamics model it uses does not match the actual performance of the airplane for the condition in question, and therefore does not attempt to find or execute an escape trajectory, thus not warning the pilot or arming the autopilot.”

4.3. Observations

The first two steps of STPA were reasonably accomplished with only a moderate familiarity with the system and the test objectives, and already provide value to the team with relatively little investment. The control structure diagram and the definition of high-level losses, hazards, and constraints facilitate deepened shared understanding of the system and streamlined communication among members of the test team.

The third step of STPA required further consultation with SMEs to assemble the lists of control actions, and further to elucidate UCAs. Many more UCAs likely exist in the Auto-GCAS testing scenario than were found and listed here; these would be more easily identified in a collaborative environment and with an expert STPA facilitator.[99] Even so, the range of UCAs identified by a preliminary solo analysis already revealed a broad range of hazard areas, and a structured way to investigate them.

It can be tempting to define UCAs in terms of logic implementation or technical parameters. This can drastically shrink the space of hazards it may capture. For example, one UCA was initially written as ‘AGCAS Logic does not sync heading bug when initiating straight-ahead fly-up. [H1],’ however this excludes the broader context wherein the heading bug should be synced wherever required to avoid collision with terrain and obstacles. As much as possible, UCA context should be framed in terms of overall system behavior. An improved version of this UCA reads, ‘AGCAS Logic does not sync heading bug when necessary to avoid terrain. [H1].’ The more general the context, the more encompassing the UCA.

The fourth step of STPA captured even multifactorial, nonlinear hazards with ease. Loss scenarios naturally span a range of categories, and no straightforward way was found to separate test unique scenarios from the others. This, however, is no worse than in THA, where test-uniqueness must likewise be evaluated on a case-by-case basis. Unlike THA, however, STPA provides contextual insight into each hazard that may permit more nuanced distinctions and more actionable follow-up. Again here, further input from SMEs, a collaborative environment, and the help of an expert facilitator could boost results beyond those shown here.

Intuitive hazards, such as crew hypoxia due to an unchecked climb following a fly-up, were used to 'sanity check' interim results. By carefully placing these into the existing STPA framework, rather than attempting to change the framework to fit the intuitions, a more complete revised result was achieved.

Finally, while achieving comprehensive coverage of UCAs and loss scenarios could prove time consuming, this may be due to the number of hazards addressed and their complexity.

4.4. Discussion

Several points stand out in this example. First, in order to exploit the full bottom-up advantages of STPA it is helpful to perform the steps of the analysis sequentially, thus building up a framework. Working backwards can result in intuitions dressed in the appearance of STPA, which are not very useful, whereas working forwards creates a holistic framework in which intuition can later be used to spot-check and improve.

The entire team and program management should agree as to the losses of interest and the top-level hazards. The on-the-ground team should all agree on a control structure model of the system. While generation and documentation of UCAs and loss scenarios may be easily coalited from disparate contributors across separate sessions, the losses, hazards, and control structure require concurrent consensus across the team.

UCAs should be stated in terms of system characteristics and behaviors as much as possible. While technical details can be important to implementation, behaviors, controls, and feedbacks are key to hazard identification by STPA.

Coming from a background of THA, it can be tempting to attempt to carry out the first steps of STPA with the objective of proving existing intuitions. This should be avoided not only to avoid biasing the process, but also because it is difficult given the nature of complex systems to derive clear direction in initial process from an expected final result. Instead, only in the final step of loss scenario generation should intuitions be compared, and then with pragmatic guidelines to improve the full analysis rather than merely 'fit' expectations. Any hazard intuited should be checked for correctness, and if correct and not yet exposed by the analysis, should be fitted into the analysis in a way that improves the entire framework of the analysis as follows, rather than creating an un-scalable one-off fix:

The intuited hazard must ultimately find a fit under a UCA. If no UCA is yet documented for which the hazard is a natural fit, then consider if there exists a documented control action and UCA pattern under which such a UCA might be logically added. Failing this, then consider control actions of the existing controllers to see if one

may be missing. If none is missing, then perhaps a controller is missing from the analysis, and the control structure should be revised.

When working with complex systems where SMEs can be drawn upon during STPA, the method of [83] may uncover additional non-obvious loss scenarios.

It is interesting to note the time-consuming nature of this analysis reflects comments from the US Air Force 412th Test Wing (see Section 2.3). The cause of this is as yet unknown. Perhaps it was due to lack of access to SMEs or lack of practice with STPA (and no facilitator), as both this analysis and that of the 412th suffered these same limitations. It could also be due to the complexity or quantity of hazards revealed by STPA, or by an inherent characteristic of the method itself.

One way to minimize time investment would be to decrease scope. This, however, may inherently exclude certain hazards from consideration. Scope should be set carefully as not to ignore hazards of relevance while still reducing total workload for the analysis.

5. Adopting STPA: Promises and Challenges

> "If I were leading a squadron to the first flight of a new X Plane today, I would be using STPA to do our scenario planning and hazard identification."
>
> – *Douglas 'Beaker' Wickert, Commander, 412th Test Wing*[27] [100]

There is power in STPA to handle new and complex systems under test. But this methodology and its theoretical framework are largely new to flight test organizations, and there are always barriers of discomfort, disruption, and unease to adopting something new. Fortunately, adopting STPA need not be an 'all or nothing' affair, and while sweeping benefits may be derived from deploying STPA broadly across design, validation, and verification processes, it is possible to reap notable benefits even at a local level by adopting STPA within individual organizations, and even for only selective aspects of analysis. Section 5.1 will discuss adoption of STPA limited to the flight test organization level.

In the context of flight test hazard management, STPA offers distinct power in hazard identification, cause and effect analysis, and hazard mitigation analysis. Adopting STPA in flight test does not necessitate changing the way risks are evaluated or managed, though it can provide stronger tools and deeper insights for doing so. Likewise, it does not demand changes in safety management processes, but rather provides richer and more detailed information to support them. And while STPA can be a powerful tool for hazard identification, it does not even necessitate abandoning the brainstorming associated with THA; the two can be merged in useful ways while providing peace of mind during transition to the new methodology.

Finally, Section 5.2 envisions what full end-to-end adoption of STPA in the aircraft development process might look like, including additional benefits and challenges doing so may pose.

5.1. STPA at the Local Level: Flight Test Organizations & Processes

While adoption of STPA in design awaits certain milestones of certification process acceptance and organizational transition, STPA in flight test may be possible immediately. Furthermore, as novel and complex systems come under test, flight test organizations already experience the need for more capable methodologies. They can already begin adopting STPA. Once test organizations adopt STPA it is reasonable to expect their consultations with SMEs in engineering organizations will help demonstrate its capability there as well.

At the level of an individual flight test organization, STPA can readily be used for flight test hazard identification. Supplemented with analysis of mitigating and recovery measures as well as a risk assessment by any existing method, STPA may be used in lieu of THA without changing present safety management processes. As trust and experience

[27] US Air Force, Edwards Air Force Base. He was not yet commander at the time of this writing.

is gained in STPA, changes can be made to safety management processes to better utilize the results delivered by STPA and to reduce or eliminate the supplementary work needed to match present process implementations.

While it would be beneficial for a flight test organization to receive existing safety analyses from the engineering organizations in the form of STPA, this is not necessary for use. As discussed briefly in Section 3.2, STPA can integrate the results of analysis by the legacy methods that may be delivered by the engineering organizations.

The transition from THA to STPA may be eased by applying the two methods in parallel while trust and experience is gained. This is resource-intensive, however, and therefore unappealing. A better solution uses THA to 'sanity check' STPA while gaining experience and trust with the methodology. If undertaken pragmatically, THA and its full complement of aids to brainstorming may be used via an approach as described in Section 4.4 to help to assure that each branch of the STPA analysis is properly constructed (UCAs) and fleshed out (loss scenarios).

STPA being a scalable methodology, it can find usefulness beyond test hazard identification even within the flight test organization. As systems become increasingly complex and coupled, flight test organizations may have little choice but to adopt methods like STPA to demonstrate the safety and performance of the systems they test:

> *Toward the collection of test data (the evidence to support an assurance or system performance argument), in a complex system the number of potential outcomes is unbounded. Consequently, a brute force approach to gathering performance data on a complex system (testing every possible outcome) becomes unachievable on account of resources and time. A different approach is required to gather data to prove performance. Rather than testing every potential outcome, it becomes necessary to test only the points of control – the limits of system authority. For this,… Systems Dynamics and STPA analyses are critical to the placement of test points.[49]*

On the other hand, some organizations have found STPA to be time-intensive to the point of limiting where they feel they can apply the methodology in their test programs, for example,

> *Boeing Technical Fellows have proposed using an adapted version of STPA for our Test Hazards and Analysis. It's difficult to visualize how the existing STPA process could be applied to the test plan level. The full method is a bit cumbersome for the rapid rate at which flight test engineers sometimes operate. However it would be interesting to see if STPA can be applied at the maneuver level for a specific airframe. The future will see a continual increase in systems and software-heavy aircraft and the number of controller interactions will increase proportionally. By setting the appropriate boundary conditions on the analysis, conducting maneuver evaluations ahead of a flight test program with STPA may become feasible.[73]*

It appears likely that some of the overhead they observe is due to the lack of STPA work products delivered to flight test from engineering, necessitating that flight test perform the full analysis from the ground up.

> *One of the reasons STPA takes a while right now is because it's not used in the design phase. Therefore testers have to take a lot of time to develop the model and then*

conduct the analysis. If design engineers incorporated STPA into their processes and it was handed to testers all testers would need to do is add the test specific aspects to the model and run the resulting delta analysis. While STPA takes longer on the test side compared to traditional methods, it is a lot faster on the design side and because it focuses on system behavior, it can be readily adapted to a flight test program. Testers seem to get the value of STPA, but we haven't had as much success pushing the rope on the design side. I think if we could get designers on board this would be a game changer.[82]

Another organization similarly highlighted the lesson that STPA should be applied "in early project development in cooperation with the program office and designer/contractor."[69]

5.2. 'Concept-to-Boneyard STPA': Possibilities & Pipe Dreams

By adopting STPA early in the aircraft development process, it is possible to envision a reality in which the flight test organization receives a very well analyzed package to start their analysis from, as just described. In fact, from the very beginning of conception of an aerospace product, each next organization involved in its development could receive from the previous organization a foundation of analysis upon which to build, and pass their further developed analysis onward to the next organization: an 'end-to-end STPA analysis relay'. For example, by beginning STPA at the ConOps development phase, and continuing to adapt the same analysis through configuration, systems architecture, design, engineering, etc., and then flight test, a situation is created in which the spool-up effort required by each organization is significantly reduced.

Work by Montes helps "standardize the applicability of hazard findings between the design and field use of the product."[17] As STAMP-based tools gain broader use on the design side, this becomes is an extremely valuable approach.

Furthermore, by continuing this STPA relay beyond certification flight test and into pilot training, line operations, maintenance safety, and return-to-service flying, it is possible the safety awareness gained at so many levels may be passed along through the life of the product.

The scalability and capacity for abstraction that STPA offers suggest that such an approach is quite feasible. Whether analysis begins with an abstract control structure and successive organizations contribute further detail, or whether organizations build a control structure out laterally, each organization may utilize the work of the previous organizations as a foundation. Adoption of a safety-guided design approach (such as described by Horney [101]) early in the development process may help to structure such an ongoing analysis productively.

This is no small undertaking, however, and significant hurdles would need to be solved for such a process to become streamlined. To the knowledge of the author, no documentation tool yet exists to easily capture, modify, analyze, and work with STPA work products, especially across teams. This is an especially challenging undertaking when each organization may desire to view only particular subsets of elements, UCAs, or scenarios that are relevant to them.

Currently, to the knowledge of the author, there is not yet a straightforward way to re-analyze a system using STPA without re-analyzing its entirety. While Section 3.4

discusses one work in this direction, more development is needed on this front. The ability to methodically cordon off areas for re-analysis following modifications will be essential to the end-to-end STPA undertaking, especially given the collaborative and ongoing work with large STPA analyses this will entail.

While increasing numbers of people are learning and becoming familiar with STPA, there is still a dearth of those who are able to lead, facilitate, and teach. Deliberate efforts to educate, practice, and master STPA will be necessary within organizations wishing to use STPA more deeply.

6. Conclusions

Exploring the theoretical compatibility and practical and systemic hurdles to using STPA for flight test hazard identification, this thesis aimed to advance the state of flight test safety by exploring tools for improved test hazard identification.

Flight test safety and process have largely evaded academic literature until now. Sections 2.1, 2.2, 2.3, and 2.4 contribute a comprehensive description of the state of the art in Test Hazard Analysis and its context in the broader processes of flight test safety, as scoured from disparate sources. Significant art has evolved around THA to bolster it against the limitations of the brainstorming methodology that underpins it, yet the limitations still leave clear gaps. When novel or complex systems are contemplated the gaps further widen. The structured output of THA forms the foundation of flight test safety management processes across organizations. Despite the limitations of such a format and the risk analyses that underpin much of flight test safety decision-making, sentiment in the field recognizes a need for tools that more comprehensively identify test hazards, but not a desire to alter present safety decision-making processes. Section 2.5 contributes a synthesis of required and desirable characteristics for an improved methodology for flight test hazard identification.

To understand the implications and requirements in adopting STPA in lieu of THA to satisfy the needs expressed in Section 2.5, a comparison of STPA and THA frameworks, constructs, and work products was undertaken. The results of this effort were described in Section 3.1. The two methods employ drastically different theoretical underpinnings and methodologies, yet by supplementing the work products of STPA using legacy methods it is possible to maintain the output needed for present organizational processes while obtaining the improved hazard coverage of STPA. It is especially important when transitioning to STPA from a THA background to understand differences in the usage of certain terminology, as the concepts they convey in the STPA context are key to the operation of the method.

Further work explored the use of abstraction to reveal test-related hazards at various scales (see Section 3.2), and examined considerations in identifying, testing, and violating safety assumptions in the process of flight testing (see Section 3.3). Abstraction and system scoping permit analysis at many levels that are of interest in flight test, including test equipment safety, test maneuver safety, real-time monitoring and decision-making safety, and organizational decision-making safety. A conscious awareness of assumptions that may be tested or violated can permit informed examination and mitigation.

The basic use of STPA for test hazard identification was illustrated by a walk-through of the process on a real flight test campaign, as described in Section 4. Constructing a control structure model of the system can already be a boost to test team productivity. Identifying the control actions available to each controller further clarified understanding. With these, the table of potential UCA patterns breaks the analysis into approachable segments while maintaining the full holistic nature of the analysis itself (i.e., without decomposing the system).

It is important to perform the steps of STPA sequentially in order to properly construct the bottom-up analytical framework. Definition of losses, top-level hazards, and control structure model of the system should be consensus-driven with the entire

team. Generation and documentation of UCAs and loss scenarios permit asynchronous contributions. UCAs should be framed in terms of system characteristics and behaviors. Identification of loss scenarios requires subject matter expertise, and a structured process can help elicit completeness.

Intuition can be used to spot-check and improve the framework. The intuited hazard must ultimately find a fit under an appropriate UCA, guiding a disciplined approach that properly builds out the analysis itself rather than merely fitting it to existing knowledge.

Adoption of STPA in flight test, and more broadly in aircraft development, will be a progressive process. As described in Section 5.1, flight test organizations can adopt STPA without changing existing organizational safety decision-making processes, by supplementing the products of STPA analysis via legacy methods. Later, organizational safety decision-making processes may be updated to better utilize STPA and increase efficiency and effectiveness.

Section 5.2 sketches a vision for end-to-end STPA throughout the lifecycle of aerospace product development, fielding, and operation. Aside from the hurdles to adoption of STPA in engineering organizations, additional hurdles of implementation will also need to be addressed. These include streamlined working tools, a methodology for efficient re-analysis following changes, and training of personnel.

By understanding key differences between STPA and THA, and the flight test context in which THA is presently used, it becomes possible to integrate STPA into existing flight test processes where it can provide immediate utility. As experience is gained and hurdles are overcome, increased adoption across the aerospace lifecycle and processes revisions to better harness the strengths of STPA can create new efficiencies in the safety workstream.

The cause of high workload experienced in conducting STPA on flight test warrants further investigation. Whether due to lack of access to SMEs, lack of practice/access to STPA, complexity or quantity of hazards revealed, or an inherent characteristic of the method itself, further study might guide more nuanced choices of implementation. If the workload is indeed due to the volume or quality of hazards revealed, then perhaps a study of the time value of hazard identification coverage could be warranted. Scope may also be adjusted to balance prospective hazard coverage versus time investment.

Further work remains to determine whether flight test organizations can use STPA to test new and emerging aircraft more safely without substantially increased workload.

7. Acknowledgements

I owe tremendous gratitude to many who have supported me in these efforts and others. To Prof. Nancy Leveson, my advisor, for her tireless quest for safe, thoughtful engineering that accounts for the human element, and for her quiet acts of support through the recent months. To my labmates in the Engineering Systems Lab, and my immediate colleagues in the System Safety Group; and special thanks to Rodrigo Rose and Justin Poh for their generosity in giving advice on this thesis. To Sandro Salgueiro also for guidance, advice, and support on this thesis and elsewhere. To Prof. Katya Arquilla for her mentorship. To Prof. Kerri Cahoy for seeing and supporting me. To Addison Tower, Dave Sizoo, Raven LeClair, Todd Lardy, Greg Lewis, George Cusimano, and others who have provided motivation, insight, and encouragement along the way. To Ayaka Miyamoto for enjoyable collaboration scrutinizing the hazards of Auto-GCAS. To Natasha Neogi for her guidance, clarity, and perspective. To my friends, who have been there for me along the way. And to my family– my brothers, my mom and dad, my grandparents, aunts, uncles, and cousins– for their support, guidance, and endless love through it all.

A special heartfelt thank you also to those who became my foundation of support at MIT in recent months; I could not have done this without your camaraderie, friendship, and care. It means the world to me. I will name only Profs. Or Hen and Tal Cohen– your kindness and generosity changed my world.

[this page unintentionally left blank[28]]

[28] Yes, this passes for humor in some circles.

8. Appendix

Partial listing of UCAs, Auto-GCAS Logic

UCA	Not Provide	Provide	Too soon / too late / out of order	Stopped too soon / applied too long
'Push' arm button (“AA”)	(1) AGCAS Logic does not push the Autopilot arm button when necessary to avoid terrain. [H1]	(1) AGCAS Logic arms the Autoflight when Airplane is not in Evasive Condition. [H1, H2, H3] (2) AGCAS Logic arms the Autoflight when Autopilot modes or setpoints cause flight toward terrain. [H1] (3) AGCAS Logic arms the Autoflight when Autopilot modes or setpoints cause flight toward another aircraft. [H2] (4) AGCAS Logic pushes the Autopilot arm button when Airplane enters Evasive Condition and Autopilot is already armed. [H1]	(1) AGCAS Logic arms the Autoflight before it sets the modes and setpoints. [H1, H2, H3] (2) AGCAS Logic arms the Autoflight too late when Airplane is in Evasive Condition. [H1]	[N/A: discrete action]
'Push' "FLC" (airspeed climb) mode button ("FLC")	(1) AGCAS Logic does not 'push' FLC button when necessary to avoid terrain. [H1, H2, H3]	(1) AGCAS Logic 'pushes' FLC button when FLC mode already active. [H1, H2, H3] (2) AGCAS Logic repeatedly 'pushes' FLC button. [H1, H2, H3]		[N/A: discrete action]
'Push' "HDG" (heading) mode button ("HDG")	(1) AGCAS Logic does not 'push' HDG button when necessary to avoid terrain. [H1, H2, H3]	(1) AGCAS Logic 'pushes' HDG button when initiating fly-up and HDG mode already active. [H1, H2, H3] (2) AGCAS Logic repeatedly 'pushes' HDG button. [H1, H2, H3]		[N/A: discrete action]
Move airspeed bug nose up ("PNU")	(1) AGCAS Logic does not move airspeed bug to Vy when necessary to avoid terrain. [H1]	(1) AGCAS Logic moves altitude bug to Vy when not in Evasive Condition. [H2] (2) AGCAS Logic moves altitude bug to Vy when Autopilot is armed and traffic is ahead and above. [H2]	(1) AGCAS Logic excessively delays moving airspeed bug to Vy after arming auto-throttle. [H3]	
Sync heading bug ("SHB")	(1) AGCAS Logic does not sync heading bug when necessary to avoid terrain. [H1]	(1) AGCAS Logic syncs heading bug when not entering Evasive Condition. [H1, H2] (2) AGCAS Logic syncs heading bug when Autopilot is armed and terrain or traffic is straight ahead. [H1, H2]	(1) AGCAS Logic does not sync heading bug before moving heading bug when entering Evasive Condition. [H1]	

Move heading bug ("MHB")	(1) AGCAS Logic does not move heading bug to desired heading when necessary to avoid terrain. [H1]	(1) AGCAS Logic moves heading bug when not entering Evasive Condition. [H1, H2] (2) AGCAS Logic moves heading bug when Autopilot is armed and terrain or traffic is nearby. [H1, H2]	(1) AGCAS Logic does not move heading bug after syncing heading bug when entering Evasive Condition. [H1]	
Move altitude bug ("MAB")	(1) AGCAS Logic does not move altitude bug to sufficiently high altitude when necessary to avoid terrain. [H1]	(1) AGCAS Logic moves altitude bug when not entering Evasive Condition. [H1, H2, H3] (2) AGCAS Logic moves altitude bug downward when Autopilot is armed and terrain is nearby. [H1, H2] (3) AGCAS Logic moves altitude bug when Autopilot is armed and traffic is nearby. [H1, H2]	(1) AGCAS Logic does not move heading bug after syncing heading bug when entering Evasive Condition. [H1]	
Arm auto-throttle ("AAT")	(1) AGCAS Logic does not arm auto-throttle when necessary to avoid terrain. [H1, H3]	(1) AGCAS Logic arms auto-throttle when not in Evasive Condition. [H2, H3]	(1) AGCAS Logic excessively delays arming auto-throttle after entering Evasive Condition. [H1, H3]	
Request local terrain from TDB ("TDB")	(1) AGCAS Logic does not request local terrain when Aircraft location changes. [H1]	(1) AGCAS Logic requests local terrain more frequently than TDB can fulfil. [H1] (2) AGCAS Logic requests local terrain for incorrect location. [H1]	(1) AGCAS Logic delays requesting local terrain when Aircraft location changes. [H1]	[N/A: discrete action]

Example listings of loss scenarios, Auto-GCAS Logic

UCA: AGCAS Logic does not push the Autopilot arm button when necessary to avoid terrain. [H1]

- AGCAS Logic does not detect the Evasive Condition[29] due to faulty AHRS data, and therefore does not attempt to find or execute an escape trajectory, thus not arming the Autopilot.
- AGCAS Logic does not detect the Evasive Condition because the dynamics model it uses does not match the actual performance of the Airplane for the condition in question, and therefore does not attempt to find or execute an escape trajectory, thus not arming the Autopilot.
- AGCAS Logic is unable to identify a successful escape trajectory, so it does not set the Autopilot modes or setpoints, and therefore also does not arm the Autopilot.
- At the moment when the AGCAS Logic tries to arm the Autopilot, the Crew is activating the disarm control in order to manually fly the Airplane out of Warning Condition, so the arm signal sent by the AGCAS Logic does not arm the Autopilot.
- AGCAS Logic does not detect the Evasive Condition because of faulty terrain data, and therefore does not attempt to find or execute an escape trajectory, thus not arming the Autopilot.
- AGCAS Logic does not detect the Evasive Condition because it is unable to compute the collision detection quickly enough, and therefore does not attempt to find or execute an escape trajectory, thus not arming the Autopilot.
- AGCAS Logic does not detect the Evasive Condition because dynamic parameters such as weight or CG in the dynamics model it uses do not match the actual parameters of the operation in question, and therefore does not attempt to find or execute an escape trajectory, thus not arming the Autopilot.

UCA: AGCAS Logic arms the Autoflight when Airplane is not in Evasive Condition. [H1, H2, H3]

- AGCAS Logic identifies a successful escape trajectory for an Evasive Condition that no longer exists, but the logic to identify and execute an escape trajectory has not terminated, so it arms the Autopilot.
- On landing, the AGCAS Logic interprets the approach to terrain (runway) as Evasive Condition, activating an escape trajectory.
- On visual approach in mountains, the AGCAS Logic interprets the approach to terrain along the approach path as Evasive Condition, activating an escape trajectory.
- At the moment when the AGCAS Logic tries to arm the Autopilot, the Crew was activating the disarm control in order to manually fly the Airplane out of Evasive Condition; when the Crew stops activating the disarm control after exiting Evasive Condition, the AGCAS Logic continues trying to arm the autopilot as it failed to do so initially.
- AGCAS Logic determines the Airplane is in Evasive Condition when it is not, due to faulty terrain data, therefore arming the Autopilot.
- AGCAS Logic determines the Airplane is in Evasive Condition when it is not, due to faulty AHRS data, therefore arming the Autopilot.
- AGCAS Logic determines the Airplane is in Evasive Condition when it no longer is, because slow computation of collision detection produces results that are out of date, and therefore arming the Autopilot.
- AGCAS Logic determines the Airplane is in Evasive Condition when it is not, because the dynamics model it uses does not match the actual performance of the Airplane for the condition in question, and therefore it arms the Autopilot.
- AGCAS Logic determines the Airplane is in Evasive Condition when it is not, because dynamic parameters such as weight or CG in the dynamics model it uses do not match the actual parameters of the operation in question, and therefore it arms the Autopilot.
- On takeoff, the AGCAS Logic comes online and computes motion derivatives differencing initial sensor readings and initialization values, such that the resulting derivatives model motion toward terrain that does not actually exist, thus identifying Evasive Condition when none exists and activating an escape trajectory.

[29] "Evasive Condition" denotes when the airplane is in a situation where immediate evasive action is deemed necessary to avoid impact with terrain.

9. Bibliography

[1] D. G. McCullough, *The Wright Brothers*. New York, NY: Simon & Schuster, 2015.

[2] "Crash During Experimental Test Flight, Gulfstream Aerospace Corporation GVI (G650), N652GD," National Transportation Safety Board (NTSB), Accident Report AAR-12/02 PB2012-910402, Oct. 2012. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1202.pdf

[3] Fred Stoliker, ed., "AGARD Flight Test Techniques Series. Volume 14. Introduction to Flight Test Engineering (Introduction a la technique d'essais en vol)." Jul. 2005. [Online]. Available: https://apps.dtic.mil/sti/citations/ADA303918

[4] K. M. Pavlock, "Aerospace Engineering Handbook Chapter 2(v): Flight Test Engineering." National Aeronautics and Space Administration: Dryden Flight Research Center. [Online]. Available: https://ntrs.nasa.gov/api/citations/20140010192/downloads/20140010192.pdf

[5] Federal Aviation Administration, "Order 4040.26C - Aircraft Certification Service Flight Test Risk Management." [Online]. Available: https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/documentID/1040360

[6] D. Ward, T. W. Strganac, and R. Niewoehner, *Introduction to Flight Test Engineering*, 3rd ed. Dubuque, Iowa: Kendall/Hunt Pub., 2006.

[7] Dennis Morley *et al.*, "Flight Test Safety and Risk Management." Science and Technology Organization, North Atlantic Treaty Organization, Jan. 2021. [Online]. Available: https://apps.dtic.mil/sti/trecms/pdf/AD1183573.pdf

[8] Col. D. "Beaker" Wickert, "Risk Awareness: A New Framework for Risk Management in Flight Test," in *Proceedings of the SETP 62nd Annual Symposium*, 2018.

[9] J. J. Bertin, *Hypersonic aerothermodynamics*. in AIAA education series. Washington, DC: American Institute of Aeronautics and Astronautics, 1994.

[10] J. J. Bertin and R. M. Cummings, "Fifty years of hypersonics: where we've been, where we're going," *Prog. Aerosp. Sci.*, vol. 39, no. 6, pp. 511–536, Aug. 2003, doi: 10.1016/S0376-0421(03)00079-4.

[11] V. S. Parsons, "Searching for 'Unknown Unknowns,'" *Eng. Manag. J.*, vol. 19, no. 1, pp. 43–46, Mar. 2007, doi: 10.1080/10429247.2007.11431721.

[12] *Exploring the Visual THA*. [Online Video]. Available: http://flighttestsafety.org/2023-wichita-ks/400-9-exploring-the-visual-tha

[13] Donald H. Rumsfeld, "Defense.gov Transcript: DoD News Briefing - Secretary Rumsfeld and Gen. Myers," Feb. 12, 2002. [Online]. Available: https://web.archive.org/web/20160406235718/http://archive.defense.gov/Transcripts/Transcript.aspx?TranscriptID=2636

[14] Mark Jones Jr., "Flight Test Safety Fact, 22-10." [Online]. Available: https://flighttestfact.com/flight-test-safety-fact-22-10/

[15] Gregory Lewis and George Cusimano, "Flight Test Principles and Practices (short course)," San Diego, CA, Sep. 18, 2023.

[16] N. Leveson, *An Introduction to System Safety Engineering*. Cambridge, Massachusetts: The MIT Press, 2023.

[17] Daniel R. Montes, "Using STPA to inform developmental product testing," Thesis, Massachusetts Institute of Technology, 2016. [Online]. Available: https://dspace.mit.edu/handle/1721.1/103422

[18] "Flight Test Safety Database." [Online]. Available: https://ftsdb.grc.nasa.gov/

[19] H. H. de Jong, “Guidelines for the identification of hazards: How to make unimaginable hazards imaginable?,” National Aerospace Laboratory NLR, NLR-CR-2004-094. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b3cbf64e441d22c16d6d1edf6a2fe770fba7ec16

[20] Ilias Maragakis *et al.*, “Guidelines on Hazards Identification,” Safety Management System and Safety Culture Working Group, European Civil Aviation Safety Team, European Strategic Safety Initiative, Mar. 2009. [Online]. Available: https://www.flighttestsafety.org/images/ECASTSMSWG-GuidanceonHazardIdentification1.pdf

[21] G. W. Hall, “Risk Management: A Test Pilot’s Perspective,” in *Safety Engineering and Risk Analysis*, Orlando, Florida, USA: American Society of Mechanical Engineers, Nov. 2000, pp. 21–26. doi: 10.1115/IMECE2000-1018.

[22] “private communication,” 2023.

[23] Michael A. Greenfield, “Risk as a Resource: Meeting the Program Management Challenge,” Goddard Space Flight Center, Mar. 30, 2004. [Online]. Available: https://www.nasa.gov/pdf/293221main_62638main_1_pmchallenge_greenfield_risk.pdf

[24] P. W. Merlin, G. A. Bendrick, and D. A. Holland, *Breaking the Mishap Chain: Human Factors Lessons Learned from Aerospace Accidents and Incidents in Research, Flight Test, and Development*. NASA Aeronautics Book Series, 2012. [Online]. Available: http://www.nasa.gov/connect/ebooks/break_mishap_chain_detail.html

[25] Ragnar Guðmundsson, “Runway excursion during flight testing (21 July 2013),” Icelandic Transportation Safety Board, Keflavik Airport (BIKF), Report on aircraft accident M-01313/AIG–09, Mar. 2016. [Online]. Available: https://reports.aviation-safety.net/2013/20130721-0_SU95_97005.pdf

[26] “AC-130J, T/N 09-5710,” US Air Force Aircraft Accident Investigation Board, Eglin Air Force Base, Florida, Aircraft Accident Investigation AC-130J, T/N 09-5710, Apr. 2015. [Online]. Available: https://reports.aviation-safety.net/2015/20150421-0_C130_09-5710.pdf

[27] “In-Flight Breakup During Test Flight Scaled Composites SpaceShipTwo, N339SS,” National Transportation Safety Board (NTSB), Accident Report AAR-15/02 PB2015-105454, Oct. 2014. [Online]. Available: https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1502.pdf

[28] Bill Jaconetti, “Flight Test Safety Risk Management and Test Hazard Analysis (THA),” *Flight Test News*, May 04, 2015. [Online]. Available: https://flighttestfact.com/flight-test-safety-risk-management-and-test-hazard-analysis-tha/

[29] Mark Jones Jr., Ed., “Flight Test Safety Fact Issue 20-06.” Test Safety Committee, Jun. 2020. [Online]. Available: https://flighttestfact.com/wp-content/uploads/Flight-Test-Safety-Fact-20-06.pdf

[30] *System Theoretic Process Analysis STPA for Flight Testers 1*, (Mar. 10, 2021). [Online Video]. Available: https://www.youtube.com/watch?v=vsgFQgLekSk

[31] A. F. Pashchenko and V. I. Akhrameev, “On Some Approaches to Development of Flight Safety Control System for Experimental and Civil Aviation,” *IFAC-Pap.*, vol. 54, no. 13, pp. 704–709, 2021, doi: 10.1016/j.ifacol.2021.10.534.

[32] S. Wu, “Evaluation and analysis of civil aircraft flight test safety management capability based on matter element extension theory,” SPIE, 2023, pp. 1277910-1277910–7. doi: 10.1117/12.2688666.
[33] “Study on Risk Assessment of Flight Test of Transport Aircraft Based on IAHP-SPA Model | IEEE Conference Publication | IEEE Xplore.” [Online]. Available: https://ieeexplore.ieee.org/document/9613072
[34] J. M. Stellman and International Labour Organisation, Eds., *Encyclopaedia of occupational health and safety*, 4th ed. Geneva: International Labour Office, 1998.
[35] J. Reason, “The contribution of latent human failures to the breakdown of complex systems,” *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, vol. 327, no. 1241, pp. 475–484, Apr. 1990, doi: 10.1098/rstb.1990.0090.
[36] J. Reason, E. Hollnagel, and J. Paries, “Revisiting the ‘Swiss Cheese’ Model of Accidents,” EEC Note No. 13/06. doi: 10.1163/1570-6664_iyb_SIM_org_39214.
[37] Ronald B. Levy, “Teaching Human Factors and Safety Culture,” presented at the FAASafety Seminar: Instructional Best Practices for Teaching Human Factors & Safety Culture, Zoom, Aug. 15, 2023.
[38] *Aviation Instructor’s Handbook (FAA-H-8083-9B)*. Federal Aviation Administration (FAA), 2020. [Online]. Available: https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/aviation_instructors_handbook
[39] “Risk Management Handbook (FAA-H-8083-2A) | Federal Aviation Administration.” [Online]. Available: https://www.faa.gov/regulationspolicies/handbooksmanuals/risk-management-handbook-faa-h-8083-2a
[40] N. G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press, 2012. doi: 10.7551/mitpress/8179.001.0001.
[41] W. F. Gerhardt and L. V. Kerber, *A Manual of Flight-Test Procedure*. Department of Engineering Research, University of Michigan, Ann Arbor, 1927.
[42] Thomas Imrich, “Safety via Regulation,” presented at the SETP PNW Section Meeting, Seattle, Washington, United States, Apr. 20, 2019. [Online]. Available: https://secure.whoglue.net/papers/SETP_Safety%20via%20Regulation%20-%2020%20April%20V9.0_Imrich%20slides.pdf
[43] Mark J. Mondt, II, *The Tao of Flight Test: Principles to Live by*. J. I. Lord, 2014.
[44] International Civil Aviation Organization (ICAO), “Safety Management Manual (SMM) (Doc 9859).” 2013. [Online]. Available: https://www.icao.int/SAM/Documents/2017-SSP-GUY/Doc%209859%20SMM%20Third%20edition%20en.pdf
[45] William J. Jaconetti, “Applying Traditional Flight Test Safety Risk Management Techniques to New / Novel Aircraft,” presented at the 52nd Annual SFTE Symposium, St. Louis, MO, 2021. [Online]. Available: https://www.sfte.org/2021_-_st_louis.php#Applying%20Traditional%20Flight%20Test%20Safety%20Risk%20Management%20Techniques%20to%20New%20/%20Novel%20Aircraft
[46] Federal Aviation Administration (FAA), “AC 23.1309-1E - System Safety Analysis and Assessment for Part 23 Airplanes.” Nov. 17, 2011. [Online]. Available: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/1019681

[47] "Job Posting: System Safety Engineer - Textron." [Online]. Available: https://careers.textron.com/wichita-ks/system-safety-engineer/D3C08C84D98C42D397EFA08494EF3167/job/?vs=6400&utm_source=RR+Saved+Search+Emails-DE&utm_medium=Other&utm_campaign=RR+Saved+Search+Emails
[48] NAVAIR, "Test Planning Manual," NAVAIR Integrated Systems Evaluation, Experimentation, and Test Department, NAVAIR M-3960.4, Dec. 2018.
[49] Ben Luther, Indra Gunawan, and Nam Nguyen, "Flight Testing Tomorrow's Complex Systems," presented at the 53rd SFTE International Symposium, London, ON, Canada, 2022. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=184354265&fac=8978935&org_id=SFTE
[50] Dominique Fournier and Antoine Van Gent, "Safety Management Systems: How to adapt ICAO based systems to the Flight Test environment?," presented at the 50th SFTE International Symposium, Toulouse, France, Jun. 2019. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=159429131&fac=8978935&org_id=SFTE
[51] T. McAteer, C. Rice, and C. D. R. C. Gavin, "Flight test safety — The U.S. navy approach," in *2018 IEEE Aerospace Conference*, Mar. 2018, pp. 1–7. doi: 10.1109/AERO.2018.8396745.
[52] "Statement from CSB Chairperson Rafael Moure-Eraso on the Passing of Noted Chemical Process Safety Expert Professor Trevor Kletz - General News - News | CSB." [Online]. Available: https://www.csb.gov/statement-from-csb-chairperson-rafael-moure-eraso-on-the-passing-of-noted-chemical-process-safety-expert-professor-trevor-kletz/
[53] Lt Col Aaron A. Tucker, "Safety, Efficacy, and Efficiency: Design of Experiments in Flight Test." Society of Experimental Test Pilots, undated. [Online]. Available: https://www.setp.org/images/stories/SETP_DOE_in_Flight_Test_-_Tucker_Aaron.pdf
[54] Stuart "Chia" Rogerson, "Textron Aviation Test Safety Risk Management." Oct. 2022. [Online]. Available: https://flighttestsafety.org/images/Textron_Aviation_Flight_Test_Safety_Risk_Management_EFTSW_Oct_2022_Mark_Purvis.pdf
[55] European Aviation Safety Agency (EASA), "Flight Test Operations Manual Guide." Apr. 2018. [Online]. Available: https://www.easa.europa.eu/sites/default/files/dfu/FTOM%20Guide.pdf
[56] United States Air Force Test Center, "Air Force Test Center Instruction (AFTCI) 91-202: Test Safety Review Policy." Nov. 23, 2022. [Online]. Available: https://static.e-publishing.af.mil/production/1/aftc/publication/aftci91-202/aftci91-202.pdf
[57] Bradley J. McKeage, "Process Based Management in Action at Cessna Engineering Flight Test," presented at the 34th Annual International Symposium, Society of Flight Test Engineers, Portsmouth, VA, Sep. 2003. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=162528123&fac=8978935&org_id=SFTEtarget=
[58] Civil Aviation Safety Authority, Australian Government, "Advisory Circular AC 21-47 v1.1: Flight test safety." Mar. 2019. [Online]. Available: https://www.casa.gov.au/flight-test-safety

[59] Federal Aviation Administration (FAA), “Order 8040.4C - Safety Risk Management Policy.” [Online]. Available: https://www.faa.gov/regulations_policies/orders_notices/index.cfm/go/document.information/documentID/1042164
[60] Rodrigo J. Huete and Greg Lewis, “Finally! A Flight Test Safety Database Available to the Public,” in *Proceedings of the 51st Annual SETP Symposium*, 2007.
[61] US Department of Defense, “MIL-STD-882E: System Safety.” May 11, 2012. [Online]. Available: https://mail.system-safety.org/Documents/MIL-STD-882E.pdf
[62] Nancy G. Leveson, “Improving the Standard Risk Matrix: Part 1.” 2019. [Online]. Available: https://a3e.com/wp-content/uploads/2021/03/Risk-Matrix.pdf
[63] 412 Test Wing (AFMC), “Air Force Instruction (AFI) 91-202, Material Command, AFRL Supplement: Test Safety Review Policy.” Oct. 11, 2023. [Online]. Available: https://static.e-publishing.af.mil/production/1/afrl/publication/afi91-202_afmcsup_afrlsup/afi91-202_afmcsup_afrlsup.pdf
[64] F. Webster and Terry D. Smith, “Flying Qualities Flight Testing of Digital Flight Control Systems.” [Online]. Available: https://apps.dtic.mil/sti/pdfs/ADA398738.pdf
[65] J. Gregory and T. Liu, *Introduction to Flight Testing*, 1st ed. Wiley, 2021. doi: 10.1002/9781118949818.
[66] Milton O. Thompson, *Flight Research: Problems Encountered and What They Should Teach Us*, vol. NASA SP-2000-4522. NASA History Division, Office of Policy and Plans, NASA Headquarters, 2000.
[67] Marianne Hörberg, “Preparing New Test Engineers for High Risk Flight Testing on the Gripen Aircraft,” presented at the 49th International Symposium, Society of Flight Test Engineers, Savannah, GA, 2018. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=159695527&fac=8978935&org_id=SFTE
[68] Michael “T-Rex” Tibbs and Lowell Bishop, “STPA Applied to the AFTC Safety Processes,” Edwards AFB, CA, Oct. 09, 2018. [Online]. Available: https://flighttestfact.com/wp-content/uploads/STPA-Applied-to-AFTC-Safety-Processes_Bishop_Tibbs.pdf
[69] Mark Jones Jr., Ed., “Flight Test Safety Fact Issue 19-02.” Test Safety Committee, Feb. 2019. [Online]. Available: https://flighttestfact.com/wp-content/uploads/Flight-Test-Safety-Fact-19-02.pdf
[70] Y. Lu, S.-G. Zhang, P. Tang, and L. Gong, “STAMP-based safety control approach for flight testing of a low-cost unmanned subscale blended-wing-body demonstrator,” *Saf. Sci.*, vol. 74, pp. 102–113, Apr. 2015, doi: 10.1016/j.ssci.2014.12.005.
[71] Ryan Bowers and John Thomas, “Safety Implications of Autonomous Vehicles – System Theoretic Process Analysis Applied to a Neural Network-Controlled Aircraft,” presented at the 54th Annual International Symposium, Patuxent River, MD: Society of Flight Test Engineers, Oct. 2023.
[72] Jordan Q. Stringfield, Dulnath D. Wijayratne, and Darren G. McDonald, “Executing the First Automated Test Maneuver on a Boeing Large Commercial Aircraft,” presented at the 53rd SFTE International Symposium, London, ON, Canada, 2022. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=184349377&fac=8978935&org_id=SFTE

[73] Dulnath D. Wijayratne, Jordan Q. Stringfield, Darren G. McDonald, and Shannon S. Clark, "Systems Theoretic Process Analysis as Applied to a Boeing Automated Test Maneuver," presented at the 53rd SFTE Internation Symposium, London, ON, Canada, 2022. [Online]. Available: https://members.sfte.org/members/secure/filearchive/file_secure_check.php?fid=184372595&fac=8978935&org_id=SFTE
[74] Rod Huete, "Op/Ed: The Subjective 2D Risk Matrix used in Flight Test," *Flight Test Safety Fact*, no. 22–02, pp. 1–4, Feb. 2022.
[75] S. Yoo, D. Gregorian, A. Kopeikin, and N. Leveson, "Improving the Risk Matrix," in *New Achievements in Unmanned Systems*, T. H. Karakoc, N. Yilmaz, A. Dalkiran, and A. H. Ercan, Eds., in Sustainable Aviation. Cham: Springer International Publishing, 2023, pp. 83–90. doi: 10.1007/978-3-031-29933-9_10.
[76] Nancy G. Leveson and John P. Thomas, *STPA Handbook*. 2018. [Online]. Available: http://psas.scripts.mit.edu/home/get_file.php?name=STPA_handbook.pdf
[77] Robert B. Kauffman, "Accidents as an Unwanted Energy Transfer," 2020. [Online]. Available: https://robertbkauffman.com/BoatingSafety/wp-content/uploads/2020/11/RM-AccidentUnwantedEnergyTransfer02.pdf
[78] Nancy G. Leveson, "CAST Handbook: How to Learn More from Incidents and Accidents." 2019. [Online]. Available: http://sunnyday.mit.edu/CAST-Handbook.pdf
[79] Mark Jones Jr., "Safety Planning using Applied Systems Theory." [Online]. Available: https://flighttestfact.com/safety-planning-using-applied-systems-theory/
[80] "The Lives of the Stars," *Cosmos: A Personal Voyage*, PBS, Nov. 25, 1980. [Online]. Available: https://www.imdb.com/title/tt0760463/
[81] John P. Thomas, "Guest post: An Introduction to STPA and its Application to Safety-Critical Technologies," Ike Blog. [Online]. Available: https://medium.com/ike-blog/guest-post-an-introduction-to-stpa-and-its-application-to-safety-critical-technologies-c869ad2998f4
[82] Sarah "Poncho" Summers, "Letter to the Editor: STPA Deep Dive." [Online]. Available: https://flighttestfact.com/letter-to-the-editor-stpa-deep-dive/
[83] John P. Thomas, "STPA Step 4: New Scenario Approach," Cambridge, Massachusetts, Apr. 12, 2024.
[84] A. Scarinci, A. Quilici, D. Ribeiro, F. Oliveira, D. Patrick, and N. G. Leveson, "Requirement Generation for Highly Integrated Aircraft Systems Through STPA: An Application," *J. Aerosp. Inf. Syst.*, vol. 16, no. 1, pp. 9–21, 2019, doi: 10.2514/1.I010602.
[85] B. Abrecht, D. Arterburn, D. Horney, J. Schneider, B. Abe, and N. Leveson, "A New Approach to Hazard Analysis for Rotorcraft," *MIT Web Domain*, Feb. 2016. [Online]. Available: https://dspace.mit.edu/handle/1721.1/114753
[86] J. Thomas *et al.*, "An Integrated Approach to Requirements Development and Hazard Analysis," presented at the SAE 2015 World Congress & Exhibition, SAE International, Apr. 2015. doi: 10.4271/2015-01-0274.
[87] Flight Test Safety Committee, "Flight Test Operational Guidance." Feb. 2017. [Online]. Available: http://flighttestsafety.org/images/Flight_Test_Operational_Guidance_v7_FTSC_020717.pdf
[88] Pete Donath, "Trip Report – the vFTSW," *Flight Test Safety Fact*, no. 20–06, pp. 5–6, Jun. 2020.

[89] Federal Aviation Administration (FAA), “AC 25.1309-1A - System Safety and Analysis.” Jun. 21, 1988.
[90] J. Sgueglia, “Managing design changes using safety-guided design for a safety critical automotive system,” Thesis, Massachusetts Institute of Technology, 2015. [Online]. Available: https://dspace.mit.edu/handle/1721.1/106224
[91] “Stanley Kubrick: Playboy Interview (1968) - by Eric Norden,” Scraps from the Loft. [Online]. Available: https://scrapsfromtheloft.com/movies/playboy-interview-stanley-kubrick/
[92] Lockheed Martin, “Auto GCAS: Collision Avoidance System,” Lockheed Martin. [Online]. Available: https://www.lockheedmartin.com/en-us/products/autogcas.html
[93] Federal Aviation Administration (FAA), “AC 61-134 - General Aviation Controlled Flight into Terrain Awareness.” Apr. 01, 2003. [Online]. Available: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentid/22907
[94] Blackbird Aerospace, LLC, “Project Specific Certification Plan (PSCP) Template for Future Certification of Automatic Ground Collision Avoidance System (Auto-GCAS), Spiral 1 Task 8,” Jul. 29, 2023.
[95] R. LeClair, B. Stone, and R. Winiecki, “Behavior Safety Analysis Framework for Enterprise Reliability (B-SAFER) Open System Architecture Technical Report.” Jan. 2024.
[96] Raven J. LeClair, “COZY MKIV [redacted] Experimental R&D Program Letter,” date unknown.
[97] “private correspondence.”
[98] “private correspondence,” May 16, 2024.
[99] J. Thomas, “System-Theoretic Process Analysis (STPA): Primer and Mini-Tutorial,” 2016. [Online]. Available: http://www.flighttestsafety.org/images/STPA_intro.pdf
[100] Douglas “Beaker” Wickert, “Don’t Rule Out STPA,” Mar. 2019. [Online]. Available: http://www.flighttestsafety.org/images/Flight_Test_Safety_Fact_19-03.pdf
[101] D. C. Horney, “Systems-Theoretic Process Analysis and Safety-Guided Design of Military Systems,” Masters Thesis, Massachusetts Institute of Technology, 2017. [Online]. Available: https://apps.dtic.mil/sti/trecms/pdf/AD1109554.pdf