Essays on decision-making in complex engineering systems development

by

Alison Olechowski

B.Sc. Mechanical Engineering, Queen's University (2010)
S.M., Massachusetts Institute of Technology (2012)

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2017

Signature redacted
Author ....
Department of Mechanical Engineering
May 5, 2017

Signature redacted
Certified by ....
Steven D. Eppinger
General Motors LGO Professor of Management
Professor of Management Science and Innovation
Thesis Supervisor

Signature redacted
Certified by ....
Warren Seering
Weber-Shaughness Professor of Mechanical Engineering
Chair of Committee

Signature redacted
Accepted by ....
Rohan Abeyaratne
Graduate Officer, Department of Mechanical Engineering

Essays on decision-making in complex engineering systems development

by

Alison Olechowski

Submitted to the Department of Mechanical Engineering
on May 5, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Mechanical Engineering

Abstract

This thesis presents three essays on the topic of tools for assessment and decision-making in complex engineering systems development.

The first essay presents an extension to the design structure matrix, used to present and analyze the suite of tests a system undergoes at multiple levels of the architecture. This method decomposes the multilevel integration test suite - progressing from component to subsystem to system - and visually represents the test coverage. We demonstrate the new method on a subsea system at BP.

The second essay presents a study of the current state of use of the technology readiness level method. We discovered, described and prioritized 15 challenges associated with assessing and using the technology readiness levels. We further discuss existing and potential solutions to these challenges. This paper is based on input from interviews at seven different organizations, and a survey of over 100 system engineers. System complexity related challenges were found to be particularly critical and currently without adequate solution.

The final essay presents an expansion of our current understanding of the options available at a phase-gate review. Beyond the typical Go and Kill options, we describe the Waiver (with and without review), Delay and switch to a Back-up plan options. We show how it is feasible to extend a simple decision tree model to analyze the expected value of this broader set of options. We demonstrate this method with four case applications from industry.

Thesis Supervisor: Steven D. Eppinger
Title: General Motors LGO Professor of Management
Professor of Management Science and Innovation

Acknowledgments

I will forever cherish the seven years I have spent learning at MIT.

I must begin by thanking my advisor Professor Steven Eppinger. Steve, I have learned so much about being an effective researcher and professional from you; about preparedness, persistence, and striving for excellence. You pushed me to do better and never ceased to deliver useful constructive feedback. You always prioritized opportunities for my growth as a researcher and future-academic, and for that I am so grateful. I am continually impressed by your passion for teaching and mentorship, and will aim to follow your example in my own career.

Thank you committee member and close collaborator Professor Nitin Joglekar. Nitin you never fail to ask interesting questions and provoke thoughtful discussion in our meetings. Not only do you care deeply about the work, but it is obvious that you also care deeply about your students.

Thank you Professor Warren Seering for opening my eyes to how much there is to offer at MIT, a place that you and I both love. You were the first person I met at MIT seven years ago, and I'm so fortunate to have done my master's work with you, and to continue to work together with you as committee chair. You are a source of calmness, reflection and confidence in the sometimes-overwhelming MIT world. Thank you for all that you have done for me.

I met Professor Chris Magee in 2011 in Singapore, where he was part of a team tasked with establishing a new university. It's been wonderful to come full circle by having Chris on my PhD committee. Thank you Chris for your insights and encouragement.

I have had the good fortune to learn from MIT faculty mentors beyond my PhD committee. I absorbed as much as possible from Professor David Wallace's brilliant teaching, his perseverance in doing things the best way, and his connection to home. Professor Maria Yang is a top-notch role model and has taught me about student-focused advising and the benefit of bringing joy to your research team.

In pursuing my PhD I worked with three capable master's students. Terence Teo started our project and set the tone of hardwork and professionalism. Katharina Tomaschek collected valuable data and was always open to problem solving. Tushar Garg tackled a neat branch of work with energy and purpose. Thank you all for the collaboration.

Our connection to our research sponsor BP was invaluable in shaping the direction and the impact of this work. I would like to thank those in Houston who supported me and this work: Tony Diocee, Brandon Webster, Trevor Smith, Russell Mitchell, David Noble, Kristen Loftus, Steve Raymer and Andrew Cockerill.

Further thank you to all of the industry professionals who participated in our interviews, survey, and case studies. This work would not have been possible without you.

In the MechE department I have been constantly supported and inspired by my peers: thank you Shreya Dave, Anders Haggman, Steve Keating, Josh Ramos, Lindy Liggett, Jeff Mekler, Sourobh Ghosh, and the "Thread".

Thank you to Josef and Brianna Oehmen for your hospitality, friendship and mentorship in my early MIT days. Thank you Chevalley Duhart for taking care of the details. Thank you Jane Kokernak for your thoughtful feedback.

It has been such a pleasure to be visited and supported by my friends from home as I have worked through the PhD. Shakiba, Matt, Eric and Dan I could not ask for a better foundation of friendship. To the Olechowski and Revington cousins, JDP, and the Queen's girls, thanks for always bringing the fun!

At MIT I found my place as a Graduate Resident Tutor (GRT) at Simmons Hall. At Simmons I found a community of smart, interesting, interested students and staff. Simmons quickly became home, and the Sims became my family. Living in the undergraduate community has brought me that much closer to the traditions and spirit that makes MIT what it is. I made what I hope will be lifelong friendships with the students, GRTs and Josh and Chansy. I'm going to miss Simmons Hall and the Sims very much.

Finally, I want to thank my O, C and R families. Jacqueline, I hope you have a post-secondary experience as fulfilling as the one I've had at MIT. Jeremy, thanks for the laughs and for holding down the fort at home. Dr. Marianne, thank you for your words of encouragement and wise advice on the PhD journey. I learned a lot from seeing you earn the first PhD in the family! Maddie, even when you're on the other side of the world I know you're on my team. You shine in your dedication to teaching and your big heart.

Granny, look, I did it! I know you're thrilled. Thanks for instilling a priority on education in our family. Grandma O, I think of you with love and thanks for your encouragement on this achievement.

Chris, your sweetness and thoughtful gestures have been buoying these past 7 years. I always appreciate your efforts to remedy my worries and make sure I'm not down.

Lilah, it's funny to think that seven years ago you were a kid and now you're an intelligent, capable, funny, caring adult. Thanks for always being willing to listen, make sense of things, and problem solve with me. Thanks for bringing joy every time you've come to visit Boston. I'm so excited to live closer to you!

Thank you Caroline for forever being a role-model for me, from grade-school student to professional career. I am lucky to know you are always there for me in tough times and good times.

Mum and Dad, I loved living in Boston as much as you said I would. It's been so fun to have a family reconnection to the city.

Dad, thanks for taking me on that first trip to visit MIT in 2009. This place really was as cool as you said it was. I've learned a lot from watching you work on cutting edge, tough problems as I study similar ones. Thanks for always believing I could do big things.

Mum: I think of my path to MIT - of Science Fair and Shad Valley and Queen's and RIM - and how integral you were in getting me along that path. I am so lucky to have you in my corner. I expect there is still plenty to learn from you as I strike out on my own academic career. Thanks for everything.

What a joy it was to have so many people I love and admire in the room (literally and in spirit) at my defense. I truly appreciate my community of support. Thank you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The successful landing of NASA's Curiosity Rover on Mars was achieved with the design and development work of 7000 employees and a $2.5 billion budget [3]; The Boeing 737 is made up of 367,000 parts [4]; The iPhone 6 has 200 components in the camera module alone, designed by a team of 800 engineers and specialists [5]. Today's complex product and system development involves the collaboration of large interdisciplinary teams, collaborating on a huge amount of interdependent activities, all working towards a quality product that is delivered on time, and within budget.

On these complex projects, difficult and coupled decisions abound: from overall strategic direction to detailed component design to supplier selection to test result interpretation. Often these decisions must be made in spite of inadequate information and high uncertainty [6]. Moreover, humans are saddled with a number of cognitive biases that impede our decision-making efficacy [7].

As a strategy to manage this highly complex undertaking, and to combat the previously described biases, engineers and project managers use structured methods and tools for evaluating information and making decisions. This thesis explores three such methods: the design structure matrix (DSM) [8], the technology readiness levels (TRLs) [9], and the phase-gate review process [10]. These tools are useful in a number of ways: they enable a shared understanding of complex information across disciplines and to decision makers; they provide structured and repeatable analysis; and they can facilitate learning across projects over time. These tools are each increasingly adopted in practice, indicating value-add to the user. Yet in these essays we demonstrate that each has room for improvement.

The essays in this thesis are motivated by real industry needs, and the findings are grounded in data and cases from industry. The following section briefly describes the industrial motivations and collaborations for the studies that initiated the three essays presented hereafter.

1.1 Industry Motivations

I joined the research team of Prof. Steven Eppinger, Prof. Nitin Joglekar, and master's student Terence Teo in June, 2013. Starting January 2013, this team was working on a research grant from BP Project 20k, an advanced technology project aiming to develop deepwater resources at 20000 psi, beyond the current 15000 psi pressures. The charge of our work was to improve the management and execution of complex engineering development projects by extending the DSM tool. In particular, BP was interested in extensions related to testing and integration tasks.

The result of this investigation is essay 1: *Improving the systems engineering process with multilevel analysis of interactions.* We present in that work an extension to the DSM method which displays the different levels of integration testing - from component, to subsystem, to system level - overlaid on the system architecture. This multilevel DSM allows the user to observe overall test coverage, as well as to see how early and late interface tests occur. This new method was demonstrated with data from a blowout preventer testing suite at BP. The BP team found the results of this DSM extension to be insightful, and further work to integrate the testing DSM with other operations data proved to be impossible due to access issues.

At this same time, the BP 20k team had performed a TRL assessment for its planned system, anticipating a great deal of new technology as being necessary for such a technically challenging operational environment change. This evaluation considered over 500 components. Yet it was unclear what should be done with this technology assessment information. This sparked the next phase of the research:

an investigation into current state-of-the-art in TRL practice. BP 20k was curious to learn how other complex systems development organizations were using TRLs to inform their technology and product development, and whether this technology maturity information could be integrated with system architecture information for additional value. Master's student Katharina Tomaschek joined our research team for this phase of the work. This investigation resulted in essay 2: *Technology readiness levels: shortcomings and improvement opportunities.* A major contribution of that work is a list of 15 weaknesses of the current TRL method and implementation and their relative criticality in different contexts.

Two promising directions for future work, motivated by critical shortcomings of the TRL method were revealed: 1) an extension of the DSM to integrate architecture and technology readiness information and 2) a model of a staged product-development process, exploring the options available when a TRL requirement is not met at the gate. We brought a master's student, Tushar Garg, onboard to handle the first direction, while I tackled the second. We soon realized that the decision dynamics did not only apply to the situation of a low TRL, but to the gate decision given any incomplete or missed requirement. This work resulted in the third and final essay: *Assessment of back-up plan, delay, and waiver options at project gate reviews.* This work includes four application cases, involving demonstration of the model with real examples from three different industries: oil and gas, automotive and medical device.

We believe that our close collaboration with engineers and managers in industry has guided this work in a direction of added value. It is our hope that these three essays provide insight to engineers, project managers, product designers, systems integrators and any person who must make tough decisions on complex projects.

The three essays are followed by a discussion of future work beyond that in each individual chapter. Here we go!

Chapter 2

Improving the systems engineering process with multilevel analysis of interactions

2.1 Introduction

Complex engineering involves multiple types of data and contexts. Figure 2-1 presents a stylized map of phases in a systems engineering V (SE-V) process for managing such a project. The SE-V consists of conceptual development, preliminary system design, detailed design, construction, integration testing and validation, startup, operations, and expansion phases [11]. An investment decision after the preliminary system design phase results in the commitment of capital to execute detailed design, procurement, construction, testing, validation, and startup activities [12]. Hence, decision makers explore ways in which cost, performance, and the impact of downstream integration tasks and risks (the right side of the SE-V) can be examined early, that is, during the decomposition stage (the left side of the SE-V). The complexity in SE-V propagates through a sequence of changes [13]. These changes have been associated with the large number of interactions involved within the system, the need for learning, and the fact that these systems involve decisions at multiple levels. Allied literature used

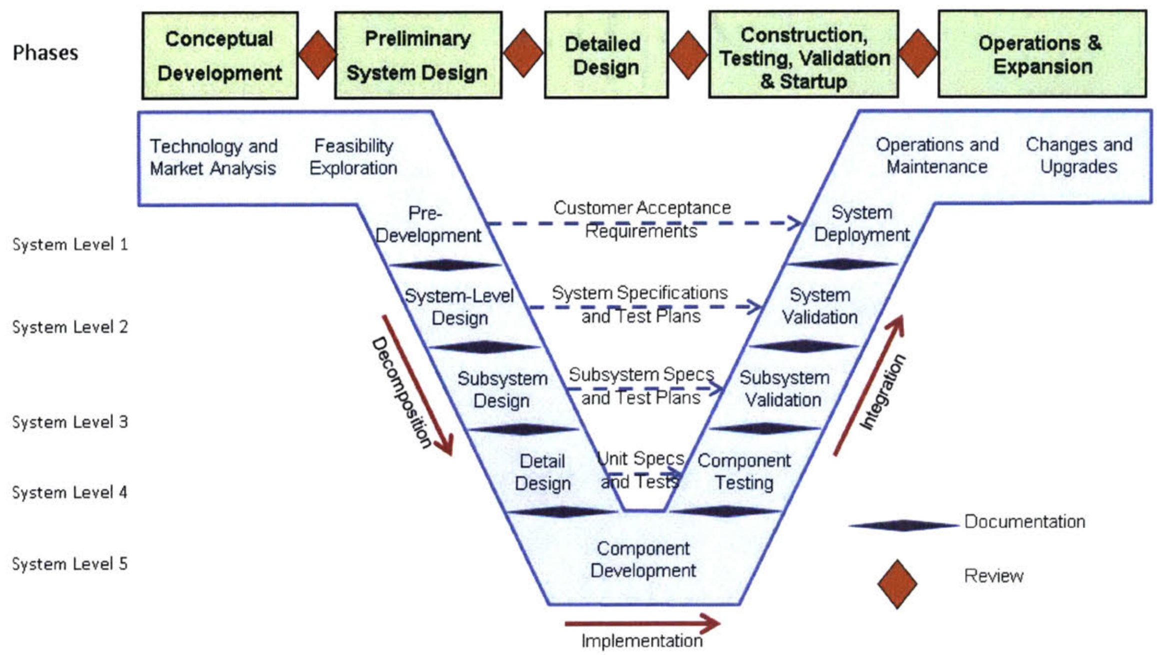

Figure 2-1: Phases and levels within a systems engineering V process.

the term "multiscale" to capture robust design involving multiple level or time phases of decisions [14, 15, 16]. We will use the term "multilevel" to identify decisions that involve two or more decomposition levels (e.g., components vs. subsystems) and/or differing time scales (e.g., time gaps and sequencing needed across preliminary system design, detailed design, and integration tests). Multilevel decisions result in poorly understood interactions. These interactions can lead to negative consequences such as cost overruns, poor startup or operational performance, and even propagation of failures [17, 18]. The design structure matrix (DSM) methodology has made many contributions toward improving complex decisions involving choice of product, process, and organizational architectures during the decomposition of systems on the left-hand side of the SE-V [8]. The SE-V diagram in Figure 2-1 indicates that a complex systems engineering process involves five levels of decomposition (concept development, system-level design, subsystem design, detailed design, and component development), specification, and integration testing. At each level of the system, the twin outcomes of a decomposition task are selection of the architecture for the next level of design and the specification for the corresponding integration tasks (shown by the horizontal dashed arrows in the SE-V). The goal of our research is to build relevant multilevel maps of DSMs involving integration tasks and corresponding component decomposition dependencies and to examine whether analysis of such maps can provide engineering managers with insights to improve the system integration process. This paper presents a method to account for multilevel data in the analysis of dependencies using DSM models. This method contributes to the DSM literature [19] by extending representation schema that incorporate multilevel and multiple-timescale test coverage data as vectors into the off-diagonal DSM cells. These vectors provide a detailed mapping between the product architecture and the SE-V integration test tasks. This mapping is richer than conventional domain mapping matrices (DMMs; see [20]). We report on the collection of a preliminary data set and multilevel analysis of 374 interactions related to a complex offshore oil industry project. Results indicate potential for unanticipated outcomes in terms of incomplete coverage of SE-V integration tasks. We also show that accounting for multilevel features using maximum and minimum function queries, readily identifies all the design interfaces associated with early and late revelations of coverage risks based on a selected suite of integration test tasks. Finally, we discuss theoretical and applied implications of the findings.

2.2 Formulation

2.2.1 Integration and testing for system failures

Failures, sometimes of the most glaring and consequential nature, can and do occur at the boundaries or interfaces between elements. These failures have often been ascribed to uncontrolled, unanticipated, and unwanted interactions between elements: in many cases between elements thought to be entirely separate [21]. For instance, based on an in-depth case study of errors in the Italian air force, Leveson et al. [22] have argued that "emergent safety properties are controlled or enforced by a set of safety constraints related to the behavior of the system components. Accidents result from interactions among system components that violate these constraints - in other words, from a lack of appropriate and effective constraints on component and system behavior." One goal of integration is to identify and resolve potential failures of the system. Among techniques commonly used to this end is failure mode and effect analysis [23, 24]. The aim of this technique is to identify not only all potential failures of the system and its parts but also the effect and the mechanism of the failure. These failures are identified based on the analysis of drawings or flowcharts of the system, an understanding of the function of the system and its component, and details of the environment in which it operates. The process involves generating solutions for how to avoid and/or mitigate the effects of these failures on the system. Alternatively, the hazard and operability study is used to identify failure risks for a given system. The identification is directed by the use of guide words [25]. The process involves the generation of solutions and treatments to address the identified risks. Potential causes of failure can be identified and understood using a fault tree analysis, whereby various failure factors are hierarchically organized and depicted in a tree according to their causal relationship [26]. This method is best performed when the team has a deep understanding of the system and the causes of failure. It is recommended that the team use detailed diagrams of the system as an aid in analysis. Presented as a fundamentally different accident model with an emphasis on systems theory, the systems-theoretic accident modeling and processes model focuses on controller or enforcement failures, not traditional component failures [27]. This method requires the analyst to conceive of the system as a control problem, and it is facilitated through the generation of the process model and control structure for the system. Some methods address failure earlier in the SE-V process, for example, the function-failure design method [28], which can be used during the conceptual design phase.

2.2.2 Hierarchical decomposition and composition

The SE-V process incorporates potential failure modes as constraints on components and subsystem integration based on hierarchical decomposition. The study of constraints on component and system behavior is a nontrivial problem, especially as the complexity of a system rises. Braha and Maimon [29] have modeled the underlying design process as an automaton and proved that the managing of such a planning problem is NP-Hard. Thus, both theorists and practicing engineers look for tools to visualize and understand the dependencies between components and subsystems within a system, especially when the complexity of the system design rises. Related work draws upon managing the decomposition based on hierarchy. For instance, Albers et al. [30] explore a "contact and channel" principle arguing that function and form emerge together during design, and therefore should be considered together in a design representation. This principle is explored in a model of the system architecture of a humanoid robot arm considering the impact of a proposed design change. Tilstra et al. [31] have introduced an extended DSM, illustrated in the context of a screwdriver design, to quantify the degree of nesting during the development of hierarchical product architecture. The DSM is the representation for capturing complex networks of dependencies used in this work [8]. Groups of tasks associated in the SE-V (Fig. 2-1) are mapped into a stylized task DSM in Figure 2-2. Several properties of this task DSM are noteworthy. Owing to the logic of SE-V, there is a regular precedence pattern between task groups as shown by "x" marks immediately below the diagonal, where each DSM mark represents information dependency. The dotted arrows depicting information flowing from the decomposition to the integration tasks in Figure 2-1 result in off-diagonal marks at each level. The "?" marks represent design iterations which may occur after integration tasks. Collectively, these marks form an X-shaped set of dependencies when tasks are grouped at each level of system decomposition. The "z" marks in the component DSM represent the component and subsystem dependencies. Mark "z" is distinct from mark "x" because interactions in the component DSM represent interfaces between the system elements (captured as spatial, energy, etc.). We define DMMs aDMM, dDMM, cDMM, iDMM, and oDMM, corresponding to linkages between the components and each of the task groups: analysis, decomposition, detailed component design, integration, and operations, respectively. The focus of this research is on the dependencies between the component architecture and the integration tasks. Thus, iDMM and corresponding

Domain	Level ID	Group ID	Description	A&F	DT1	DT2	DT3	DT4	CD	IT4	IT3	IT2	IT1	O&M	C1	C2	C3	C4	...	Cn
Analysis Tasks	1	A&F	T& M Analysis & Feasibility											?	aDMM					
Decomposition Tasks	1	DT1	Concept Development	x									?		dDMM					
	2	DT2	System Level Design		x							?								
	3	DT3	Subsystem Design			x					?									
	4	DT4	Detailed Design				x			?										
Component Tasks	5	CD	Component Development					x		?					cDMM					
Integration Tasks	4	IT4	Component Testing					x	x						iDMM					
	3	IT3	Subsystem Validation				x			x										
	2	IT2	System Validation			x					x									
	1	IT1	System Deployment		x							x								
Operations Tasks	1,2,3, 4 & 5	O&M	Operations & Maintenance	x									x		oDMM					
Component Dependecies	1,2,3, 4 & 5	C1	Component 1	aDMM	dDMM				cDMM	iDMM				oDMM		z				z
		C2	Component 2												z					
		C3	Component 3													z		z		
		C4	Component 4													z				
		...	....													z	z			
		Cn	Comonent n																z	

Figure 2-2: A multilevel design structure matrix of systems engineering V tasks and components dependencies.

task and component DSMs are highlighted with chain dotted borders. Extending this representation schema, a single DSM can capture multiple types of interaction data if each off-diagonal cell contains a vector [19]. For instance, these data types might be spatial, information, energy, and material dimensions of component interactions [32]. In contrast, these vectors may capture different types of task interactions [33]. Within this context, two types of gaps are evident in the DSM literature:

1. Conventional DMMs map the elements in one domain to another. For example, component-task DMM maps a component DSM (that captures the complex interaction in product architecture) to a task DSM (that captures the complex interaction among system integration tasks, such as subsystem validation or a subsequent system verification test). However, such DMM mappings [20] have not accounted for the amount of coverage available at each interface within the product architecture based on a selected suite of integration tasks.

2. The importance of accounting for multilevel evolution of complexity has been recognized in the complex engineering literature. For instance, the law of req-

uisite variety [34, 35] postulates that aggregation can absorb variety, where the term variety refers to the total number of possible states of a complex system. A simple example for the application of this law is a patient in a hospital with temperature fluctuation (i.e., uncertainty) associated with fever. Aggregation of some kind is needed if the doctor is not to sit all the month staring at the thermometer. Action must be taken immediately to isolate the patient, such that the root cause of the temperature fluctuation may be explored and understood based on different units of analysis (e.g., either fluctuation in food intake or exposure to environments with different types of germs). In complex engineered systems, analogous decisions may involve situations where subsystem tests during a software development suite fail to reveal a bug, even if a test engineer suspects that a bug exists based on failure history. The test team may have to resort to higher level integration tests, with a sufficient variety of stimuli, to replicate this failure.

Based on the requisite variety law, Bar-Yam [36] argues that "Modularity and abstraction are generalized by various forms of hierarchical and layered specification ... these two approaches either incorrectly portray performance or behavioral relationships between the system parts or assume details can be provided at a later stage." This builds the case for taking a multilevel view of potential integration problems. Multilevel methods, such as logarithmic transformation and filtering of data, enable system design teams to understand patterns of emergent behavior as the complexity of their system rises [37]. For example, data analysis on system-level tests may reveal unique insights about coverage on certain components that may be missing in subsystem-level test data. Conventional DSM models have typically not aggregated, or disaggregated, product architecture and process dependency data based on their levels of decomposition. Our premise is that both of these gaps can be addressed by appropriate data mapping and analysis at each and every interface within the product architecture DSM based on multilevel views of the SE-V process. Hence, we develop a method for data collection, query, and aggregation that accounts for differing levels of testing to examine if different types of integration risks may be evident at different times during the integration process. Integration risk in this instance refers to the potential that any interface covered by a suite of tests during the SE-V integration process may reveal a failure mode within a system design. Data associated with this method grow quickly with increase in the rank of the system DSM, the number of measurement dimensions, and the size of the integration test suite. We have developed a vector representation scheme to capture all interactions from a suite of integration tests that are relevant to a particular DSM cell. Further, in order to isolate the contributions of multilevel analysis, we assign the interactions associated with different interface dimensions (e.g., structural vs. information interactions) at relevant levels (e.g., component vs. subsystem) with unique codes. Thus, the relevant interaction at any level can be queried, analyzed, and displayed as a DSM map. A number of multilevel data aggregation and analysis techniques, ranging from renormalization using finite element analysis to optimal control, have been reported in the literature [38, 39, 40]. Many of these multilevel implementations have been limited to either stylized data or small-scale problems. In our case, we have implemented multilevel analysis in a complex DSM context using maximum and minimum value filters in Section 2.4.

2.3 Research Context and Data

We are working with a research sponsor in the offshore petroleum industry to study a deepwater development project, with focus on the blowout preventer (BOP). The primary function of the BOP is to manage well pressure during drilling by completely sealing off the well bore and circulating out the influx in the event of high-pressure hydrocarbons entering the drill hole. Data collection was performed in three stages. First, we assembled data to create the system architecture DSM. Second, we collected data regarding integration testing. Third, we documented interactions in the system architecture DSM that were tested in each type of integration test. Data were collected over a period of 3 months based on review of engineering documentation and onsite interviews with subject matter experts. These onsite interviews were conducted during 2 weeklong visits. These interviews were followed by e-mail and phone conversations to clarify open issues. Each interface included in the data set was reviewed through this process. Our experience with this data collection process, and allied literature [41], indicates that DSMs are sparsely populated and the size of the data collection scales linearly with the rank of the DSM. In order to manage the data collection effort, some of the subcomponents were grouped into a single component, based on inputs from this review.

2.3.1 System architecture

The BOP system architecture describes its decomposition into subsystems and components.We placed our focus on including those primary components that affect system functions and are critical for system reliability. Ancillary parts (e.g., shuttle valves, piping, and hoses) were grouped with their corresponding components. An initial list of 93 components was created based upon company and industry documentation. These 93 components were classified into eight subsystems. The component list and subsystem boundaries were reviewed with company subject matter experts. The list was refined to 67 components in the following six subsystems:

- lower marine riser package (LMRP)
- blowout preventer (BOP)
- auxiliary lines (Aux Lines)
- choke and kill system (C&K)
- hydraulic power unit
- surface control system

Because the surface control system has minimal interactions in the types of DSMs we will show,we omit this subsystem for clarity, resulting in five subsystems in our analysis here. The next step in data collection was to identify interactions between pairs of components.We were interested in interactions in five dimensions critical to reliability and function, as advised by the subject matter experts. These dimensions are

- *spatial*, involving the physical connection or adjacency of two components;
- *structural*, involving a load or pressure-transferring interaction between two components;
- *energy*, involving the transfer of hydraulic or electrical energy between two components;
- *information*, involving the transfer of information between components by means of electrical signals or hydraulic pilot signals; and
- *materials*, involving the transfer of material (principally drilling mud, but also gas and other wellbore fluids) from one component to another.

All possible pairs of interacting components were identified using engineering documentation. These data were then reviewed with the subject matter experts. The presence of an interaction in any of the five dimensions was recorded. Interaction data are recorded on a binary scale, "0" (*no interaction*) or "1" (*required interaction*). These interaction data for each pair of components form a 67x67 system architecture DSM. We considered the five interaction dimensions separately and created a distinct DSM in each dimension. An entry of "1" indicates the presence of an interaction between the component pair, while a blank indicates a lack of interaction. Figure 2-3 shows the system architecture DSM for the structural dimension, including 56 of the 67 components and their interfaces. (For clarity, we omit the remaining 11 components having no interfaces in the structural dimension. DSM data showing the other four dimensions are also excluded here, for brevity.) The DSM is symmetric, because the interactions are nondirectional. It is possible for interactions to occur within a subsystem or across subsystems. The five areas of possible subsystem interactions, occurring within blocks along the diagonal, have been shaded gray for visual clarity. For example, in Figure 2-3, we see that a within-subsystem interaction exists between components 1 and 2 (i.e., LMRP frame and junction box). Illustration of a cross-subsystem interaction is evident between components 6 and 24 (i.e., the pod hydraulic section within the LMRP and the pod hydraulic section receptacle within the BOP). Some subsystems have more interactions associated with them than others. In general, there are more interactions within a subsystem (in the gray areas along the diagonal) than across subsystems. In the five interaction dimensions combined, 279 interactions are within subsystems and 62 are across subsystems.

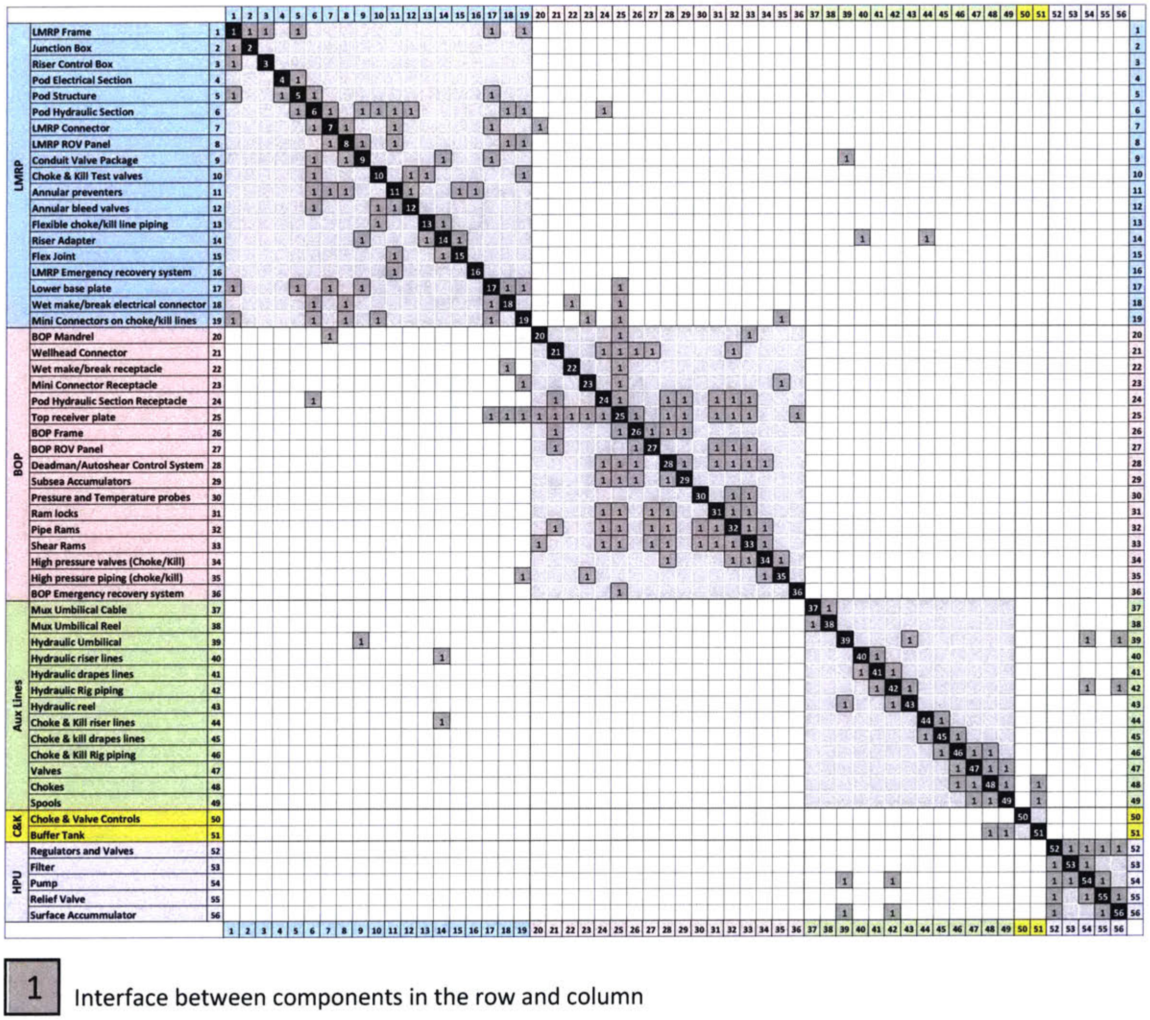

Figure 2-3: System architecture design structure matrix representation of structural interactions between components. Marks in off-diagonal cells identify interfaces between components in their row and column. Five subsystems are highlighted with gray background: lower marine riser package (LMRP), blowout preventer (BOP), auxiliary lines (Aux Lines), choke and kill system (C&K), and hydraulic power unit (HPU).

2.3.2 Integration tests

The second stage of data collection focused on integration test data. Company documentation was first consulted to assemble a list of 57 integration tests. Given that our DSM analysis focuses on integration issues, only tests of interactions between components and subsystems are illustrated. Tests of isolated components are excluded from the analysis in this paper. In addition, while it is possible to conduct tests using digital models, such tests are also excluded when they occur during the decomposition tasks on the downside of SE-V, as opposed to the integration upside of the SE-V. Upon consultation with the company subject matter experts, the list of integration tests was reduced to 25 tests important to system function and reliability, as presented in Table 2.1. It is worth noting that the data shown here are representative and not exhaustive. Thus, it is possible that a test suite in the current analysis might show that an interface is not tested, while it might be tested later through a test that is not discussed here. Each test included in this analysis falls into one of three test levels. Each subsystem is assembled and tested separately in "subsystem-level tests." Next, the full system is assembled and tested in "dock-level tests." Finally, tests of the complete system are conducted in the deployed environment in "subsea-level tests." These tests are sequenced within each level and are temporally separated. In some instances two subsystems are assembled together before the first level of testing, in which case interfaces across these two subsystems are reported to be tested at the subsystem level.

2.3.3 Interactions addressed by integration tests

The third stage of data collection sought to identify which interactions, and which dimensions, were tested in each of the integration tests. Each interaction-test combination was reviewed with the subject matter experts in order to identify these data. Given that there are 25 tests and 374 total interactions across the five dimensions in the full data set, there was a challenging number of combinations to review. To facilitate the subject matter expert consultation process, we developed a data table

Table 2.1: Integration tests.

Index	Test	Level
T1	Mechanical FIT check	Subsystem
T2	Function verification check	Subsystem
T3	Actuator leak check	Subsystem
T4	Shear test	Subsystem
T5	Emergency systems test	Subsystem
T6	EDS test	Subsystem
T7	BOP pressure test	Subsystem
T8	Mechanical FIT check	Subsystem
T9	C&K control and pressure test	Subsystem
T10	HPU function an performance test	Subsystem
T11	Panel function test	Subsystem
T12	System setup and verification	Dock
T13	HPU function test	Dock
T14	Panel/BOP function test	Dock
T15	High pressure mud system test	Dock
T16	Signature test of operators (performance)	Dock
T17	Power and communications redundancy test	Dock
T18	Emergency systems test	Dock
T19	EDS test	Dock
T20	BOP pressure test	Dock
T21	BOP drift test	Dock
T22	Function test	Subsea
T23	BOP pressure test	Subsea
T24	Emergency systems test	Subsea
T25	EDS test	Subsea

where the integration tests could be mapped to the component-component interactions in an efficient manner. An example of this data input is shown in Table 2.2. Each test-component combination was assigned its own row in the table. There is an entry in the row corresponding to an interaction in each of the five dimensions. If no interaction exists in the corresponding dimension, the cell is shaded gray and the combination does not need to be reviewed. If the interaction does exist, the cell is white and the subject matter expert identifies whether that interaction is tested by the test under review. If it is included in the test, an "X" is marked, and if it is not, the cell is marked with an "O." For example, in Table 2.2, test T2, the function verification test, is a test of the valve and actuator functions within the LMRP and

BOP subsystems. For this test, the spatial and structural interactions (shown by X's in Table 2.2) between the LMRP connector and the pod hydraulic section are tested and verified. However, T2 does not test the integrity of the connection between the LMRP and BOP; the BOP mandrel's spatial, structural, and material interactions with the LMRP connector are not tested in T2 (as shown by O's in Table 2.2).

Table 2.2: Test coverage data.

Index	Test	Component 1	Component 2	Material	Information	Energy	Structural	Spatial
T2	Function verification check	6 Pod hydraulic section	7 LMRP connector				X	X
T2	Function verification check	20 BOP mandrel	7 LMRP connector	O			O	O

2.3.4 Vector representation

An effective way to represent the data set is to envision a vector of 25 tests associated with each of the off-diagonal entries in the 67 x 67 DSM. Abstracting to a higher level, and given that each test is classified into one of three levels (subsystem, dock, or subsea), we set a three-dimensional vector behind each interaction in the DSM. We label each test level numerically; subsystem is Level 1, dock is Level 2, and subsea is Level 3. We further add details so that a vector in each DSM cell captures the integration test sequence coverage (i.e., for individual interaction, for each of 25 integration tests spread across three levels), for five types of dependency dimensions (spatial, structural, energy, information, and flow). This yields an augmented 67 x 66 x 25 x 5 (i.e., 552,750 potential interactions in the full data set, most of which are null because the matrices are sparse) vector data set that captures the multilevel complexity associated with the system development and integration test architecture.

2.4 Results

We have explored several alternative data aggregation mechanisms to visualize these data vectors. In order to improve the ease of visualization during multilevel information comparison, we present these data by levels, and then use the maximum and minimum filters to construct maximum and minimum integration level DSMs for each of the five dimensions in our data set. For ease of exposition, we present the results on only one dimension (structural) out of the five dimensions of interactions. As explained in Section 2.3.1, these structural interactions are only presented for a subset of the data (that form a DSM with rank 56) instead of the full data set (with rank 67).

2.4.1 Interactions by levels

Figures 2-4 through 2-6 depict the results of queries by different levels in DSMs. For instance, "1" (and "0") in Figure 2-4 show the structural interfaces that are (or are not) addressed by subsystem-level tests. There are a total of 190 structural interactions within our five subsystems. They are identified in the gray segments of Figure 2-4. A total of 126 of these interactions are addressed during subsystem level tests and they are marked "1," and the other 64 are marked "0." A small number of interfaces across subsystems are also tested at the subsystem level because those subsystems are tested together; there are 10 such interfaces between the LMRP and BOP, and 4 between the C&K system and Aux Lines. Similarly, the "2" (and "0") in Figure 2-5 show the interfaces that are (or are not) addressed by set of dock system-level tests. Finally, Figure 2-6 uses "3" and "0" to identify the interfaces tested in the subsea system-level tests. In principle, every interface can be tested at the subsea level because the full system is installed in its operational condition. It is clear that not all tests are relevant to each interface. It is also evident that the test suite we analyzed has very different distribution of coverage at the subsystem, dock, and subsea levels of tests.

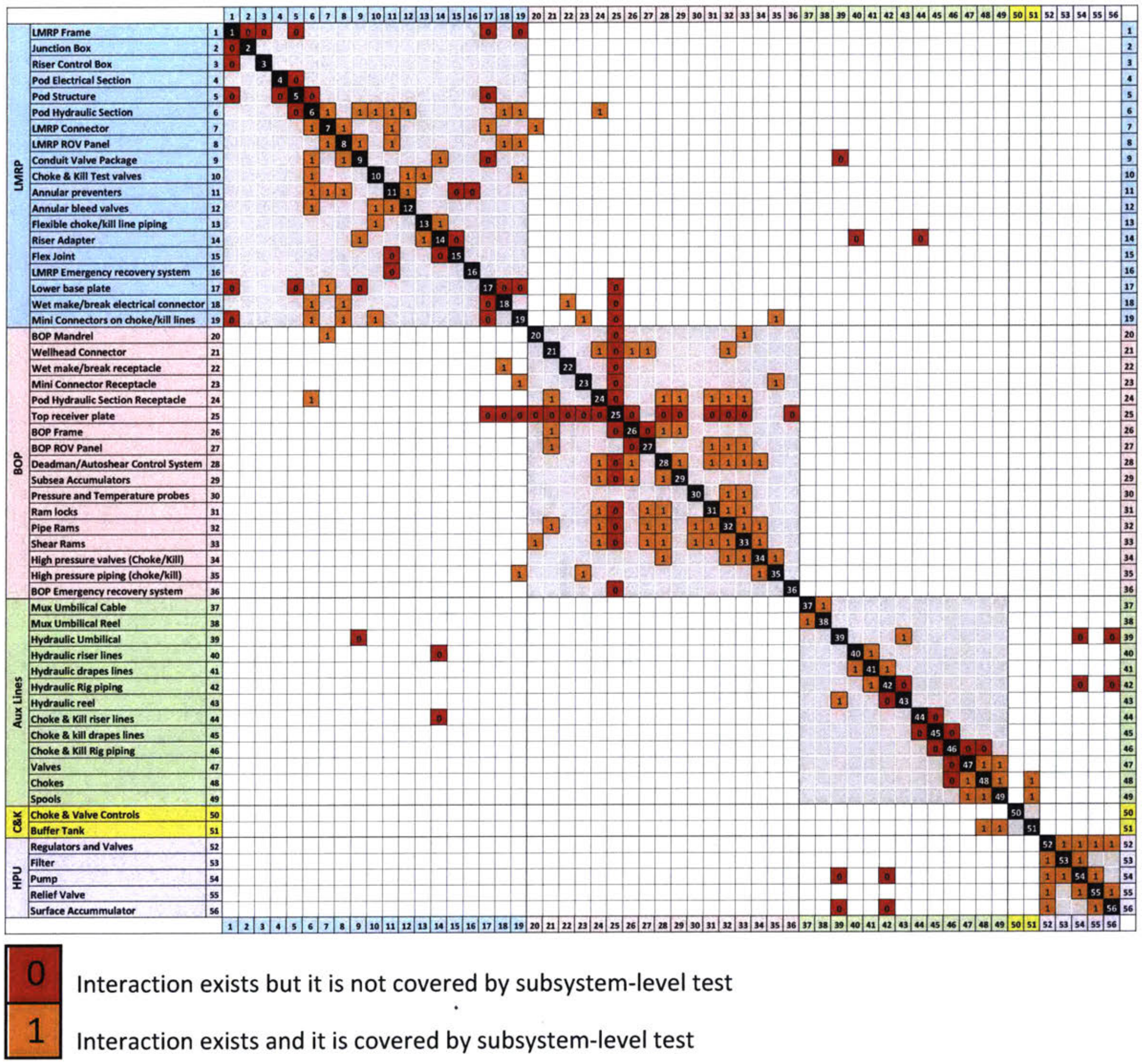

Figure 2-4: Multilevel structural interaction design structure matrix showing the subsystem test level.

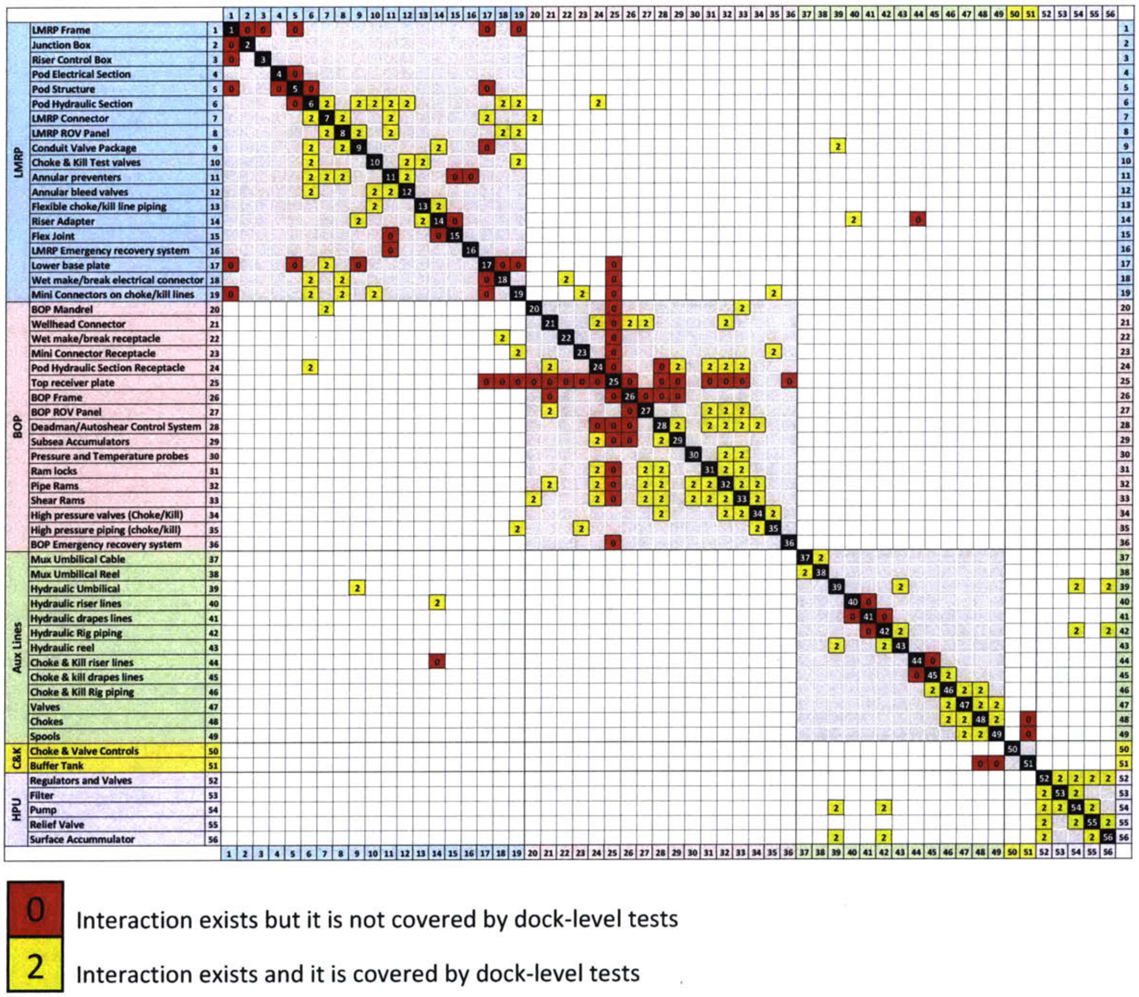

Figure 2-5: Multilevel structural interaction design structure matrix showing the dock test level.

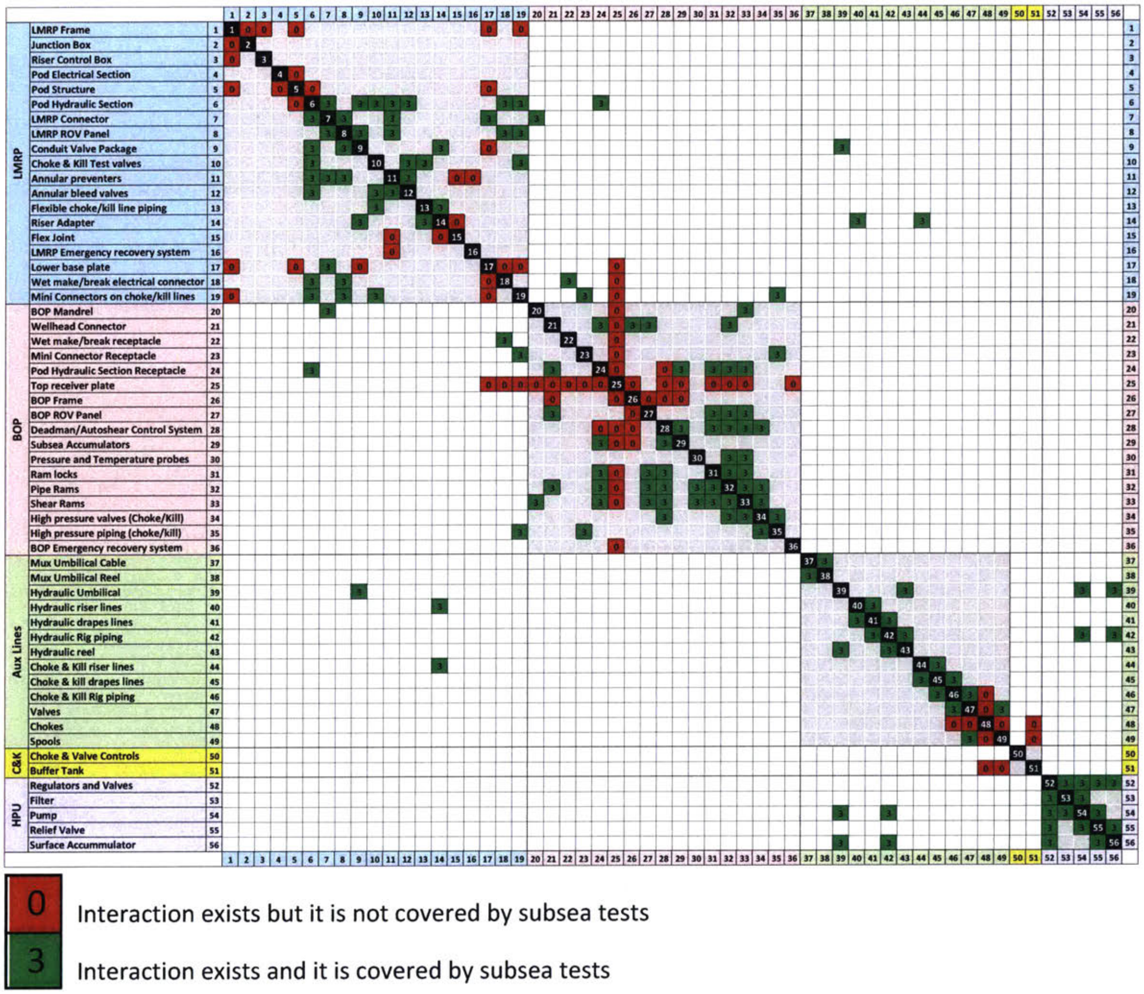

Figure 2-6: Multilevel structural interaction design structure matrix showing the subsea test level.

2.4.2 Multilevel output: Maximum integration level

Figures 2-7 and 2-8 combine interaction marks from multiple test levels. The off-diagonal terms in the DSM can be filtered out of the data across multiple levels to reveal the highest test level at which each interaction is tested, per dimension. We map the largest index of a positive test level in the vector corresponding to each of the interactions onto the system architecture DSM. The maximum integration level DSM for the structural dimension is presented in Figure 2-7. For example, a "1" in the maximum integration level DSM indicates that that particular interaction is last tested at the subsystem level and is not at all tested at the dock level or subsea level. Given no constraints on resources, an ideal system validation procedure would have all interactions tested at the final test level in the sequence. In this way, all interactions are tested in the most completely assembled configuration and in the most realistic setting to actual operational conditions. A system that is fully tested at the subsea level would lead to a maximum integration level DSM in Figure 2-7 with every interaction entry a dark green "3." A red entry of "0" indicates that the interaction does exist but is not tested in any of the integration tests in this data set.

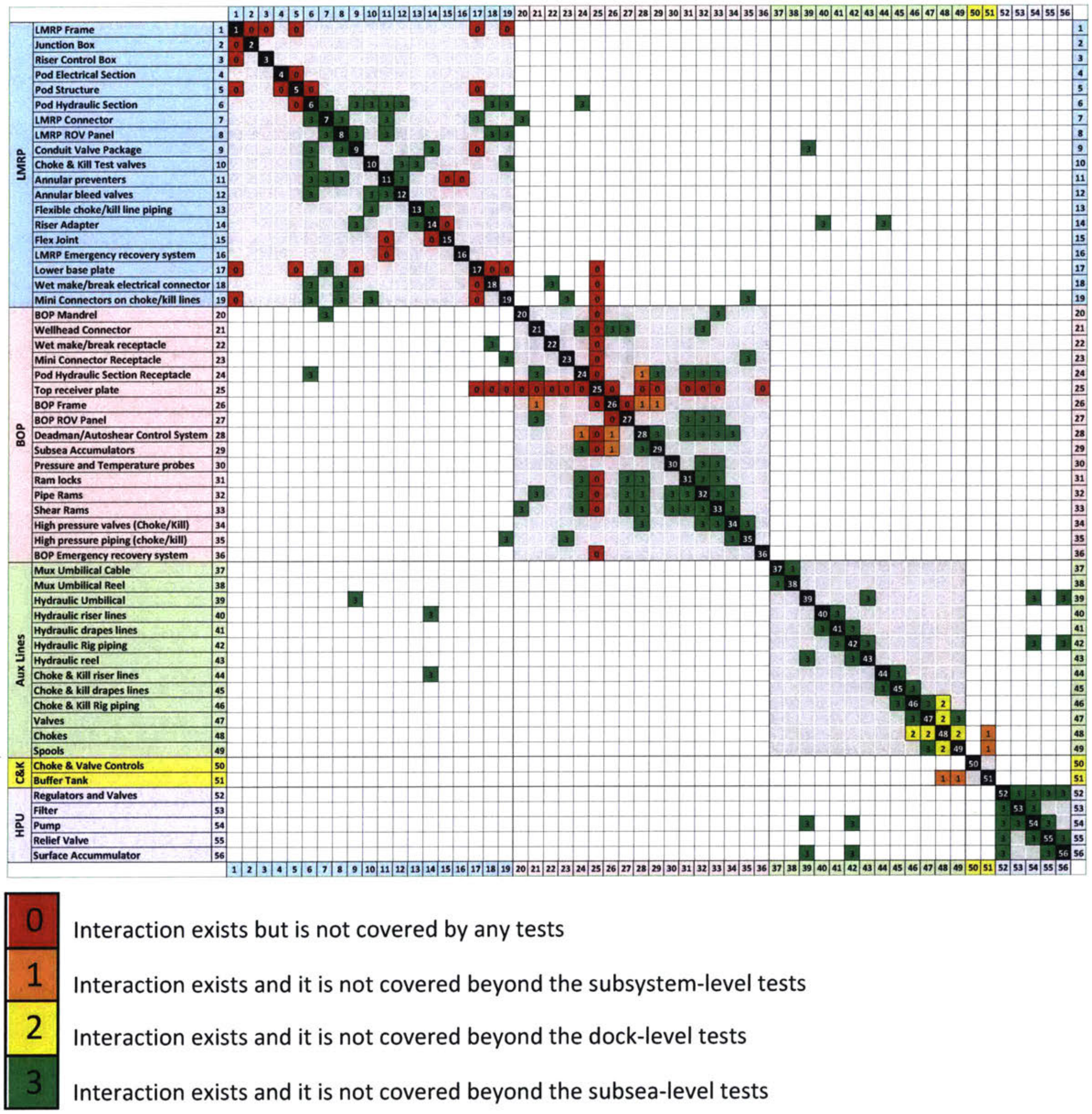

Figure 2-7: Maximum structural integration level of the design structure matrix.

2.4.3 Multilevel output: Minimum integration level

A second useful way to present the integration test data is the minimum integration level DSM. Such a DSM shows the first level at which each interaction is tested in each dimension. The minimum integration level DSM for the structural dimension is presented in Figure 2-8 . The data displayed in the DSM are the result of a minimum search of the test-level vector for each interaction. A red entry of "0" in the DSM indicates that the interaction is not tested in the integration test sequence in any assembled configuration. From the minimum integration level DSM, we would ideally see that each interaction within a subsystem would be first tested at the subsystem level. This area is shaded gray for clarity of visualization. Therefore, all of the entries in the gray shaded area along the diagonal should ideally be "1." In Figure 2-8 we see that, within the Aux Lines subsystem, 14 interfaces have the ideal "1" value, 8 are "2," and 2 are "3." For interactions between those subsystems that are assembled prior to subsystem level tests (BOP and LMRP, C&K and Aux Lines), we also see a "1." We would also expect that any intersubsystem interaction could not be tested until the second or third (dock or subsea) levels, because those interactions do not exist for testing before the subsystems have been assembled. Therefore, the DSM entry for those interactions outside the gray shaded area would be a "2" or a "3." For example, an interaction between a component in the LMRP subsystem and the Aux Lines subsystem could only first be tested at the dock or subsea level. This DSM is a map of when information regarding interaction performance is revealed within the SE-V process. An ideal testing protocol would reveal as much information about the performance of the interactions as soon as possible, revealing issues and risks early to allow time for mitigation, rework, or redesign. From this interpretation, the ideal minimum integration level DSM would show that all intrasubsystem entries are tested at subsystem level (all entries are "1") and the intersubsystem entries are all tested when assembled (all entries are "2").

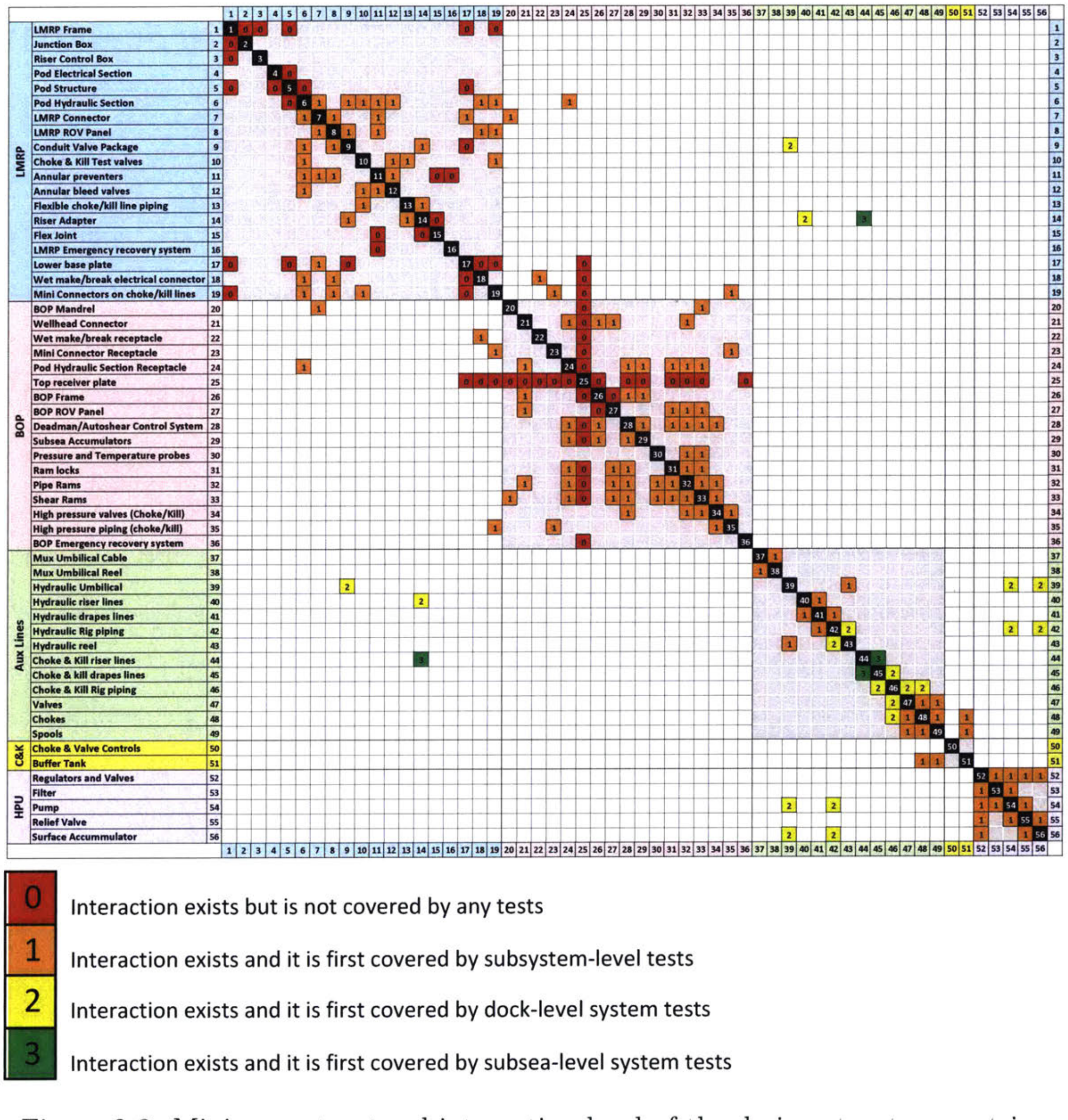

Figure 2-8: Minimum structural integration level of the design structure matrix.

2.5 Discussion

In many industries, test procedures are based on regulatory requirements and industry standards. Such standards do not tend to specify tests from an interaction point of view. The DSM-based query of interactions is a different lens through which the completeness of the test set can be considered. Thus, this analysis has the potential to reveal previously undiscovered information and insights to systems engineers.

2.5.1 Potential for unanticipated outcomes

Upon examination of the maximum integration level DSM, we see in Figure 2-7 that two-thirds (66%) of the interactions are tested to the highest test level (subsea) in the structural dimension; however, a quarter (26%) of the interactions are not being tested in the integration test set at all. For instance, we observe that all of the interactions involving the top receiver plate and all of the interactions involving the LMRP frame are not structurally tested during system integration. This is because these two components are not instrumented with strain gauges during these tests. We presume such instrumentation would require costly or time-consuming procedures in order to check these interfaces after assembly. Thus, it is possible for the multilevel analysis proposed in this paper to yield outcomes that can point to opportunities to improve the integration stage of the SE-V process. A deviation from the ideal test level discovered through the maximum and minimum integration level DSMs may either prompt a redesign of the interface or call for additional instrumentation on the existing interface so that it can be tested. Furthermore, it may induce the development team to introduce additional integration tests. One caveat to these findings is that the quality of output in terms of completeness of coverage is predicated upon the completeness of the chosen integration test suite. In many complex systems ranging from offshore oil operations to mission critical software development [42], it is difficult to include all the test conditions and their combinations. It is therefore common to use a range of test cases (sometimes known as regression tests) to create adequate test coverage. In any case, DSMs (shown in Figs. 2-4, 2-5, and 2-6) provide useful maps for designing test coverage and for debugging structural failure modes. Such findings are not limited to the structural dimension. We have studied the maximum and minimum integration level DSMs for the other four dimensions (not shown here). For instance, the information dimension DSMs show that, within the scope of the 25 tests we considered, the interface between the pod hydraulic section receptacle and the deadman/autoshear control system is not tested beyond the subsystem level. This analysis of integration-phase testing raises the possibility of potentially revealing unanticipated failure modes and when additional tests should be performed, either at the subsystem or system level.

2.5.2 Insights from multilevel analysis

A key contribution from this paper lies in the manner in which test and integration data are represented within the DSM. The use of maximum and minimum functions is merely one analytical approach for improving outcomes based on this representation. Other analytical formulations are also possible. The choice of query and formulation function depends on the question being asked. For instance, we have examined the data generated by alternative multilevel queries (one set for each dimension of the 25 tests, disaggregated by levels, listed in Table 2.1) to figure out either how early or how completely a particular test may address integration issues at a given level of analysis. We have also examined the failure modes associated with an aggregate (i.e., a single level) map of the product architecture by querying the DSM representation that yielded measures, such as "network centrality," and provided insights on whether the network position of a component contributed to system failure. Such results are not presented in the current manuscript for brevity. The minimum integration level DSM reveals that in the structural dimension, some interactions are not tested until the subsea level, even though these interactions are present earlier in the test sequence (assuming that subsystems are assembled first). Many of the auxiliary lines interactions exhibit this behavior, likely because they are not yet assembled for dock tests because they are too physically large. Further, we see that some intersubsystem interactions are not tested until the subsea level despite the fact that the interacting components may be fully assembled, although not in the deployed environment, in the second (dock) level. There is only one example of such an interface, that between the choke and kill riser lines and the riser adapter. The maximum integration level DSM (see Fig. 2-7) reveals that in the structural dimension, some interactions are tested at the subsystem and then are not tested as the system progresses through integration. For example, the interactions within the BOP subsystem between the BOP frame and the wellhead connector are tested at the subsystem level but are not tested at the dock or subsea system-level configurations. Thus, the multilevel timing information revealed in the maximum and minimum integration level DSM analyses shows which of the interfaces are tested early and late in the integration process. Based on their coverage of interfaces, a design team can assign different levels of risks to the integration plan. This observation gives rise to questions of how the dock testing and subsea testing scope are decided. For instance, we found that in the material dimension minimum integration level DSM, that all of the intrasubsystem interactions are tested at the ideal time, as soon as possible, except for those involving the flex joint, which are not integration tested through the set of tests examined in this work. The interaction information in the DSM representation is restricted to our review of engineering documentation, followed by inputs provided by subject matter experts. It is possible that other interactions exist, but they are neither reported in the documentation nor anticipated by an expert. It is also possible that some potential failure modes might precipitate through a combination of interactions. This heightens the need for careful design of the integration phase in the SE-V through a series of tests to uncover unanticipated interactions or combinations of interactions. The rigor of the method described in this paper is restricted by the representation schema and data that we have captured. It does not guarantee completeness of the test coverage. It also does not rule out the possibility of unanticipated failures during integration tests. The DSM representation can inform failure model and effect analysis [24] in terms of interaction pattern identification and coverage while exploring the causes for unanticipated failures. INCOSE [43] recommends an integration process that "verifies that all boundaries between system elements have been correctly identified and described." DSM representation and allied maximum and minimum integration level analyses can complement several useful alternatives for investigating system integration: the hazard and operability study [25], network reliability modeling [44], and so forth. Our initial field study has restricted the scope of the work to five dimensions of dependencies: spatial fit, structural load, energy flow, information flow, and material (fluid) flow across only two domains (component and testing) from a list of five domains shown in Figure 2-2. The current analysis is preliminary and limited to demonstrate a proof of the multilevel analysis concept. Thus, we have restricted the analysis of the interactions to a single dimension, in this case, structural, as shown in Figures 2-4 through 2-8. In reality, there can be significant interactions across the five dimensions. For instance, a structural load may cause deflections that could create spatial misalignment while making hydraulic line connections. It is possible to augment the analysis, by constructing combinations of interaction measures. We leave such an analysis as an extension for follow-on work.

2.6 Conclusion

The research underlying this project, and the method outlined in this paper, are at an early stage of development. Multilevel analysis of DSMs developed in this study contributes to the design of complex engineered systems by addressing two gaps: first, it develops a data collection and mapping methodology to account for the amount of coverage available at each interface within DSM representation of complex SE-V processes; and second, it offers a theoretical basis and a method for data aggregation and query that accounts for differing scales, in terms of both level and timing, to explore if different types of integration risks may be evident at different time scales. Design and analysis of complex engineered systems is a growing research area that calls for systematic and rigorous approaches based on advances in complexity and behavioral sciences [45]. Augmented vector DSM data and visualizations presented in this paper can lend themselves to further analysis. For instance, multilevel data can be used to inform the development of system architecture decomposition options and optimal sequencing of the integration tasks based on design for testability and design for reliability considerations. Developments based on detailed understanding of interactions at each interface, captured in the off-diagonal cells of a system architecture DSM, may yield novel integration risk metrics, algorithms, and behavioral research opportunities for improving complex system design early in the SE-V process.

Chapter 3

Technology readiness levels: shortcomings and improvement opportunities

3.1 Introduction

Many of today's highly innovative products and systems are built around new technologies. These technologies can enable operation in even more extreme environments, and can lead to the introduction of radically new products to open whole new product categories. Some believe that emerging technologies will be the foundation of solutions to the most challenging global problems of the 21st century [46]. Given this potential upside of new technologies, a huge amount of resources is spent on R&D and technology development activities to progress the readiness of technological ideas to the point where they are embodied in a component that can be incorporated into a new product or system. These activities are risky; a great deal of uncertainty exists around the new technology, such as its performance, integration and reliability [47, 48]. Engineering managers must carefully plan processes that not only develop the technology itself, but also prepare the technology to be integrated with the rest of the product in a project at the right time [49, 50, 51, 52, 53]. Technology immaturity has been shown to cause project cost and schedule growth [54, 55, 56, 57]. How does one know when a technology is ready to transition from research and technology development to a product- or system-development project? How does one measure the technology's readiness progression? When faced with these questions in the 1960s and 70s, NASA engineers developed the technology readiness levels (TRLs). Since its inception, the use of the TRL scale has spread beyond NASA, and it is now applied not only on aerospace and defense projects but also projects in industries such as energy, transportation, and electronics. Despite this increase in uptake of practice, little research has been formally conducted to understand and describe TRL application. In this work, we acknowledge that the TRLs deliver benefits as a maturity assessment tool, but also bring challenges of which practitioners should be mindful. In this paper we share the findings of a broad, two-stage study of state-of-the-art use of TRLs. We present an investigation, assessment and discussion of the challenges of modern use of the technology readiness level method. We start with a brief background on TRLs including a history of their use and development. Next we present the findings of a series of interviews with TRL practitioners which revealed 15 challenges of modern TRL use. We classify these 15 challenges into three different categories: system complexity, planning and review, and validity of assessment. We then expand the scope of the study with a large-scale survey of TRL users, which not only provides a better picture of who uses TRLs today, but also prioritizes the 15 TRL challenges. We find that the challenges related to system complexity are particularly critical to TRL users. We share little-discussed practices from industry that address some challenges, as well as solutions from the academic community. We conclude by identifying improvement opportunities and directions for future research related to these challenges.

3.2 Background

The technology readiness level scale was developed at NASA in the 1970s to be a consistent measure of technology maturity. NASA first externally published a readiness scale in 1989, in the form of a 7-point scale [58], and later in 1995, published the familiar 9-point scale with detailed descriptions of each level [9]. Continued refinements have resulted in the most modern form of the TRL, shown in Table 3.1, with procedures available online [1, 2]. At TRL 1, the scale starts with a technology in a very basic scientific form, and progresses to a technology proven in the operating environment at TRL 9. The technology readiness assessment provides an evidence-based evaluation of the progression in maturity of the technology towards eventual operation. The assessment uses data generated during the course of development, such as test results. In the TRL literature and at NASA, the term "technology" is most commonly conceptualized at the level of a component technology featuring new materials, scale or working principles. The component technology of interest could be a new invention or an adaptation to an existing technology. The TRL is used to assess a technology from when it starts as a research finding, through its embodiment in a component, until eventually it is integrated into a system. We adopt the same conceptualization of "technology" for this paper. The TRLs provide a number of benefits to their user. In the initial TRL publication [58], the author describes that the TRLs were motivated by a need for a shared understanding of technology maturity across NASA. This shared understanding is achieved through a standard language that can be used across disciplines, organizations and functions in order to better communicate and assess risk. The TRL assessment is complementary to existing project management, systems engineering, and oversight processes, and can inform other assessment and planning activities. TRL assessments can be used to monitor progress of a developing technology, and to identify and manage technology-related risks. A TRL assessment does not eliminate risk in and of itself, but it does illuminate risk. When selecting between multiple alternative technologies fulfilling the same function, their relative maturity assessed via TRL can be very informative. The TRL is also particularly useful at times of technology hand-offs as it can facilitate information exchange between different groups in the same organization (for example, between R&D and projects) or between technology supplier and customer. Additionally, the progressive levels provide a systematic approach and model for technology-intensive system development, with the TRLs serving as guideposts. The achievement of TRLs can trigger additional resource investment, scaling this investment in line with the technology's readiness. Some view the TRLs as a path from science to marketable technology, through a series of increasingly complex physical artifacts as embodiments of technology [59]. A useful technology readiness assessment follows a repeatable process, relies on sufficient evidence, and results in a credible output that is useful to all involved parties, including project managers, technology developers, and governance bodies [60]. TRLs have recently been increasingly applied outside of the aerospace context; new government and commercial implementations of the TRLs are often similar to NASA's 9-point embodiment, typically with slight changes to tailor the language to the context. Examples include guidelines customized to the contexts of defense, oil and gas, and infrastructure [61, 62, 63, 64]. The most comprehensive publically available technology readiness guideline published to date is from the US Government Accountability Office (GAO) [60]. The guide is currently in exposure draft form and will be updated based on feedback submitted through August 2017.

Table 3.1: NASA Technology Readiness Level Scale. [2]

TRL	Definition
9	Actual system "flight proven" through successful mission operations
8	Actual system completed and "flight qualified" through test and demonstration
7	System prototype demonstration in an operational environment
6	System/subsystem model or prototype demonstration in a relevant environment
5	Component and/or breadboard validation in relevant environment
4	Component and/or breadboard validation in laboratory environment
3	Analytical and experimental critical function and/or characteristic proof-of-concept
2	Technology concept and/or application formulated
1	Basic principles observed and reported

One popular means of implementing the TRL scale is through mapping TRL goals to the organization's system development process, i.e., target TRL requirements are assigned to the gates in a development process. Such a mapping exists for the US Department of Defense (DoD) System Acquisition Process, as represented in Figure 3-1 [62]. These TRL mappings facilitate consistent and explicit expectations of technology maturity across projects. This practice also ensures that the technology maturity requirement is considered in the decision to pass the gate reviews. These reviews are shown as milestones A, B, and C in diamonds in Figure 3-1. Although limited, there exists some evidence to suggest that the mapping of TRLs to the system development lifecycle is a helpful best practice. A GAO study of 62 DoD acquisition programs found that those programs which reached TRL 7 or higher by Milestone B (the start of Engineering & Manufacturing Development in Figure 3-1) generally finished on time and on budget, whereas those programs with technologies below a TRL 7 showed, on average, development cost growth of 32%, acquisition unit cost increase of 30%, and schedule delay of 20 months [55]. More formal academic studies have confirmed these findings: Studying TRL and schedule slip data from 28 NASA programs, Dubos provides a formal analysis to support the GAO's TRL 7 recommendation [56]. Another study of 37 DoD weapon systems showed that GAO's technology maturity guidance had a statistically significant effect on the schedule overrun of these systems [57].

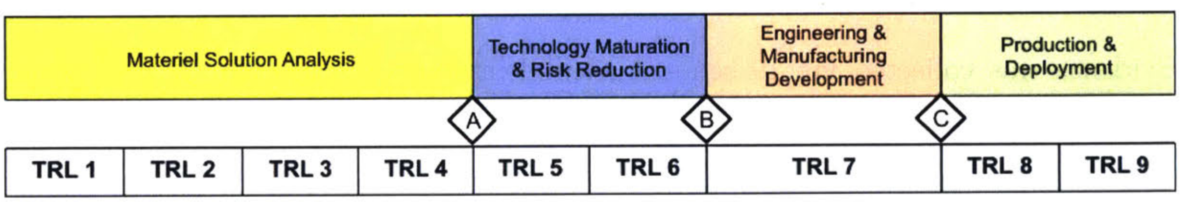

Figure 3-1: Mapping of technology readiness levels to US Department of Defense System Acquisition Process. Technologies are expected to achieve TRL 4 by milestone A, TRL 6 by milestone B, and TRL 7 by milestone C.

Despite the increasing maturity of and confidence in the TRL method at NASA, we see evidence of challenges to implementation and effectiveness. In a 2009 TRL retrospective, Mankins, who first published the 9-point TRL scale, concluded with a short description of two TRL challenges based on his experience in the aerospace context [65]. The first is "achieving the right level of technology maturity across multiple subsystems and components is an ongoing challenge to development success," while the second is a lack of "practices and metrics that allow assessment of anticipated research and development uncertainty." Our study builds on this work by systematically describing and prioritizing an array of challenges of TRL use.

3.3 Data Collection

We present in this section the details of our data collection process. Our study followed an exploratory sequential mixed methods design Creswell2014. We began by exploring TRL shortcomings through qualitative semi-structured interviews. We analyzed the interview transcripts to create a complete set of challenges revealed by our participants. We used these findings in the next phase, a large-scale survey of TRL users in industry, where we measure the relative criticality of these challenges. Through this multi-step method, we can see if the findings based on detailed data from select technology developers can generalize to a more complete sample of TRL users. In the Results section, we present the results of these two phases in an integrated manner.

3.3.1 Interviews

Evidence was collected via 19 semi-structured interviews with employees at seven different organizations.

Sample

The set of organizations was selected on the basis of diversity across a number of measures, so that we could be exposed to a complete perspective on the TRLs. Our selection reflects a diversity of development lifecycle lengths, degree of regulatory oversight, and competitive environment. Additionally, the organizations are at varying degrees of maturity in their own TRL processes and experience; Google X, for example, was relatively new to using TRLs, whereas NASA is where TRLs were developed in the first place. Details of the interview participants are available in Table 3.2. In each case, our interviews were conducted with the part of the organization responsible for hardware development (so at Google X, for example, the consumer electronics group). Where possible, we interviewed a variety of roles within the organization (from engineer to project manager to executive) in order to receive a more complete perspective. In each case, we interviewed employees with TRL experience, whether they were responsible for, involved in, or use the outputs of technology maturity assessment. In four cases, we interviewed the employee responsible for establishing and maintaining TRL guidance at the organization. In addition to semi-structured interviews, we reviewed company- and industry-specific technology development guideline documents.

Table 3.2: Interview participants.

#	Organization	Industry	TRL Experience	Role of Interviewee
1	NASA	Space	> 8 years	Office of the Chief Engineer
2				Office of the Chief Technologist
3	Raytheon	Defense	> 8 years	Director of Engineering
4	BP	Oil & Gas	2 - 8 years	Technology Leader
5				Engineering Manager
6				Technology Manager
7				Engineering Manager
8				VP Technology
9				Project Manager
10				Independent Consultant
11	Bombardier	Aircraft	2 - 8 years	Senior Engineering Specialist
12				Systems Integrator, Advanced Design
13	John Deere	Heavy Equipment	2 - 8 years	Systems Engineer
14				Systems Engineering Manager
15	Alstom	Power Systems	<2 years	System Engineer
16	(now GE)			Risk Expert
17				Process Expert
18	Google X	Electronics	<2 years	Program Manager
19	(now X, Alphabet)			Product Design Lead

Analysis

We transcribed and coded the interviews, identifying references to TRL-related challenges. We started with predetermined codes (challenges) from the literature, and added or removed based on emerging evidence from the transcripts. Next we grouped the evidence by similarity of concept. We developed several different groupings, and eventually chose a set of 15 concepts (challenges) that were thought to be internally consistent and conceptually distinct. In the Results section, we provide a description of each challenge that was synthesized from this evidence. Additionally we include specific data in the form of illustrative quotations from interview participants. A table of additional illustrative quotes for each challenge is also provided in the appendix.

3.3.2 Survey

Design

In order to learn further about the generalizability and prioritization of the challenges discovered in the qualitative stage of this research, we next surveyed a large sample of TRL practitioners on the criticality of the challenges. We also gathered information regarding the characteristics of TRL users, and TRL-using organizations. The survey was structured in three parts: Part 1 consisted of background information about the respondent in order to characterize the TRL user community. The second part was the core of the survey and asked about the relative prioritization of the challenges using the best worst scaling (BWS) method, which we will describe in more detail below. Part 3 asked open-ended questions about the completeness of the list of 15 TRL challenges. For respondents who had not used TRLs in their work, the survey would end after the demographic background questions in part 1, given that we targeted informed TRL users for our survey sample. The time needed to complete the survey was typically 15 to 20 minutes.

Sample

The sample for this survey was drawn principally from members of the International Council on Systems Engineering (INCOSE), a worldwide professional organization dedicated to the advancement of systems engineering with approximately 9800 members. All regional INCOSE chapter presidents were approached individually by email and phone. In total, 34 chapter presidents agreed to aid in recruitment for the TRL study by emailing a link to the survey to their chapter members. The survey was distributed to approximately 5370 INCOSE members worldwide. In order to reach the widest audience possible, a survey link was also posted in TRL-related discussion groups on professional online platforms (e.g. LinkedIn) and snowball sampling was allowed by giving subjects the chance to forward the survey to other users with TRL experience. INCOSE members make up 76% of the total survey responses. 190 responses were received during the two-month data collection period (July to September 2015). Responses from non-TRL-experienced subjects, multiple responses from the same IP address, and those completed in unreasonably short time were excluded from the analysis. Overall 113 responses were used in our analysis. Respondents were primarily from North America (68%) or Europe (31%). A majority of the responses came from either aerospace (31%) or defense and military (38%) industries. Respondent organizations include for example Airbus, Baxter, Boeing, Carl Zeiss, Dana, Fujifilm, General Electric, Lockheed Martin, NASA, Rolls Royce, SpaceX, and the US DoD.

Analysis

The core of the online survey consists of a BWS experiment, a method that elicits and quantifies relative importance. The data were analyzed to reveal users' placement of the previously identified 15 TRL challenges on a latent, subjective scale of "criticality". BWS is a choice-based measurement method which aims to take advantage of respondents' abilities to reliably and accurately identify extremes in a set. The method, motivation and theory behind BWS is well described in a text by Louviere, Flynn and Marley [66], but has little been applied in engineering and operations contexts [67]. In a BWS survey, respondents are shown a subset of the total set of objects, and are asked to identify the "best" and "worst" objects. In our case, the objects of interest were the 15 challenges, and the respondents were sequentially shown 15 different sets of five challenges, and asked within each set to identify the most- and the least-critical challenge. It is then possible through analysis to derive aggregate preference scores from these individual choices. The BWS is based on random utility theory [68], and therefore the survey must be designed such that each object appears equally as often, and co-appears with other objects equally as often. We used the balanced incomplete block design [66] from Sawtooth software (Sawtoothsoftware.com) to satisfy this requirement. BWS surveys are chosen over a more common Likert rating scale or a ranking task for a number of reasons, including acquiescence bias, and extreme response bias [66]. Further, from the respondent task point of view, it is easier to process and make judgments about subsets than about the full list of objects - for example, it would take a great deal of concentration and effort to rank all 15 challenges from 1 to 15. BWS provides the improved reliability of paired comparisons, without requiring the subject to evaluate all 105 pairs. We collected demographic and descriptive information about our TRL user sample. In the next section we report differences in the criticality scoring of the challenges between subgroups created from the following descriptive variables: industry, size of organization (number of employees), respondent function within the organization (engineering, research, consultant), respondent's personal experience with TRLs, respondent's frequency of TRL use in current position, organization's TRL experience, inclusion of TRL in internal standard, whether TRL use is required by a customer, and respondent's TRL responsibility on the scale. These differences in mean best-worst criticality scores are found via either a two sample t-test or ANOVA if more than two subgroups exist. We observe similar variances in the samples. We report differences at the 5% significance level. Correlations between descriptive variables were performed and are detailed in a conference paper [69].

3.4 Results: 15 TRL Challenges

We have grouped the 15 challenges into three categories for clarity of presentation, as shown in Table 3.3. Some challenges bridge categories, and could have been placed in more than one category; in these cases we have chosen to place them in the most appropriate section based on the evidence. The challenges are numbered from most to least critical on average for the survey sample. Within each category, we present these challenges in order of most to least critical overall according to our survey results.

Figure 3-2 shows the mean criticality scores based on the 113 complete survey response sets. The summary data are available in the appendix. Criticality scores are calculated by taking the number of times a challenge is picked as Best and subtracting

Table 3.3: Challenges encountered in modern TRL implementation.

A: System Complexity	**B: Planning and Review**
A1. Integration and connectivity	B5. Confidence to progress
A2. Interface maturity	B7. Effort to progress
A3. Influence of new components or environment	B10. Aligning TRLs with system development gates
A4. System readiness	B11. Waivers
A8. Scope of the TRL assessment	B13. Back-up plans
A12. Prioritization of technology development efforts	B14. Product roadmapping
A15. Visualization	**C: Assessment Validity**
	C6. Imprecision of the scale
	C9. Subjectivity of the assessment

the number of times it is picked as Worst, then dividing by the number of times the item was seen. A criticality score of 1 would indicate that every respondent selected the challenge as most critical each time it was presented in a set; a score of 0 corresponds to a challenge that was picked an equal number of times Best and Worst; and a score of -1 would indicate that it was selected as least critical each time it was presented in a set. Marley and Louviere have shown that these scores are strongly and linearly correlated with outputs of a more complex conventional multinomial logistic regression or conditional logistic regression choice models, and thus are adequate for judging overall relative criticality measures for the sample [66, 70]. Nevertheless, the results from the counting analysis were confirmed with the more complex hierarchical Bayesian estimation, the details of which can be found in an earlier publication [71].

We further analyzed the survey results to reveal differences in prioritization of the challenges based on the descriptive statistics collected. In Figure 3-2 we also graphically depict the findings from this further analysis by group, including any results found to be significantly different at the 5% level. No significant differences were found when the prioritizations were investigated by number of employees at the organization and whether using TRL is required by customer or not. In the following sections we describe each of the fifteen TRL challenges with evidence from our interviews and provide criticality ranking data from the survey. Each of the challenges described in this paper can also be seen as an opportunity for improvement, and we believe that potential solutions may exist for all 15 challenges. We have collected and will share creative and effective practices from industry, as well as from the literature, alongside our description of the challenges themselves.

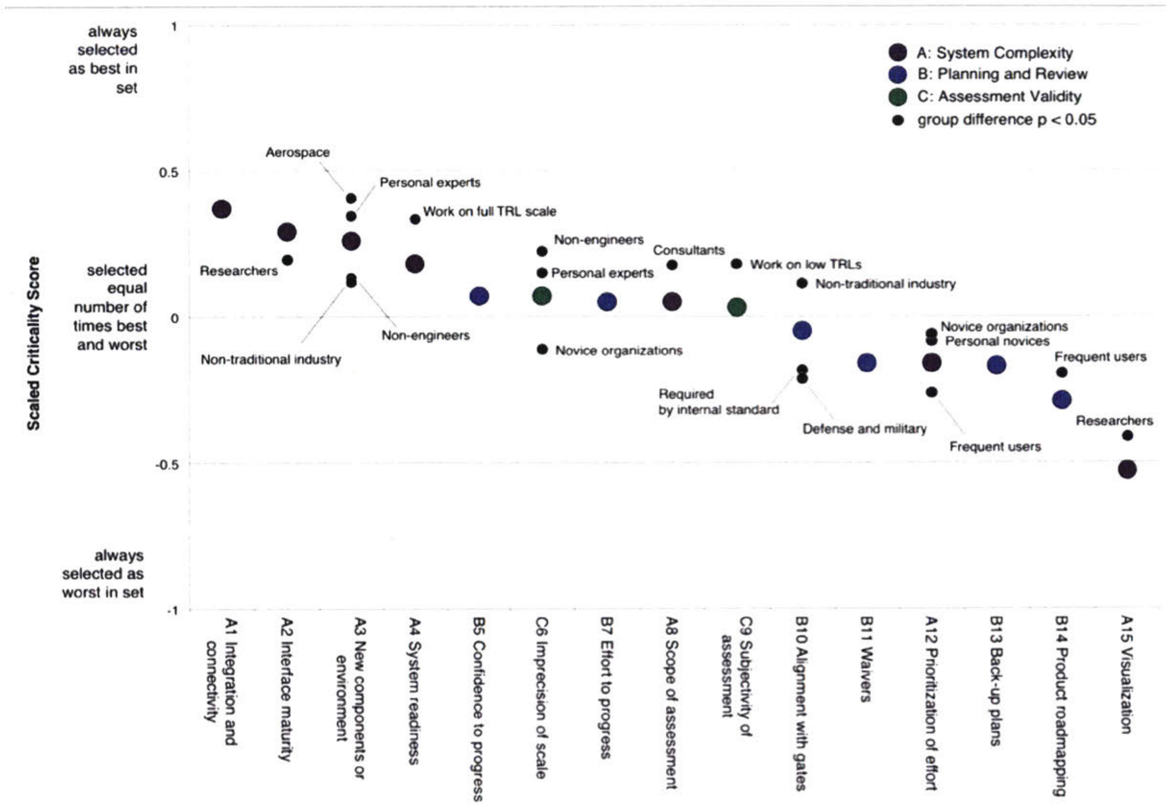

Figure 3-2: Large circles represent average criticality scores of 15 challenges based on responses from 113 survey respondents. Smaller circles indicate significantly different (p-value $< 5\%$) criticality scores for sub-groups of the survey population.

3.4.1 System Complexity

In this section we present challenges emerging from the inclusion of new technologies in increasingly complex systems.

Challenge A1 - Integration and connectivity: The technology readiness levels were developed with the intention of assessing an individual technology. Although the higher TRLs acknowledge that a component progresses during development from being on its own, to being part of a subsystem, to a final system, the levels offer limited insight into integration, which is a key challenge faced by development programs. There is no acknowledgement of the component as a part in a connected network with dependencies, or architecture, where a change to one component would affect another. A systems engineer at Alstom explained that despite the specific TRL definitions, "as far as we're concerned, the technology simply is not at a suitable readiness unless it integrates with what's around it" and the TRL does not support this. When we interviewed at Raytheon, we learned the organization had just launched a business improvement project to look into improved integration readiness assessment. The need for a solution to address this challenge was further emphasized in the survey results; integration and connectivity was found to be the most critical challenge overall, and most critical for all subsets of the survey respondents - no significant difference was found regardless of grouping by TRL experience, industry, function, or any other descriptive statistic collected. Jimenez and Mavris, researchers in the aerospace community, propose an extension to the traditional TRL language in the form of integration- and architecture-focused additions to the descriptions at each level [72], which our interview partners at Bombardier had seen and hoped was a promising approach for tackling the integration and interface challenges.

Challenge A2 - Interface maturity: Component technologies are connected to one another in the system architecture through interfaces. The TRLs do not explicitly assess the maturity of the interfaces, despite the fact that there may be new and novel ways to connect two components. Two mature technologies may interface through a novel immature interface, resulting in an overall system that is not mature. A consultant in the oil and gas industry explained the importance of interface maturity, stating that "it is not unusual that two pieces of equipment mismatch or require several iterations to work together." Yet interface maturity gets no attention in the current technology readiness assessment process. The survey revealed that this challenge is seen as particularly critical amongst all TRL users; it was rated second most critical overall. Only those respondents who work in a research function were found to be significantly different in rating this challenge third most critical, not second. Sauser and his colleagues have introduced the integration readiness level (IRL) scale to the systems engineering community [73, 74]. This work recognizes that technologies are connected to each other through interfaces in the system architecture, and that these interfaces have independently assessable maturities. The IRL uses a 1-9 scale in the style of the TRL to assess the readiness of the interface connecting two components. Although some interview partners were aware of the IRL, none had found it appealing enough to adopt and use.

Challenge A3 - Influence of new components or environment: Often a proven (TRL 9) technology component is chosen for use in a new system which will operate in a different environment or feature a modified architecture. In these cases, assessment of the TRL can be non-obvious. It can seem unfair to discount the TRL of a proven technology, yet the technology is truly only proven in the configuration and environment in which it has successfully operated; this is discussed only in aerospace best practice guidelines [1, 75]. Yet NASA is still wondering about this question internally, as we learned: "An item we're considering [at NASA] is the effect of hierarchy to TRL - that is, if you have a TRL 9 vehicle but you replace something at the assembly, component or even piece part level with a lower TRL, does the entire system then become that lower TRL?" Is this too severe of a reaction to a modification? Careful thought regarding modifications to architecture or changes to environment should lead to guidance in such cases. This challenge was rated third most critical overall by the survey sample. Further analysis reveals that practitioners who are new to TRLs (self-identify as novices or only having some experience) find this challenge to be less critical, ranking it fourth. Those not in the aerospace and defense industries rate this challenge as one rank less critical (fourth), as do those who identify their function as non-engineering (sixth). This challenge has been addressed at NASA through a flowchart used for assessment (Figure 3-3) with architecture and environment explicitly called out. This flowchart makes it clear that unless the architecture and environment are identical, the integration context is not identical, and thus the TRL is only at level 5.

Challenge A4 - System readiness: There is strong interest in an expansion of the (component-level) TRL assessment to a system readiness level measure of maturity. Such a measure would allow managers to reflect on the maturity of the system as a whole, to compare to other current projects in the portfolio or past projects, and even to set system readiness requirements in the technology and product development process milestones. This measure would consider not only the TRLs of the components, but also include some measure of integration or interface maturity to reflect

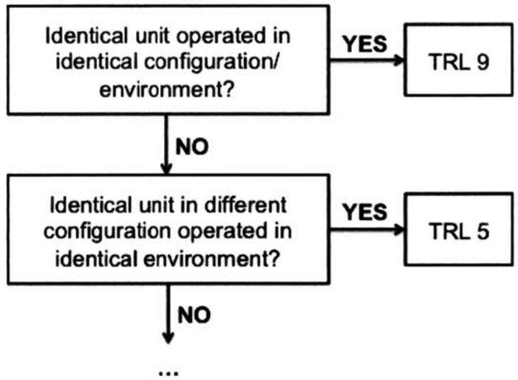

Figure 3-3: Top section of the decision flowchart from NASA Systems Engineering Handbook [1] emphasizing the change in a component's TRL resulting from a change of operating environment.

on the system's full architecture. A technologist at NASA reiterated the desire for an improved system maturity measure, explaining: "System readiness level... everybody wants to use it. People would like to be able to characterize the maturity level of the system. Nobody has come up with anything that's useful yet." Rated fourth most critical overall in the survey, this challenge was found to be particularly critical to those who identified their job responsibility as touching on all 9 of the technology readiness levels (they rated this challenge second most critical). There does exist an academic attempt in the literature to address this challenge, namely the systems readiness level (SRL), accompanying the previously described integration readiness level. The SRL is a 0-1 value computed from the system's TRLs and IRLs. The need for the SRL concept certainly resonates with TRL users, as evidenced by a growing awareness of the IRL and SRL in industry. However our interview partners at John Deere and NASA had tried to use the SRL and did not find it to be tractable. A systems engineer at John Deere explained: "In the literature they were trying to calculate [SRL] by doing some matrix multiplication between the rating, which fell apart right away. We started to try to apply some numbers [...] and it didn't work." The SRL also has academic critics, citing mathematical invalidity [76, 72].

Challenge A8 - Scope of the TRL assessment: When initiating a technology readiness assessment, how does one decide which technologies to consider? Technology readiness assessments can be very resource intensive, yet the assessment is only effective if it captures the components with high technology risk, and a systematic and thorough approach to TRL assessment would consider all component technologies. A consultant in the oil and gas industry described a downside to an oversized assessment scope: "operators use TRL not to manage key technologies, but for tracking readiness of all equipment for installation. Every nut and bolt of every equipment is included in an Excel sheet. You can imagine such a spreadsheet will become very large," implying it would also become hard to manage and less effective as a tool to identify risks. It is a challenge for new adopter of TRL processes to decide how to scope their technology readiness assessment. The survey responses reveal that this challenge is of medium criticality - eighth overall. We also found that those in consulting functions rate this challenge as more highly critical (fifth) than those outside of consulting. We found that at organizations with long-term TRL experience, practice is predominantly split into two approaches. The government and defense industry formal guidelines suggest focusing on a small set of critical technology elements (referred to as CTEs). The GAO guideline includes a chapter addressing scope (challenge A8), called "selecting critical technologies" [60]. The guide elaborates on the definition of critical technology element, and suggests strategies for the decomposition of a system into appropriate technologies for assessment. Yet alternatively we see that NASA instead assesses TRLs for all the components identified in a product breakdown structure (PBS). Those following the PBS approach argue that it is important to have a baseline of technology maturity and not to omit a possible maturity risk or challenge. Yet there is a trade-off in size of scope and cost of TRL-intensive assessments. With arguments in favor and against both approaches, it is a challenge for new adopter of TRL processes to decide how to scope their technology readiness assessment.

Challenge A12 - Prioritization of technology development efforts: TRL assessments are ultimately just that - assessments. What should managers or developers do based on the assessed TRLs? How should one focus efforts and management attention to make the biggest impact on overall maturity? Should the lowest TRL components always be the priority of development resources? As a director at Raytheon explains, "It quickly draws a red line around the low TRLs and suggests the program manager has to put some resources against them to fix it." Yet this approach may narrow the attention of management and hide other, higher risk issues (other TRL advancement hurdles, for example). In summary, currently available TRL guidance does little to address how to prioritize technology development efforts based on TRL information. Survey analysis revealed this challenge to be twelfth overall in criticality. Those who identified using TRLs frequently in their day-to-day work and those who have more experience with TRLs think this challenge is even less critical (fourteenth). We have found the most common method for achieving a system readiness measure is called the "weakest link" method, where the subsystem is assigned the minimum TRL of its components, and the system is assigned the minimum TRL of its subsystems. Action is prioritized to address the weakest link through additional technology development efforts, and increase its TRL. The GAO guideline presents guidance related to a technology maturation plan (TMP) [60]. A TMP establishes a roadmap of the testing or engineering activities required to mature immature technology. The GAO guideline provides a template for TMP reports, allowing other organizations to adapt this best practice to their own context.

Challenge A15 - Visualization: When TRL assessments are complete, what is a useful way to visually present this information? This becomes a challenge if there are many component TRLs that have been assessed. The common approach is to list all components and their TRLs in a spreadsheet. But this representation does not leverage any information on architecture, alternatives, difficulty, or confidence. It also may not draw the attention of decision makers in the most effective way. A Bombardier engineer shared with us that, "we generate lists [of TRLs], and then pretty much use them listlessly." This challenge was ranked least critical overall and nearly universally amongst the survey subgroups, with the only minor exception being those who work in a research function perceiving this challenge to be second least critical. Current practitioners who choose not to use the singular number SRL as a measure of overall system readiness (challenge A4) may instead be well served with improved visualization of all the component TRL information (and possibly integration readiness level information), since a concise visualization of the full system technology readiness could provide the manager with an understanding of the system's overall maturity. NASA's handbook shows a TRL spreadsheet with color coding to highlight low TRLs [1]. The limited academic TRL visualizations that do exist have been based around the product architecture, using the design structure matrix [8]. Brady introduced the technology risk design structure matrix with case study demonstrations of NASA's Mars Pathfinder and Near Earth Asteroid Rendezvous, [77]. In the context of the oil and gas industry, Yasseri presents Sauser's IRL in a system architecture design structure matrix [78]. While we do think these are useful approaches, powerful additions to the architecture view might incorporate information about development risk, for example effort or confidence to progress and change effects that ripple through the architecture.

3.4.2 Planning and Review

In this section we present challenges related to the integration of TRL assessment outputs with existing organizational processes, particularly related to planning, review, and decision making.

Challenge B5 - Confidence to progress: Similar to the challenge previously discussed as effort to progress, the likelihood of achievement of future TRLs should also inform planning and risk assessment. The confidence to progress should be based on an understanding of the obstacles and tests required to mature the technology embedded in the particular component. It is possible that some TRLs pose particularly new or uncertain steps, while others are simple steps that do not require significant investment. As an engineering manager at BP explained, "if you have a small part at a low TRL, you don't want to flag that to threat level red if you know it's a low risk component." Because risk is directly related to confidence and likelihood, there are many ways in which confidence to progress would be useful in technology management decision making. The survey analysis revealed that this challenge is ranked fifth most critical overall, and the highest amongst any challenges in the planning and review category. This prioritization is consistent across all the subsets, as no significant deviations were found from any subset of responses. We found no existing metrics or assessment methods for confidence to progress (B5), despite the perceived criticality of the challenge by TRL users. At Alstom, a high-level proxy for confidence to progress is used in long-term planning to reflect on the makeup of the portfolio, an example of which is shown in Figure 3-4. Alstom, and others, would like a straightforward formulation for confidence to progress.

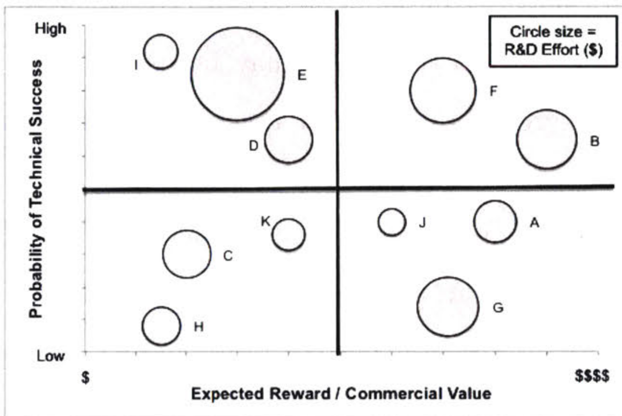

Figure 3-4: Example of Alstom's risk-reward chart, demonstrating a use for the confidence to progress measure. Each prospective technology is represented by one circle.

Challenge B7 - Effort to progress: An assessed TRL tells you the degree to which a component has been demonstrated. It is a snapshot of the current state. The TRL does not provide any information about the effort (time and resources) that will be required to achieve subsequent TRLs. The effort to progress to the next and subsequent TRLs may be important in making planning and technology-selection decisions. Some TRL advancements may require little effort - the testing apparatus already exists, for example - while others may take a lot - a long-term reliability test, for example. A product design lead at Google asked specifically about effort: "What do we need to do to get it to the next stage of readiness? What do we really need to do to really have it secured in our back pocket, and put it on the shelves." Additionally, we learned from NASA that they are not consistently using an effort-to-progress assessment in their standard process, and are planning to investigate and develop one. This challenge was prioritized as seventh most critical, again without any significant differences found across subpopulations. Some measures of effort to progress do exist in aerospace technology development guidance: the advancement degree of difficulty (referred to by NASA and Bilbro [1, 75]) and the research and development degree of difficulty (referred to by Mankins [79]). We have learned, however, that neither of these measures has become widely used, and even though NASA refers to the advancement degree of difficulty, it is not consistently used there and in fact was raised in a recent internal report as a standard to be developed. We are hopeful that improvements can be made to this type of measure that will increase adoption in industry. In the meantime, it may be useful for organizations to look into the advancement degree of difficulty or the research and development degree of difficulty as a basis for their own metric for effort through TRL progression. Figure 3-5 shows the type of graphic provided to management of a BP project with over 500 component TRLs. This bar chart proved to be the most insightful visualization we encountered in our study, differentiating not only by TRL but also by development difficulty rated on a three-level scale. Even this coarse effort-to-progress rating is informative.

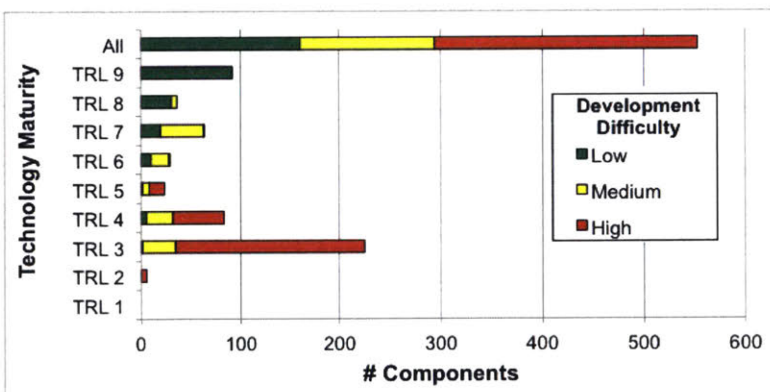

Figure 3-5: Summarized component TRL information of a complex project at BP. Components are classified by TRL and by development difficulty. (Note that the TRL distribution has been adjusted to maintain confidentiality.)

Challenge B10 - Aligning TRLs with system development gates: A major way in which organizations use the TRL scale is to include the achievement of specific TRLs as deliverables in the technology or system development process gates. These gates are key decision points that often coincide with major resource commitments. There is no general guidance available, however, to establish alignment between an organization's major development milestones and the TRLs. Should the alignment be the same for every project? Should it differ by industry, by product cycle time, or by extent of innovation? A Google product design lead explained why it is a challenge to establish this mapping: "Often times the product roadmap has a more regular cadence and a relatively short cycle. A lot of times the technology development that feeds into that, the cadence is not in synch." Ranked tenth most critical overall, we find that those who work at organizations that do not have an internal TRL standard find this challenge to be more critical (eighth overall). Those who are neither from the aerospace or defense industries rate this challenge as significantly more critical, fifth overall. Although identified as a challenge by some users, experienced TRL-using organizations have successfully mapped their system development lifecycles to the TRLs (for example see [62]). The DoD alignment of TRLs and gates is presented as an example, in Figure 3-1. Organizations looking to establish their own mapping can study this mapping for insight related to the tradeoffs between selecting immature technologies (with high technical potential at high development risk) or mature proven technologies (with limited technical potential at low development risk). However, more work should be done to better understand appropriate mapping in industries with very fast cycle times, unique risks or different technology characteristics.

Challenge B11 - Waivers: Despite careful planning, it is sometimes the case that a project arrives at a gate with one or more components at TRLs lower than what is required by their formal process - whether it be because of issues with suppliers, or unexpected test results, or new information being learned. We discovered that in these cases it is common practice to consider waiving the TRL requirement (also sometimes called dispensation). There is little guidance as to what factors should be considered when making this decision, or the nature of risk tradeoffs. This challenge ranked eleventh overall in criticality according to our survey results. No significant differences in prioritization were found amongst any subgroups of the sample. To address the challenge related to waivers, some organizations, such as BP, are simply extending their existing gate requirement dispensation process, and tailoring the reporting and information input to technology risk. The US Department of Defense does include a limited discussion on dispensation in the Technology Readiness Assessment Guidance [62]. The milestone decision authority can grant a waiver based on the project manager's plan for risk mitigation. Examples of risk mitigations are working with the customer to relax requirements and including alternative (more mature) technology.

Challenge B13 - Back-up plans: It is good practice to identify and plan for transition to back-up plans for risky critical technologies. These back-up plans are however not necessarily mature themselves. Should back-up plans and alternatives be TRL-assessed in addition to the chosen technology? There are cases when the high maturity or low criticality of the chosen technology suggests that it is unnecessary to have a viable back-up plan; there are other cases where it would be very risky not to have a back-up plan. If one is considering triggering the back-up plan, what tradeoffs should be considered? There is no such guidance in existing TRL best practice. A technology manager at NASA explained that "[having] the fallback or alternative path or plan B is a 'best practice' but not a requirement at NASA. Many projects don't develop [such] exit ramps." There is no formal link between back-up plans, their trigger points, risk, and TRLs. This challenge was found to rank thirteenth in criticality overall in the survey, across all categorizations. Regarding this challenge, we learned in our interviews that a common engineering practice is to have back-up plans in mind, however guidance regarding back-up plans and TRLs is lacking.

Challenge B14 - Product roadmapping: Technology roadmaps allow an organization to do long-term planning of product lines and innovation pipelines. Given that the TRL scale starts with observation of a physical phenomenon (TRL 1), the maturity of prospective technologies as captured by a TRL is a useful piece of information to be included in technology roadmaps. Yet current best practice in technology readiness assessment provides no guidance on integration with technology roadmaps and long-term planning. A design lead at Google suggested ways in which they see useful integration of TRL information into long-term technology planning: "Right now it's pretty haphazard, where we're like 'ok that looks great, do you think it'll be ready by the time we do the next product?'" further stating that "we're looking for ways to validate when the technology is ready for dropping into the product roadmap to help us guide our development plans." The survey respondents on average ranked this challenge second least critical overall. Respondents who frequently use TRLs in their day-to-day work find this challenge to be slightly more critical (thirteenth), as do those who work at companies which are new to TRLs (twelfth most critical). Those interested in integrating technology readiness information with technology roadmapping may find the Technology Landscaping roadmapping technique of Tierney et al. useful [80]. This technique uses a technology readiness assessment as a means of systematically addressing the maturity of current technologies, and the expectation of future technology development.

3.4.3 Assessment Validity

In this section, we present challenges related to the reliability and repeatability of the TRL assessment and scale.

Challenge C6 - Imprecision of the scale: The TRL scale simplifies the technology and system development process to 9 steps. This simplification is partly what gives the TRL such communicative power, but it also can frustrate those who would like more precision in the assessment. For example, a subsystem demonstration likely requires more than one test, and the test results themselves inform the maturity of the technology. Were multiple tests required before the technology passed? Did the technology pass marginally or easily? In particular, this lack of precision is evident at the high end of the TRL scale, where integration tests are represented. A systems engineering manager at John Deere explained that standard TRL descriptions are not practically useful: "If you're going to have an assessment and use [TRLs] to make decisions, you're going to need criteria that are not only industry specific, but even product-line or product-type specific." Yet there is little guidance on how to customize this language in a useful way. This challenge was rated on average sixth most critical overall by the survey respondents. Those who identify as not being part of the engineering function find this challenge to be much more critical (second overall). On the contrary, those who are personally new to using the TRL find it much less critical (tenth overall), as do those who work at a company that is new to the

TRLs (eleventh). The imprecision of the TRL scale relates closely to the relationship between testing specifications and TRLs. It should be possible to map out the test sequence that a technology will undergo and decide ahead of time which test results correspond to which TRLs. We see that it is common for organizations and industries to tailor their TRL definitions from the NASA baseline with more relevant language including specific internal test names.

Challenge C9 - Subjectivity of the assessment: The TRL assessment process typically requires consensus among a number of stakeholders in the project. These different stakeholders may hold different perspectives when it comes to how mature the technology is, and to what degree it has been demonstrated. Perhaps someone from R&D would like more time to spend on perfecting the technology, and so champions a lower TRL, while the project manager would like to see the technology included in the project to deliver on a performance requirement, and so argues for a higher TRL. As explained by an engineering manager from John Deere, "inevitably the person who favors the technology will interpret the TRL higher than everybody else." Some TRL users feel that the scale can sometimes seem subjective rather than objective, and that power and influence bias the ultimate assignment of TRLs. The survey analysis revealed this challenge to be ninth most critical on average. We also found that those respondents who work on the low TRLs rate this challenge as more critical (fourth overall). There exists some preliminary academic work that explores the application of modern computing advances - such as computer document classification and big data - in assigning TRLs in a more automated way [81, 82]. Practical implementations of that approach may help address assessment subjectivity concerns. Two academic works, from the aerospace and new product contexts respectively, introduce alternatives to the TRL scale based on engineering requirements, technical specs, and failure modes [83, 84]. These works reflect a criticism of the TRL scale as overly qualitative, imprecise and subjective. Further, the GAO guideline suggests strategies for documenting dissenting views, a transparent way to handle difference of opinion on an assessment team [60].

3.5 Discussion of Challenges and Prioritizations

The survey provides empirical evidence that the identified challenges are not equal in impact, and are perceived to have a range of criticalities. We found that even experienced TRL-using organizations are meeting challenges for which solutions do not yet exist. TRL users from new industries are encountering those same, as well as different challenges. We discuss the challenge- and criticality-related findings of the interviews and survey in this section.

3.5.1 Insights from overall prioritization

The four most critical TRL challenges on average fall into the system complexity category: A1- integration and connectivity, A2 - interface maturity, A3 - new components or environment, and A4 - system readiness. Further, we are not aware of significant useful current practice from industry or the literature to address these challenges. It is notable that regardless of experience, industry, function, or any other descriptive variable, these four challenges were of highest criticality. These challenges reflect that the TRL is a component-by-component measure whereas the reality of modern engineering involves integrated components that must perform in systems. TRLs were designed to assess the maturity of an individual technology and include the transition from component-level to system-level at TRL 6 [58, 9]. On the other hand, the survey also points out which challenges are relatively less critical for practitioners, regardless of context: Challenge A15, visualization of TRL data, and challenge B14, integrating TRLs with roadmaps. The challenges in the assessment validity category were ranked in the middle of the pack in terms of criticality, while the challenges in the planning and review category tended to be on average of low criticality. Qualitative comments and reactions to the survey reveal that these 15 challenges are truly relevant in industry to at least some practitioners, and TRL users would embrace solutions to these challenges. However, some respondents pushed back against the challenges, arguing that the TRL should be used for exactly what it is, a technology maturity assessment tool, and not be pushed beyond the original intended application presented by NASA.

One respondent explained:

> Many of the [challenges] stem from people wanting TRL assessments to be more than they were meant to be. In particular, people expect them to provide clarity in decision choices and risk assessments, which they can play a part in, but were not meant to accomplish.

- Technology Readiness Manager, Aerospace Industry

Although we understand this point of view, our interviews revealed that the TRL is increasingly included in project, system and technology development guidelines (for example [62]), and required by regulators ([61]) and customers, in many cases expecting the TRL combined with other knowledge to inform development decisions and risk management. We think it is appropriate to embrace this industry trend and to seek solutions addressing these challenges. Our survey was primarily circulated through a professional engineering organization; thus, 77% of our respondents identified themselves as being a part of the engineering function. Our survey may therefore under-represent the voice of non-engineering product and system developers who think the TRL scale is overly technology-centric and ignores customer, market, business, and other important factors. We do not wish to suggest the TRL should be the one and only input for project decisions. We do however think it can be an informative technology-risk related input.

3.5.2 Differences in prioritization by group

Our survey results reveal that the perceived criticality of some TRL challenges depends on the context of the TRL work and experience of the user. Figure 3-6 shows the criticality scores of the 15 challenges by respondents from organizations that have five or more years of TRL use experience ("experienced") plotted against the scores from those who have less than five years experience ("novice"). Figure 3-6 reveals that the new and experienced TRL users have some overlap, as evidenced by the four challenges highlighted in the top right corner of the figure. Despite some significant differences in rating of challenges A1, A2, and A3, those challenges remain among the top four in ranking criticality for both novice and experienced respondents.

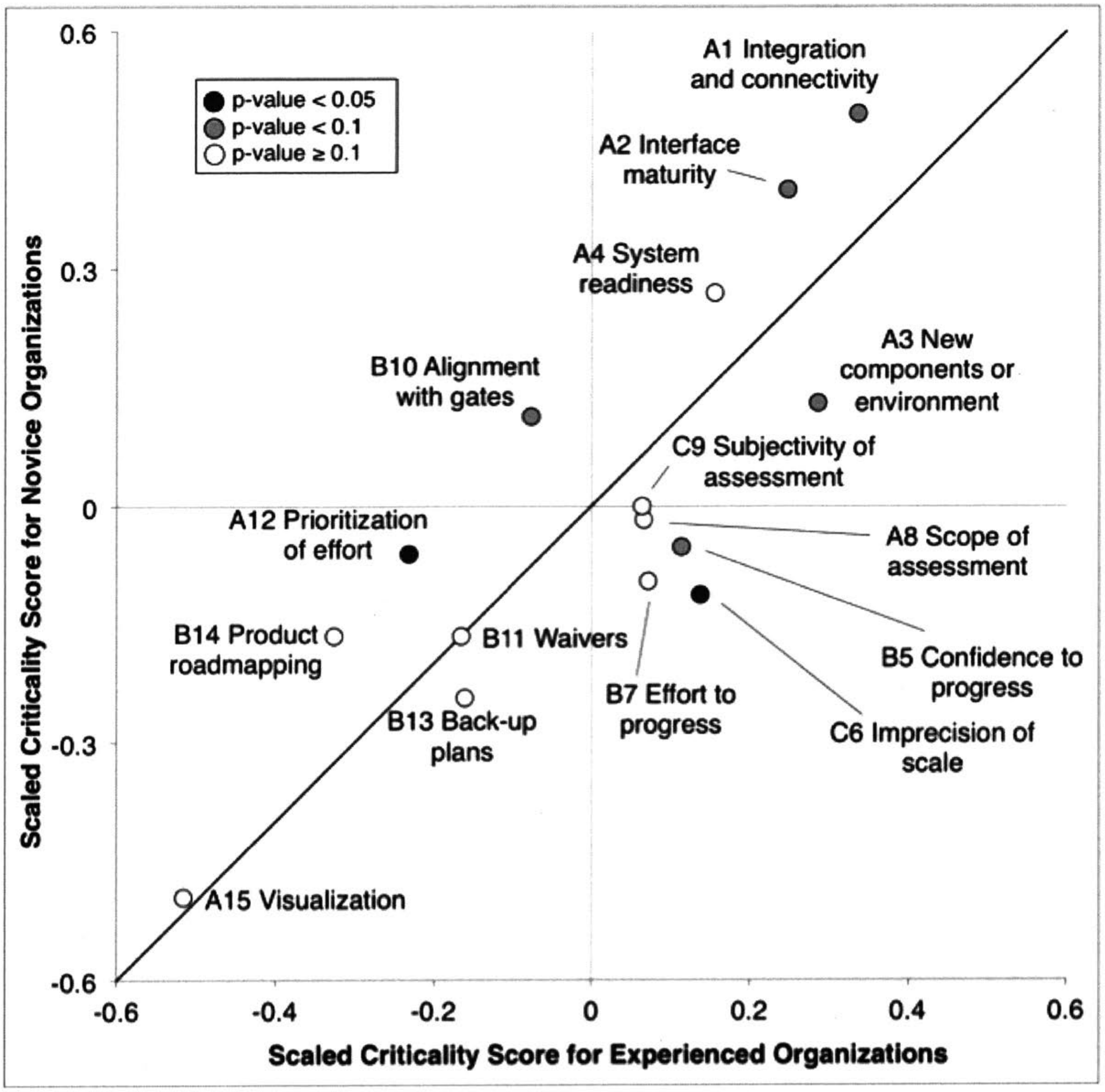

Figure 3-6: Plot of 15 challenges with their criticality score by individuals from experienced organizations (N = 23) versus those from novice organizations (N = 76).

But there also exists some differences in their perception of the criticality of TRL challenges. Much of our current understanding of TRL implementation and effectiveness results from the NASA and GAO (aerospace, defense and government) contexts. The major institutions in these industries have established standards, procedures and rules surrounding TRL usage. Our survey reveals that it is now being adopted in an even broader context, in such areas as automotive, healthcare, energy, consumer goods and electronics, as evidenced in Figure 3-7. Individuals at these novice organizations may not even know that best practices exist to address some of their challenges. On the other hand, the application of best practices from the established organizations may be impossible or unreasonable given the difference in development context; the novice organization may have a shorter development cycle or less reliance on formal and systematic decision-making. It is further possible that some challenges are highly critical initially, but once norms are established or experience is gained, become less critical. Novice organizations rate alignment (B10) and prioritization of effort (A12) as more highly critical than experienced organizations; one could argue that these challenges deserve initial attention within the organizations but can be solved with tailored, company specific-processes and standards. Challenge B10, for example, is ranked more highly critical at novice organizations (fifth most critical, versus tenth at experienced organizations). We know that the Department of Defense and NASA have existing alignments as part of their standard process. In fact, 47% of those respondents from the defense or military industries stated that TRL are included in their internal standards, 33% from aerospace, and only 21% for respondents from other industries. We do not yet know to what extent and in what cases the publically available mappings from experienced organizations are appropriate for organizations new to TRL processes, such as Google X. Moreover, respondents from TRL-experienced organizations rated confidence to progress (B5) and imprecision of the scale (C6) as more highly critical. Confidence to progress is a non-obvious extensions to the TRL method, and perhaps is only realized by the user after some experience is built with TRL practice. Similarly, it is possible that concerns regarding the imprecision of the scale may result only once a steady state and reliable TRL assessment method is established in the first place, and only then will the lack of precision be at the forefront of assessment validity concerns.

Opportunities exist to further study this new wave of TRL users, to better understand the differences between their technology risk and that of the TRL pioneering organizations, and so to better develop methods for process and standard establishment and tailoring. We imagine that the achievement of some major generalizations to TRL processes, and flexible solutions to some of the 15 challenges, would result in the

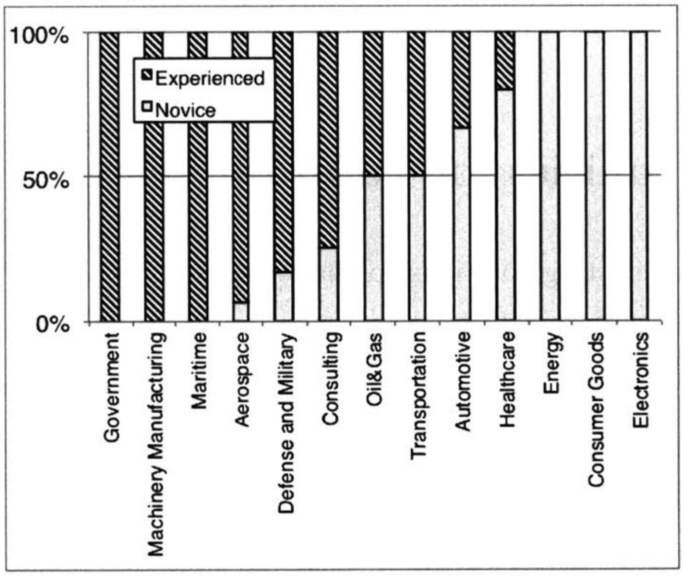

Figure 3-7: Organizational TRL experience by industry, collected in survey.

adoption of TRLs in new industries. For example, we can envision reluctance amongst those in fast-moving development industries, such as Google X, to map TRL requirements to development gates (challenge B10), since this would require investments in formal, resource intensive, and potentially burdensome assessments of technologies. In the aerospace and defense contexts, technology change happens at a slower pace, and requires significantly more stakeholder agreement and documentation. Therefore the information gained from a technology readiness assessment remains relevant and useful longer, and there exists a broad audience of stakeholders with a shared understanding of what the results mean. If a development project typically only includes one new technology, as is the case of many incremental innovation projects, we imagine the appeal of TRLs, especially in their current state of lacking integration relevance, is diminished. For example in the traditional automotive sector, TRLs at their current level of precision (challenge C6), and without major integration guidance (challenge A1, A2, A3, A4), provide little additional information than their own internal testing requirements. In summary, solutions to the 15 challenges presented here have the potential for differing impact; some would result in increased benefits to those already using the TRL, others would lead to more efficient establishment of TRL practices and norms amongst new users, and finally others may open the TRLs' relevance to whole new industries, and increase adoption.

3.6 Opportunities for Future Research

We presented some existing solutions to challenges that may be solved by the sharing and implementation of effective and creative practices across industries or from solutions found in the literature; other solutions are not obvious and present opportunities for academic research. We see multiple major directions for future research to address subsets of the challenges. The first is a study of the system implications of technology readiness. This research area should addresses challenges in the system complexity category, which were found to be ranked highly critical in the survey, yet are without effective industry or academic solutions. This research direction should take advantage of systems engineering and architectural knowledge, and build on previous work related to the system readiness levels and integration readiness levels [73]. There is a clear desire amongst practitioners for a measure of system readiness (challenge A4). However, we question the usefulness of a single number to characterize a complex system's readiness. We are doubtful that such a number would be informative to project decision making, nor would it be a comprehensive way to track a system's technology progress. We wonder why there is interest in such a simplified number. Is the way in which we currently process TRL information too complex to be helpful? Rather than a computed system readiness number, we see promise in research to expand our understanding of the system aspects of technology readiness. The integration and connectivity realities of complex systems development (challenge A1) may be effectively explored through a cascade model, building on work done by Clarkson in change prediction for product redesign and customization [85]. The work of Smaling and de Weck on technology infusion is particularly relevant to this future work, as it examines architectural invasiveness of technologies, and introduces the delta DSM architectural view of the system [50]. Also of interest is a DSM extension by Eppinger et al., which combines system architecture and integration testing information, resulting in a "multi-level" view of the test suite [86]. We imagine that such an architectural work would benefit from some elaboration on the role of interfaces on technology readiness, and on interface readiness itself (challenge A2). Another promising research direction relates to the scenario when an organization's development or acquisition process requires a minimum level of readiness of a technology in order to move to the next phase. What happens when a technology has not yet achieved the required readiness? Although these decisions are sometimes viewed as go/kill decisions, it is only sometimes the case where a project relies significantly on a critical technology and so the project would be cancelled if the technology were immature. In reality there are other options that can be negotiated in the interest of managing risk and optimizing value, including providing more time or more money, adjusting the performance requirements, or switching to an alternative technology. Both the challenges related to waivers and back-up plans (challenges B11 and B13) are possible reactions to immature technologies at a gate decision point, along with delays and cancellations. Additional information regarding confidence and effort to progress (challenges B5 and B7) would be particularly relevant. This topic could be explored via a choice model of a gate decision with the risk-weighted value of each option calculated. Such a model might be able to answer questions such as: what factors should a manager consider when deciding whether to approve a waiver? When should a back-up plan be triggered? Shishko et al. have explored real options in technology decision making should inform this line of thinking [87]. This information could potentially answer questions related to when parallel technology development of alternatives versus focused technology development is more effective, or less risky. Future work related to alignment (challenge B10) could explore whether TRLs can help qualify the hand-off between an organization's technology development and product development process. This key transition is often accompanied by a major commitment of resources. Is it appropriate to have one alignment of development gates and TRLs, as is the case in the DoD guidelines? Should the complexity or criticality of technology shift the TRL requirement higher or lower? How should the modularity of the technology affect the required TRL? We believe that these questions could be well addressed by a system model with iteration, where the likelihood of future new information discovery requiring rework, and the consequences of that rework, are captured in the model. This work could build on previous work that analyzed GAO data on technology maturity at the transition, and schedule overrun [55, 56, 57]. The tradeoff between a short time-to-market and minimized likelihood of high-impact cascading technology issues could be explored. There exists an interesting opportunity to dig into the assessment validity assessment challenges (C6 and C9), which would strengthen TRL methods overall. One approach could be to study the ad-hoc conversations, power dynamics, and negotiations that occur at TRL assessment meetings. Studies of repeatability and reproducibility of TRL assessment would provide insight into whether the perceived measurement weaknesses and biases are truly as impactful as practitioners think.

3.7 Conclusions

In this chapter we presented an in-depth investigation of the current practice of the technology readiness levels. We learned that the TRLs are used in increasingly many industries, beyond NASA and the aerospace industry where the TRLs originated and where they are extensively used. The TRLs have a significant influence on industry's technology development, investment and management decisions. We see evidence to suggest that TRL adoption will only increase, given their use imposed by industry regulations, in research and grant applications, and by risk-aware customers. We have provided a broad-based discussion of the state-of-the-art in TRL practices, and use empirical evidence to identify 15 challenges of TRL implementation. We uncovered these challenges through semi-structured interviews with 19 practitioners from seven organizations. We then broadened our base of opinion through a survey of 113 industry practitioners. We found that although some organizations are lagging behind best practices, even the most advanced TRL users face difficulty related to three categories of challenges: system complexity, planning and review, and assessment validity. We saw that in particular, the challenges related to system, interface and architecture are especially critical to current TRL users. TRL users from different contexts and with different histories, such as from different functions or of different experience levels, have differences in challenge prioritizations. We believe that awareness of these TRL implementation challenges will help technology developers and project managers avoid common pitfalls. We also shared current state of the art in best practices to address some of these challenges, with the goal of enabling process improvements. Finally, we discussed opportunities for future work motivated by our findings with the aim of inspiring other researchers to pay attention to these real and impactful challenges faced in industry today.

3.8 Appendix

We present additional illustrative quotations for each challenge in Table 3.4.

Table 3.4: Illustrative quotes of TRL shortcomings from interviews and survey.

Challenge	Illustration
A1- Integration and connectivity	"Even though we didn't necessarily have a problem with using TRL in a classic NASA way, we believe - whether it's true or not - that commercial jet aircraft and business aircraft are particularly integral products. So many things affect other things and so integration is a big deal." "We do take a view within the TRL on the level of integration. It's probably not written down somewhere in a very clear manner. We do do this. But maybe not systematically or in a way that's written down." "For example, we were looking at the more electric technology for aircraft, like electric environment control system, electric eye protection... so you need to change the electrical system but it's not necessarily new technology. But you have an impact on the engine, and on the thermal environment on your structure, even if you use a conventional structure. Basically not only assessing the maturity of the electric ECS and electric eye protection, but also how they interact with each other, with the conventional part of the aircraft that doesn't change (system level, structure, thermal environment) and this gives us an evaluation of all that needs to be done in terms of integration to be mature for the aircraft." "But a company like [ours] does not invent a lot of technology. We integrate technology from suppliers. And we sometimes integrate technology that we do ourselves, but it's usually about integrating a new technology that's been proven somewhere into a new application. Rarely are we creating a truly new technology. It happens, but it's rare. So for us, when you think about TRL, the challenges that drive our day-to-day jobs are integration"
A2- Interface maturity	"In my experience, having worked in the industry for 18 years, a lot of the big mistakes or problems I've seen have been due to poor interface management." "In defense acquisition, it is a best practice to choose mature technologies, but this inhibits cutting edge capability and often the application and interfaces aren't mature even though the technology selected may be."

A3- New components or environment	"As soon as you get in the region of TRL 5-6-7, you need to have ideas of potential application, otherwise [your demonstration] might be completely wrong. If you put [your component] in a different environment, it might not work." "[An additional challenge is] the interpretation between groups/agencies regarding what types of changes in the technology or using environment requires a change in the TRL level. Some people are more strict about the similarity to the using environment."
A4- System readiness	"In the literature they were trying to calculate [SRLs] by doing some matrix multiplication between the rating, which fell apart right away. [...] We started working on the math and it didn't work." "The challenge of a true composite TRL rating when multiple components have a lower rating, what is the overall rating, should it be lower because of the integration challenge?"
B5- Confidence to progress	"TRL is only a point in time - you need a risk weighted probability of success." "How do you identify the cost and risk associated with migrating between TRL levels?"
C6- Imprecision of scale	"We go into a TRL review, and everyone comes in and has read the definition, but has not necessarily interpreted it in the same way. One might ask for a specific test to be done, while another says 'no no, we can just do a simulation.'" "Lack of specific detailed maturity criteria for each maturity level." "A company that wants to use TRLs to assess readiness, does need to develop a set of somewhat formalized criteria. If you're going to have an assessment and use this to make decisions, you're going to need criteria that are not only industry specific, but even product-line or product-type specific. Those detailed criteria of what's ready may matter differently if you're at the tractor level vs. a transmission level vs. the gear level; at the controller level vs. the software level vs. the circuit board level. If somebody's going to use this it requires commitment to it." "At the moment what we try to do it to harmonize, as we get more specific definitions of what is required at each TRL level, and to harmonize between the people who are doing the peer reviews - to have the discussion of interpretation."

B7- Effort to progress	"The effort can differ. [...] Depending on the technology, you have to evaluate how much effort it will be to go to the next [TRL]." "The methods itself does not indicate the time to progress from one TRL to the next."
A8- Scope of assessment	"If the system architecture has changed in a manner that introduces new subsystems or new implementation methods, then the team is required to look into those areas and determine if there are new CTEs." "We generally do a top level assessment of the total system and then identify and assess critical subsystems. We usually do not have time to do a bottoms up assessment of every subsystem or component and then reach a combined assessment for the entire system." "Our approach to evaluating technology readiness starts with identifying Critical Technology Elements (CTEs). The danger in doing this, I feel, is that this tends to identify system configuration items, subsystems or even components that may have some degree of technical immaturity but may not address the obvious technical immaturity of the new system (in total) being developed." "I've never seen us do anything more than a TRL assessment (in engines) for a subsystem or a chunk of parts. Like the TRL for a system, for example - we were trying to push a power level of a specific engine in a specific application, so we were looking at TRL for the application. That devolved into certain engine tests to prove that the engine could live in that environment."
C9- Subjectivity of assessment	"[An issue] is the common agreement of the different stakeholders on the maturity level, e. g. an advanced engineering manager may call a technology production ready and a manufacturing manager may disagree." "Technologies may be well established (TRL8+) elsewhere, they are disbelieved in the local domain so become TRL2-3 !" "And now what we're facing is that if we go into a TRL review, and part of that is a peer review, and everyone comes in and has read the definition, but has not necessarily interpreted it in the same way."

B10- Alignment with gates	"[A mapping] helps people understand that they could not afford to be inventing new technology in a new product development program. [However, we were] on a new product development timeline and we were still doing technology invention." "TRL not linked to project phases reviews"
B11- Waivers	"The incentives are real to go ahead and stuff some development into something that's capitalized. We pay for it though because as you can imagine, when you do technology development on the critical path, the critical path gets long and expensive and surprising. And the risk manifest, and you wind up with big cost overruns." "Recently, a project went into the [detailed design] phase with 6 or 7 items not at the required TRL, that was all signed off and discussed with management."
A12- Prioritization of effort	"You need new choke-and-kill outlet valves, which haven't been tested subsea, so they're at TRL 3. That puts the whole blowout preventer at TRL 3. Management gets a minor stroke. But you could just put the valves in a hyperbaric chamber and move those up to TRL 5 very quickly, without much effort." "[A new] tool should give you a way forward. Given the TRL information, perhaps we can come up with rules for prioritization or strategy or action."
B13- Back-up plans	"In the case of emission compliance, [readiness risk] was the fear, because there was no step-down." "[If you only consider TRL as a number it] devalues the time and effort to pilot and mature alternative architectures or software development tools ('because that's not technology')"
B14- Product roadmapping	"Another would be taking too long to evolve the technology through the TRL cycle such that it is no longer value add by time the technology is mature...this happens ALL the time." "The developers seek to perfect a technology with respect to its application in a system, whereas the eager intended users/operators are willing to accept to accept a degraded or lower level of performance because of an urgent need to meet a serious requirement."

A15- Visualization	"I like the idea that [a matrix] is visual. But the level of integrality of the technologies in commercial jet aircraft means that this will be 50 ft by 50ft matrix. We need an interaction between tools and display which is easy to build, use and evaluate." "Our tool captures CTEs. But no data reduction or presentation."

Best-Worst Choice survey data for challenge rating and ranking is presented below in table 3.5. The Mean Count is the average of the Best-Worst Count for the 113 survey responses (could range from -5 to 5). The Mean Score is the Mean Count normalized on a -1 to 1 scale.

Table 3.5: Challenges with aggregate choice counts, ordered from most critical to least critical.

Challenge	Best Count	Worst Count	Best - Worst Count	Mean Count	Standard Deviation	Mean Score
A1 Integration and connectivity	237	30	207	1.83	1.99	0.37
A2 Interface maturity	193	27	166	1.47	1.85	0.29
A3 New components or environment	188	41	147	1.30	1.79	0.26
A4 System readiness	135	35	100	0.88	1.58	0.18
B5 Confidence to progress	120	80	40	0.35	1.83	0.07
C6 Imprecision of the scale	116	76	40	0.35	1.99	0.07
B7 Effort to progress	130	101	29	0.26	2.34	0.05
A8 Scope of assessment	97	70	27	0.24	1.56	0.05
C9 Subjectivity of assessment	123	106	17	0.15	2.30	0.03
B10 Alignment with gates	105	131	-26	-0.23	2.37	-0.05
B11 Waivers	60	151	-91	-0.81	1.97	-0.16
A12 Prioritization of efforts	62	155	-93	-0.82	1.87	-0.16
B13 Back-up plans	59	154	-95	-0.84	1.94	-0.17
B14 Product roadmapping	38	204	-166	-1.47	2.10	-0.29
A15 Visualization	32	334	-302	-2.67	2.39	-0.53

Chapter 4

Assessment of Back-up Plan, Delay, and Waiver Options at Project Gate Reviews

4.1 Introduction

Much of today's industrial product and system development follows some type of a phase-gate process. (This process is also known as phase-review, stage-gate, toll-gate, or by other terms.) Work completed during each phase is reviewed at the subsequent gate. The gates are commonly considered Go/Kill decision points. At these points, before a commitment is made to invest in the next phase of development, decision makers review the project's progress and decide whether it is worth continuing to the next phase. A careful review of the gate deliverable checklist is prescribed. The gate decision is a critical component of project control, and thus has been studied previously in some depth, as presented in Table 4.1. There exists a large amount of guidance and discussion of this Go/Kill model of the phase gate process, principally by Cooper who has written about the "stage-gate" process since the late 1980s [10, 88, 89]. Christiansen and Varnes present a set of case studies of organizations using a phased-process and learn that organizations follow the gate rules to varying degrees of formality, exhaustiveness and elaborateness, yet do not describe alternatives to Go/Kill [90]. Another body of work has explored aspects of the sometimes-irrational Go/Kill decisions made in practice. The escalation of commitment literature explores the observed phenomenon of decision makers' tendencies to continue projects (not kill), despite evidence to suggest a nonviable outcome [91, 92]. In this work we explore the gate decision that occurs when the work is not complete by the gate. The incomplete work may be, for example, a failed or incomplete test, a delayed integration task, or unavailable market data. In these cases, a Go decision is not necessarily appropriate - it could be highly risky, given the nature of the missing deliverable. At the same time, a Kill decision may be an overreaction - to cancel the project based on an addressable problem may be forgoing a large amount of value. In reality, decision makers consider more options than simply Go and Kill: they can grant a waiver and proceed to the next phase, they can delay the project's progress to the next phase, or they can switch to a back-up plan. All of these options are exercised in practice, however they are not presented in the well-known and accessible phase-gate literature. Some studies have expanded the model of the Go/Kill gate decision. The real options view of product development includes the abandonment option, analogous to Kill, continue option, analogous to Go, and adds an improve option - a "midcourse correction" described as delayed design freeze, engineering changes, or a change in the project team [93]. Krishnan and Bhattacharya model a product development effort aiming to integrate a prospective technology [52]. When the new technology is not fully validated by the required gate, the options available include committing to the new technology despite its risk (a Go decision), switching to a proven technology, or deferring commitment until later.

Van Oorschot et al. describe various interventions as means of recovering from a delayed project [94]. Several interventions heuristics are explored: do-nothing heuristic, analogous to Go; time heuristic, which involves an acceleration via increased team size; cost heuristic, which involves de-scoping of performance; and performance heuristic, where the delay is compensated for by increasing performance. These works provide useful insight on these specific options, however the complexity of their mod-

Table 4.1: Related literature and context of this work.

	Grounded in modeling, experimentation, large data sets	Grounded in industry cases; useful as a decision support model
Classic consideration of go vs. kill	Escalation of commitment [91, 92]	Cooper's stage-gate literature [10, 88, 89]; Formal rules [90]
Additional gate options considered	Real options [93]; Technology selection [52]; Interventions [94]	*This work: understanding gate options and decision heuristics with real examples*

els results in an output that is hard for practitioners to consider and integrate into their own decision-making processes. We present a more comprehensive explanation of the reality of gate decision options, with the addition of Waiver (with and without re-review), Back-up plan, and Delay, along with Go and Kill. We also show how it is feasible to extend the simple decision tree modelling approach currently used for the Go/Kill choice to analyse the expected value of the broader set of options available. Finally we demonstrate this new approach with studies from industrial application of the method. These case studies show that it is possible to estimate the parameters needed to conduct the decision tree analysis.

4.2 Realistic Options at the Gate

Informed by the previously presented literature, and discussions with practitioners, we have identified a more comprehensive set of gate decision options considered when a deliverable is incomplete at the gate. The options are shown in Figure 4-1 and elaborated upon below.

Waiver: The project can be granted a waiver for the missing gate deliverable, acknowledging that the work is not complete but nevertheless allowing the project to move into the next phase so that the investment can be approved, and the rest of the team can move on with development. Often applying for the waiver requires the generation of a plan for how the team will catch up in the next development

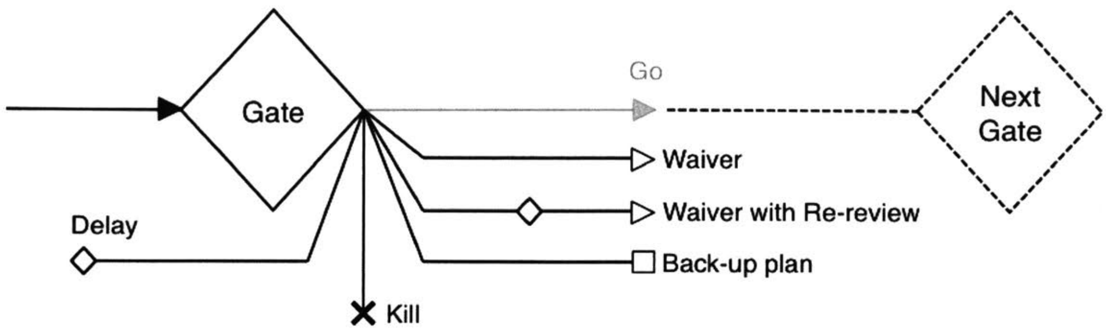

Figure 4-1: Gate options available when a deliverable is incomplete.

phase, achieving the deliverables for both phases at the next gate. A successful waiver (when the work does get caught up) avoids both a delay to the project and a performance scope sacrifice. A failed waiver may induce time- and resource-intensive rework. Within this option we also consider "passive" waivers, where the requirement is not met, yet the gate is passed without an explicit acknowledgement that there is additional risk being taken.

Waiver with re-review: A variant of the waiver is a waiver with re-review. In this case, an interim date is set for review of the incomplete work. At this re-review, action can be taken and a mid-phase adjustment can be made. This option allows the same progress of the project as the waiver, but provides an earlier opportunity for reviewing the outcome of the waiver, and making additional choices.

Back-up plan: Some projects identify a back-up plan for risky aspects of their project; for example, the back-up plan for a new technology may be a proven technology used in a previous model. Sometimes the back-up plan is a de-scoped, less desirable - and less risky - option; for example, it has less performance capability, or is more costly to develop or acquire. Other times the back-up plan is a riskier option, and is only a last resort. Back-up plans may have been considered as alternatives initially in project planning, or they may be identified only once the risk of development failure has been identified. Sometimes there is no explicit preference between Plan A and Plan B, and both options are pursued in parallel until a choice is triggered by new information.

Delay: At times the project will choose to delay entry into the next phase.

The gate decision will be to remain in the current phase until the deliverable is complete. This option allows information to be generated before the commitment to the next phase is made. It also typically includes a delay to the project timeline, or alternatively an increase in resource cost to compensate for the compression of a future phase.

Kill: When the work is incomplete at the gate, the kill option may still be appropriate. This will depend on how critical the work or deliverable is, and how much confidence the team has in being able to recover value from the project. The decision to kill may be more likely if there is no viable back-up plan option available. There may be some salvage value to killing a project - organizational learning, technology progress, or selling of capital equipment, for example. A variant of Kill is Hold, where a project is put on hold until conditions change, for example until the market prices go up or the enabling ecosystem is more fully developed. In the next section, we provide a means for decision makers to analytically consider these options.

4.3 Decision Tree Analysis

The Go/Kill model is typically accompanied with a decision-tree-style analysis, which Cooper calls Economic Commercial Value [89]. In this style of analysis, estimates are made for development costs, future earnings, and probabilities of success. Based on expected values, the decision maker can calculate whether there is greater expected future value in either going forward with the project or killing the project. We expand this decision tree modelling to include the additional gate options using the same structure and inputs, as shown below in Figure 4-2. Development costs involve engineering, tool development, rework, and capital costs. For each uncertain development activity, probabilities of success are assessed. Payoff values are assessed for both a successful outcome (S) and a failure outcome (F): they are the resulting financial impacts based on the timing, quality, cost, and revenues associated with each outcome.

This method allows the decision-maker to compute the expected values of the

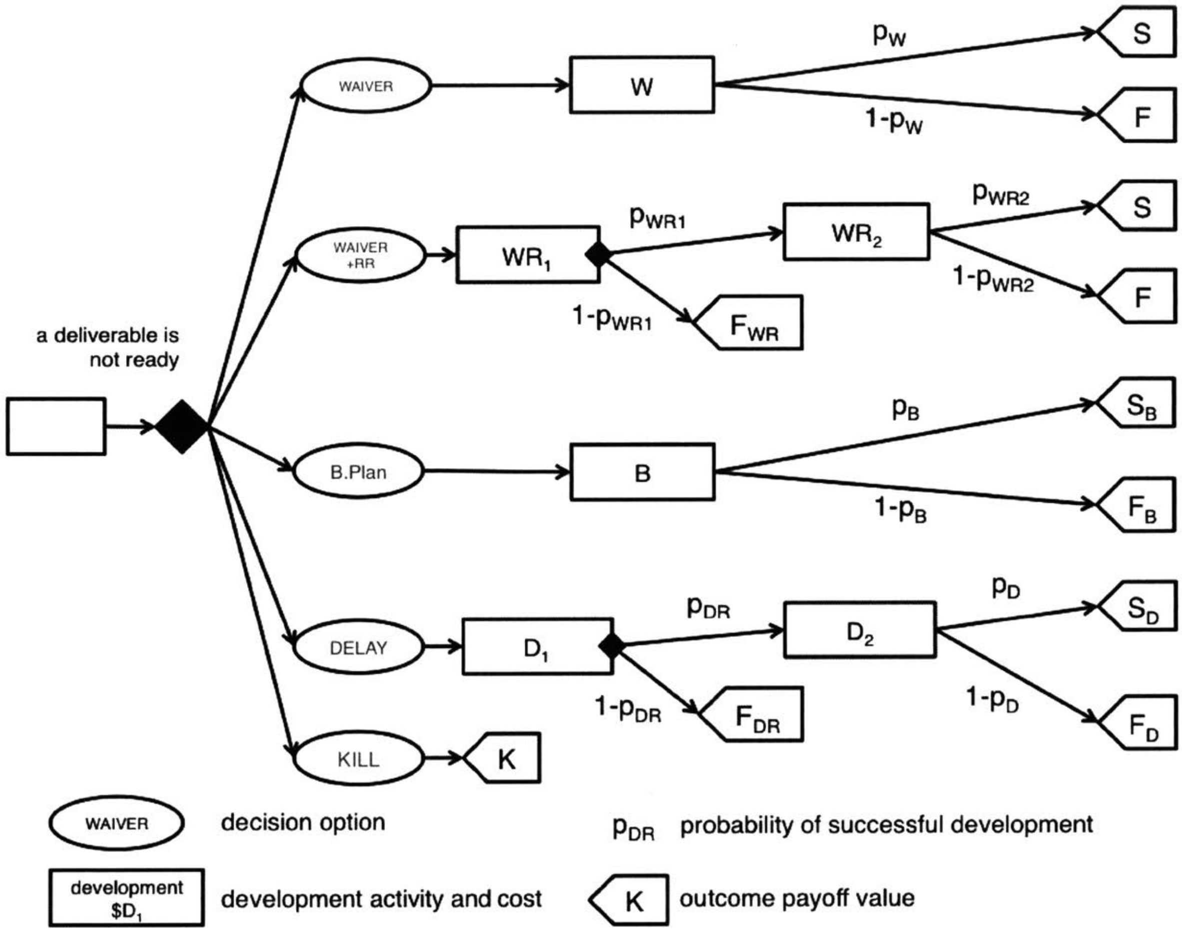

Figure 4-2: Decision tree model of expanded set of gate options.

available options and select the option with the maximum expected value. Recognizing that confidence assessments are difficult to make without bias, we envision the following graphical representation as a means of presenting information to the decision maker in the form of a broader probability space. Each option is represented as a plane in the selected probability/confidence space, as shown in Figure 4-3a as an example of a gate decision where a back-up plan, waiver and delay were considered. For a useful graphic we parametrize based on two confidences, in this case the confidence of the back-up plan and the confidence of the waiver. The optimal choice is one that maximizes value, and thus Figure 4-3b shows a two dimensional view of optimal choice for each probability combination.

We envision this analysis also to be useful as a model-based input during the decision process at the gate. Teams can assess confidence in the options, then check

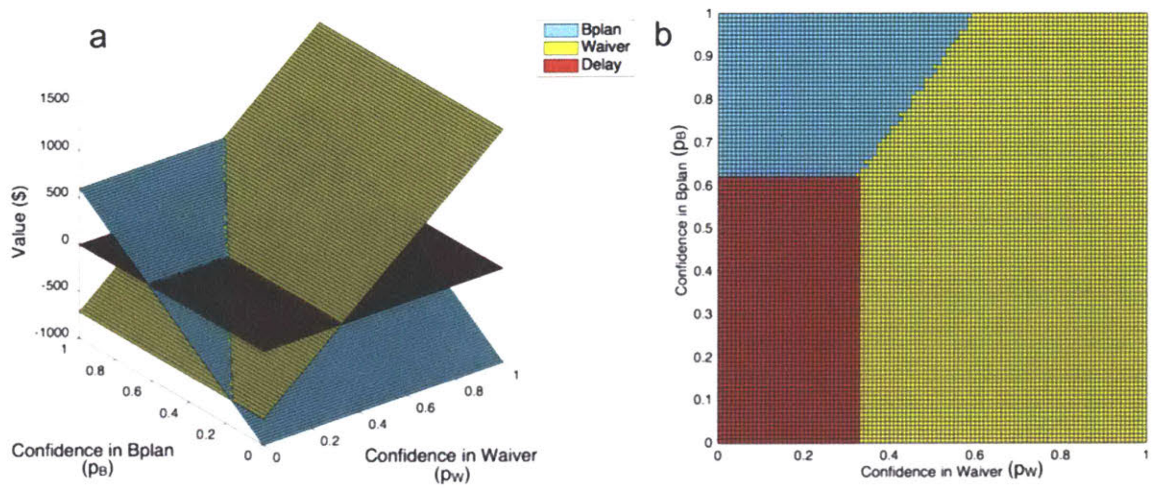

Figure 4-3: Decision space represented in terms of confidence in each option.

the model to see where the optimal choice is, based on the inputs considered. We would expect that if the confidence estimates placed the team close to the edge of any zone, a more thorough conversation would ensue to reach the decision. In order to demonstrate the use and test the limits of the model on real projects, we conducted four case studies which are presented in the next section.

4.4 Application Examples

4.4.1 Case 1: BP In-Line Inspection Tool

We worked closely with BP in Houston to apply our gate decision analysis method to two cases within one major offshore oil and gas project. Thunder Horse is BP's largest production and drilling platform in the Gulf of Mexico. Thunder Horse was designed to access a 1-billion-barrel reservoir, and achieved start of production in June 2008. Still, 58M barrels of oil are inaccessible from the original project; the Thunder Horse Expansion (THSX) project aims to drill four new wells and install additional subsea infrastructure to access this stranded oil. Regulators require oil flowlines be inspected by an in-line inspection tool (ILI tool) shown in Figure 4-4. An ILI tool must be used to conduct a baseline inspection prior to start of production. This tool uses ultrasound technology to inspect the integrity of the interior of the flowlines without interrupting the flow, and operates via a spring-loaded linkage against the pipe wall. It is propelled through the pipeline by the flow of the product. The design of the THSX ILI tool proved to be technically challenging because the original Thunder Horse subsea system has 12-inch diameter flowlines, however the expansion is designed with 10-inch flowlines. A non-standard dual-diameter ILI tool was therefore under development for this project. One requirement to be completed during the Preliminary Engineering and Definition phase is an operational environment test, i.e. technology is to be tested in the future operational environment. For the case of the ILI tool, this would involve demonstration in a dual-diameter subsea test loop. The project arrived at the scheduled exit gate of Preliminary Engineering and Definition, which precedes entry to the Detailed Engineering and Execution phase. The ILI operational environment test was not complete. The decision whether to proceed to the next phase was performed by a gate authority, a committee of BP employees of various areas of expertise, some from the project and some external to the project. Since the gate deliverable was incomplete, a Go decision was not appropriate. Accessing this stranded oil was of very high value to BP, and so killing the entire THSX project did not make sense. Instead the project considered three options:

- Apply to the gate authority for a waiver - explain how there is high confidence that by the next gate, the operational environment test will be successfully completed in addition to the next gate's requirements.

- Switch to the back-up plan - qualify two ILI tools of different diameters for the two flowline sections.

- Delay entry to Detailed Engineering and Execution until the test is complete - delay production.

The decision tree for this case is shown in Figure 4-5. The detailed values for each option are presented below.

Figure 4-4: An in-line inspection (ILI) tool. Image used with permission, source: http://www.atcopipelines.com/upr/Media/.

Waiver: The development required to pass the test and perform further development to the next gate was estimated to be \$5M. Successful development of the dual diameter tool would be considered the project baseline. If the waiver failed (i.e. development of the dual-diameter tool was incomplete by the next gate) the team would use two ILI tools, one for inspection of the 10-inch section and one for the 12-inch section. Developing and operating these two tools has an expected cost difference of \$19.5M. A condition of waiver was to include representatives from the operations team with ILI experience on other projects in the development of the tool - this effectively increases confidence in the waiver's likelihood of success.

Back-up plan: The team could decide to use two tools now at this gate. This decision would result in a \$7M development cost and \$10M operational efficiency penalty for deferred production due to increased inspection time over the field life. If the two tools are not successfully developed, the contingency is to do a more difficult "reverse inspection" which would require periodic shutdowns to do inspections. This option has double the operational efficiency penalty, equal to \$20M and would involve \$3.5M in development costs.

Delay: The team could delay passing the gate until the subsea test is complete. The estimated 3-month delay to production would have a \$25M impact. The other

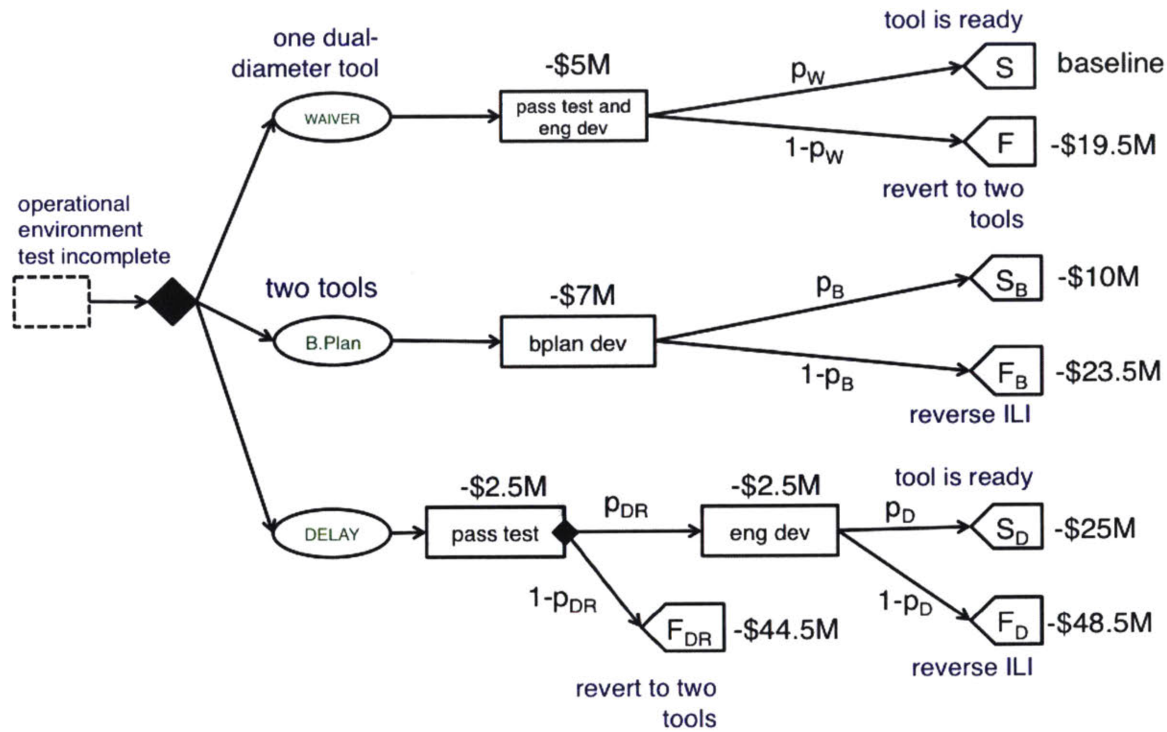

Figure 4-5: Decision tree model for in-line inspection tool case.

outcomes in the delay option are components of previously described scenarios. Discussions with the team revealed total confidence in the success of the testing that would occur as a result of the delay ($p_{DR} = 1.0$). In other words, if the whole project was delayed because of this one tool, there would be such an increase in attention and resources that there is no doubt the test would be successfully passed. Therefore we prune the F_{DR} branch of the decision tree. We asked two separate functions of the project team to assess probabilities of success. The project managers had 70% confidence in the dual-diameter tool development (p_W), and 85% in the development of two tools (p_B), versus 75% and 90% respectively for the technology specialists. These confidence estimates are shown on the output graphic of Figure 4-6. We see in the output graphics that the delay option is entirely dominated by the other two options. The optimal choice is the back-up plan only in the case of very low confidence in the waiver and high confidence in the back up plan. The actual confidence estimates of both functions place the optimal decision squarely in the waiver zone.

In this case, the team did apply for and receive a waiver to the next phase.

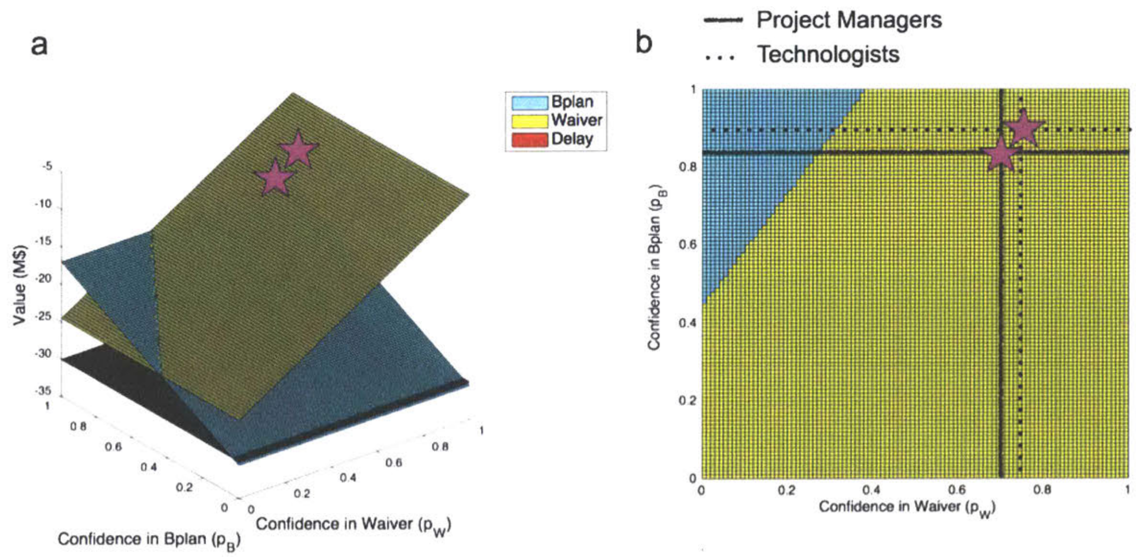

Figure 4-6: Model output for in-line inspection tool case.

Within two quarters, the operational environment test had been passed and work was proceeding on the deliverables of the next phase.

4.4.2 Case 2: BP Subsea Injection Valve Control

The 58M barrels of stranded oil will be accessed from four new wells as part of the THSX Project. These wells are threatened by asphaltene deposition, which may plug the tubing and valves in the well. To cope with asphaltene deposition, the THSX project will inject chemical inhibitors which keep the asphaltenes dissolved and avoid damage. The chemical injection metering valves (CIMVs) are controlled and monitored by an auxiliary control module (ACM) which is installed subsea, as shown in Figure 4-7. The project arrived at the scheduled exit gate of Preliminary Engineering and Definition without having completed the operational environment test on the ACM - in this case a pressure test in a hyperbaric chamber. At the gate, the team considered:

- Applying to the gate authority for a waiver - explain how there is high confidence that by the next gate the hyperbaric test will be completed in addition to the next gate's requirements.

- Switch to the back-up plan - use a subsea control module (SCM), a proven technology used in many other subsea control applications.

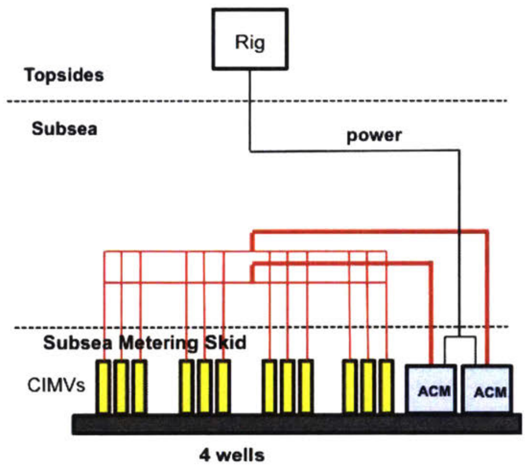

Figure 4-7: Subsea set-up of ACM for subsea injection valve control.

The decision tree for this case is shown in Figure 4-8. The values for each option are presented below.

Waiver: Developing the ACM would cost \$2.1M to demonstrate in hyperbaric testing and complete development. We benchmark the model to the successful ACM development as baseline. If the ACM fails to develop by the subsequent gate, the project will choose to switch to an old technology, a communications hub (CH), which would require topsides rework (\$10M) and would not be ready for one additional quarter, delaying production, with a \$25M impact. The reason the team could not *then* switch to the back-up plan is that the SCM development requires much more time to develop.

Back-up plan: The project could switch from the ACM to the SCM *now*. The SCM has the same performance as the ACM, and could be developed in the same timeline. It would cost \$5.8M to develop. A failed SCM development would result in minor schedule slip for rework, estimated to be negligible cost.

Again we had two separate functions of the project team assess probabilities of success. The project managers had 50% confidence in the ACM, and 100% in the development of the SCM, versus 80% and 100% respectively judged by the technology

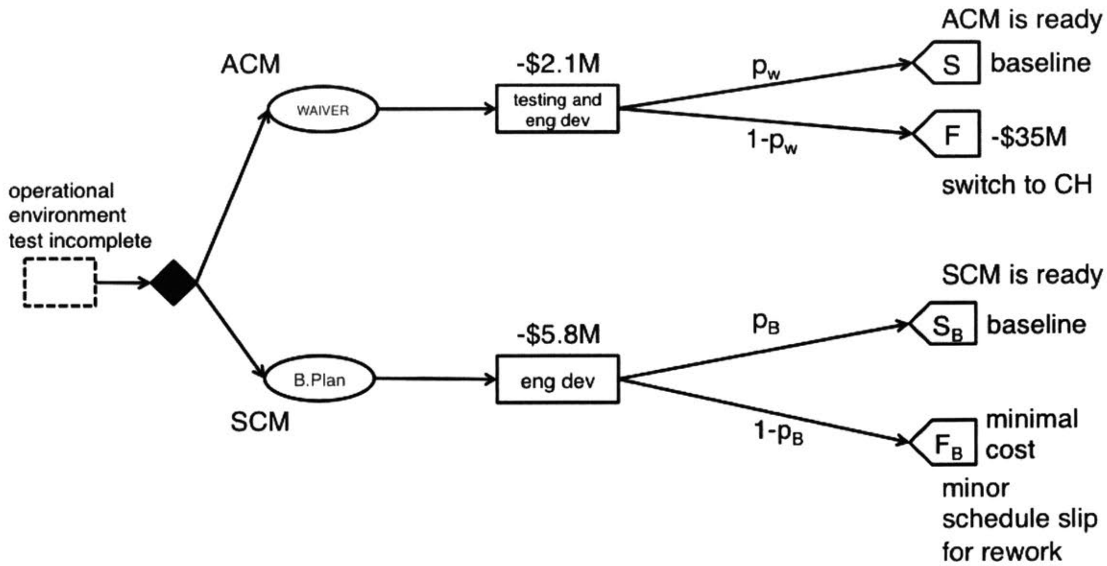

Figure 4-8: Decision tree for subsea injection valve control case.

specialists. These confidence estimates are shown on the output graphic of Figure 4-9. Discussions with the team revealed no uncertainty in the successful development of the SCM, as it was a well-proven and understood subsystem used in previous BP projects. Therefore we prune the FB branch in the analysis output in Figure 4-9. We see that the model would suggest switching to the back up plan (SCM) except if the confidence in the waiver (ACM) is greater than 90%. The technologists' confidence estimate places the optimal decision close to the edge of the zone, perhaps indicating that a more thorough investigation should be conducted. The project managers' estimate is squarely in the back-up plan zone.

In this case, the team disagreed with the model: they applied for and were granted a waiver for the ACM. In discussions with the team to understand this difference, three factors were revealed:

- A myopic scope fixation: The logic given for choosing the ACM over the SCM was their concern over difference in development cost: the ACM would cost roughly one-third the cost of the SCM to develop. The team was managing to their own current phase budget, and discounting value consequences later in the project.

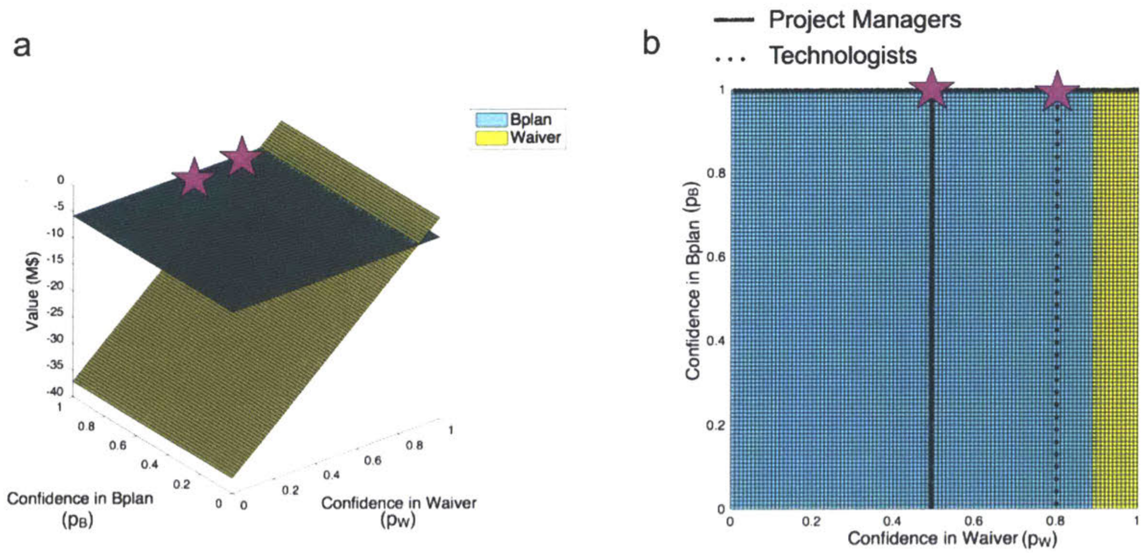

Figure 4-9: Model output for subsea injection valve control case.

- Overlooking the ACM failure outcome consequence: The back-up plan dominates the optimal decision map because the waiver option failure scenario has a very high loss outcome. When asked to make their confidence explicit, team members estimated 50% or 20% perceived chance of failure, which are not insignificant and thus appreciably lower the expected value.

4.4.3 Case Study 3: Major North American automotive program

We worked with a technical director at a major vehicle program to build the third case, capturing the decision details of a now complete major vehicle program. This program was proceeding in the execution stage when management placed a change request to include wireless charging capability in the center console. Being first to market with wireless charging would enhance revenue and lift the brand. The design and engineering team began a new design solution to incorporate the charging into the existing center console design, which required physical accommodation of the wireless charging package, relocation of the cup holders, and electrical wiring to supply power. Figure 4-10 shows one example of wireless charging technology in the center console.

Figure 4-10: Example of wireless charging technology in center console. Used with permission from GM Media Archive.

The new console design was approved, a physical prototype was developed, and part tools were developed. Then the physical prototype was tested. The center console did not pass thermal testing needed for physical prototype approval and entry to the next phase. The team did not consider killing the program, because of market expectations and supplier obligations. It also did not consider granting a waiver since thermal test results are serious safety considerations. Therefore, given this failed testing requirement, the team considered the following options at the gate:

- Switch to back-up plan - revert to original console design; lose the opportunity to be first to market with wireless charging
- Delay entry to the next phase - aim to achieve physical prototype approval with wireless charging capabilities by 90 days, resulting in a delay to Start of Production, but allowing the program to launch with wireless charging

Figure 4-11 shows the decision tree for this wireless charging case, with detailed values estimated from conversations with the technical director. The organization anticipated a first-year volume for the wireless charging model of 500,000 units, with a manufacturer's suggested retail price increase of $150 based on marketing clinic data. The material cost of the technology was $107, leaving a $43 per car profit. Each option is presented in more detail below.

Delay: achieve physical prototype approval by 90 days to integrate a cooling fan to the charging unit. A supplier would be contracted and physical and electronic design and prototyping would be completed. The additional hardware would result in a $30 increase per car in costs, further reducing the wireless charging profit to $13. This 90 day rework period would also result in a delay to Start of Production, estimated to cost $1M per day, totaling $90M. The rework during the delay may not result in a complete solution, which would require an additional $5k in tuning of the electronics.

Switch to the back-up plan: revert to original console design. In this case the team would immediately pass the gate, and have no costs over baseline in the tool development. In the case that the tool development was unsuccessful, there would only be "minor costs" ensuing from standard tweaks. This path would also lose the additional opportunity for revenue from the wireless charging (and thus the on-time, baseline positive outcome is set to the baseline of $0).

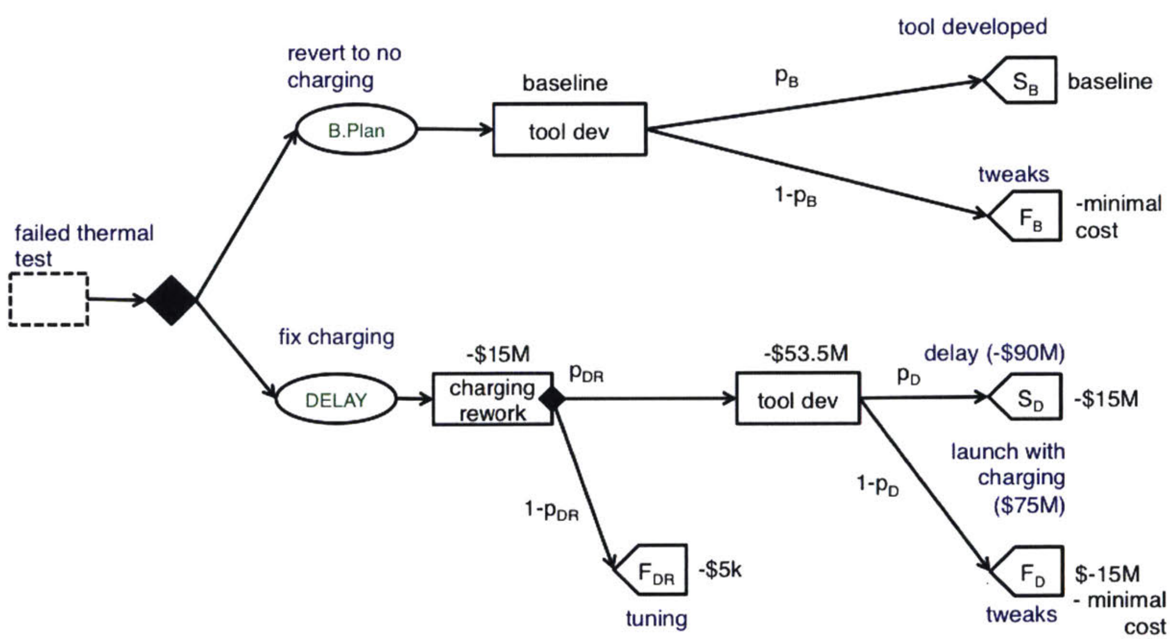

Figure 4-11: Decision tree for major automotive program wireless charging case.

We asked the technical director to assess probabilities for this example. It was estimated that there was an 80% chance the fan integration would succeed without issue (p_{DR}). With a reversion to the original console design, the project anticipated a very high 99.5% on-time success rate of the subsequent prototyping and tool development (p_B and p_D). We see from the model output in figure 4-12 (and perhaps from the decision tree already) that given these financial estimates, the delay option is entirely dominated by the back-up plan option, regardless of confidence estimate.

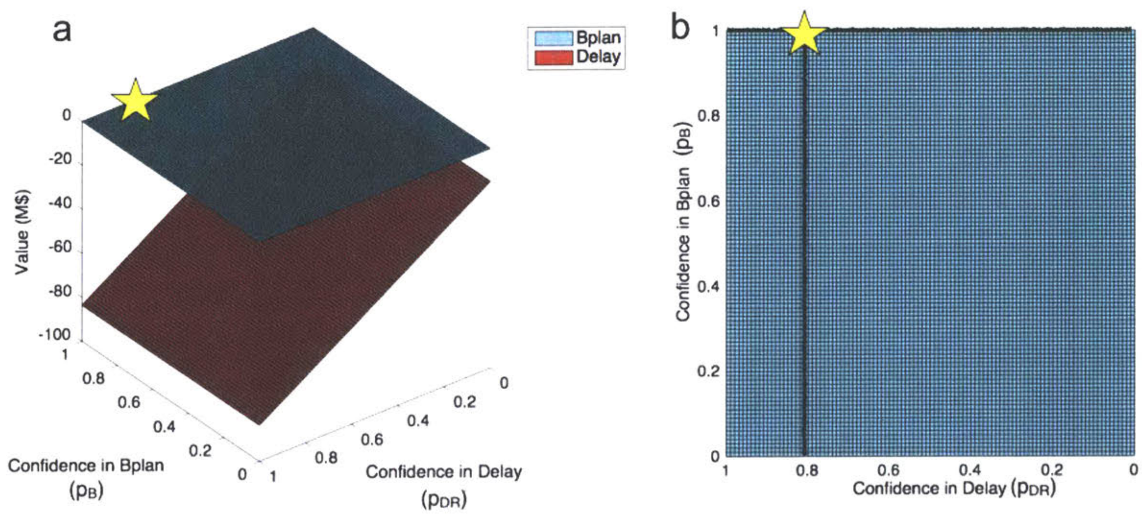

Figure 4-12: Model output for automotive wireless charging case.

Leadership chose the back-up plan, i.e. to proceed without wireless charging on this model. It should be noted that the technical director chose not to monetize brand value for the purpose of this model, and therefore we do not capture the loss of market leader status in terms of integrating wireless charging technology.

4.4.4 Case Study 4: Injectable drug delivery system

We worked with an innovation team developing an injectable drug delivery system for our next case. An example of such a system is shown in Figure 4-13. This device would be the first in a future platform of solutions. The immediate aim for this medical device development is to expand the user pool of a drug, open up a larger market, and improve competitive advantage over competitors with similar drugs. However in the case that the injectable delivery device was not ready on schedule, the organization could still launch this project's drug with a step-down delivery method. The team was working towards a Proof of Concept (POC) gate for the injectable delivery system at the end of the preliminary design phase. To pass the POC gate, a dose accuracy requirement of +/- 5% had to be demonstrated. The team discovered that with the planned design and corresponding manufacturing method, +/-5% was not achievable. Therefore awarding a waiver was not an option. The team briefly considered delaying entry to the next phase, however there was pressure to continue on the current schedule and so that option was never formally considered and the financial estimates needed for this model were not made. Instead, the team considered:

- Back-up plan 1 which could reach the dose accuracy requirement by adjusting 100% of product on the manufacturing line, but at a considerable scrap rate, and thus high unit cost.

- Back-up plan 2 which would be a compromise between the original design and Back-up plan 1.

- Kill the project, foregoing the potential revenue for this project and for future platform opportunities.

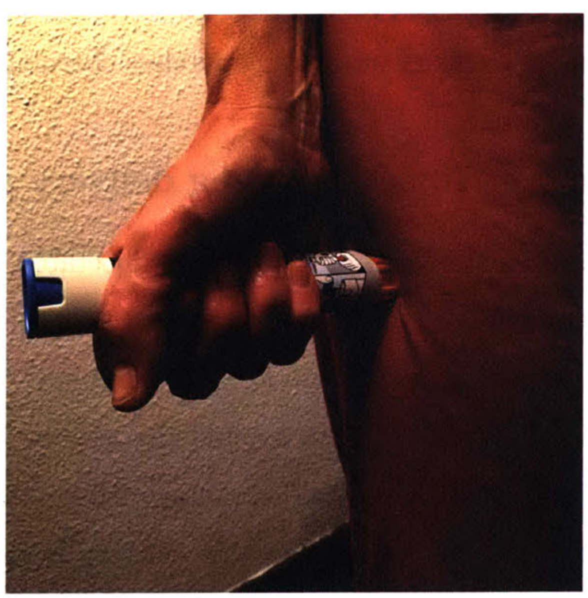

Figure 4-13: Example of an injectable drug delivery system.

The decision tree for this case is presented in Figure 4-14, and the financial estimates are explained below. Anticipated sales volume and revenue values were not readily available to our research team for the case company, however we were able to estimate these values based on expert industry reports. The potential incremental revenue was estimated to be $2B, however we acknowledge more uncertainty in this estimate than those from previous cases in this work.

Back-up plan 1: The team had previously estimated a capital cost of manufacturing equipment for this option to be $600k. The team also converted a scrap rate to an equivalent unit cost using an anticipated yearly volume, resulting in a $5 per unit cost over the baseline injectable option, and $5M total material cost based on anticipated first-year volume of 1M. If the detailed design phase failed, the drug would launch without this medical device, and thus the baseline of $0 would be the pay-off, however the device would be complete within the year and so only one year of revenue would be foregone.

Back-up plan 2: The team had previously estimated a capital cost of manufacturing equipment for this option to be $450k. Based on the same anticipated volume, the total incremental material cost for this option would be $2.5M per year. As with back-up plan 1, if the detailed design of this option failed, the outcome would be the baseline of $0, however the device would be complete within the year.

Kill: the organization considered whether to kill the project altogether, and proceed with the step-down drug delivery method in the short term, and in the long term perhaps license a device from another organization. It is uncertain whether that device would be appropriate for a platform, or how much this type of agreement would have a major impact on future revenues. It is also uncertain which drugs will make it through clinical trials to result in revenue. Given all of this uncertainty, the project manager facilitated the estimate of this value to be a present value of $4B, acknowledging a number of assumptions made.

We collected estimates from two different perspectives at the organization: project management and manufacturing. The project management team had an estimate of 75% success for Back-up plan 1, while the manufacturing team only had 25% confi-

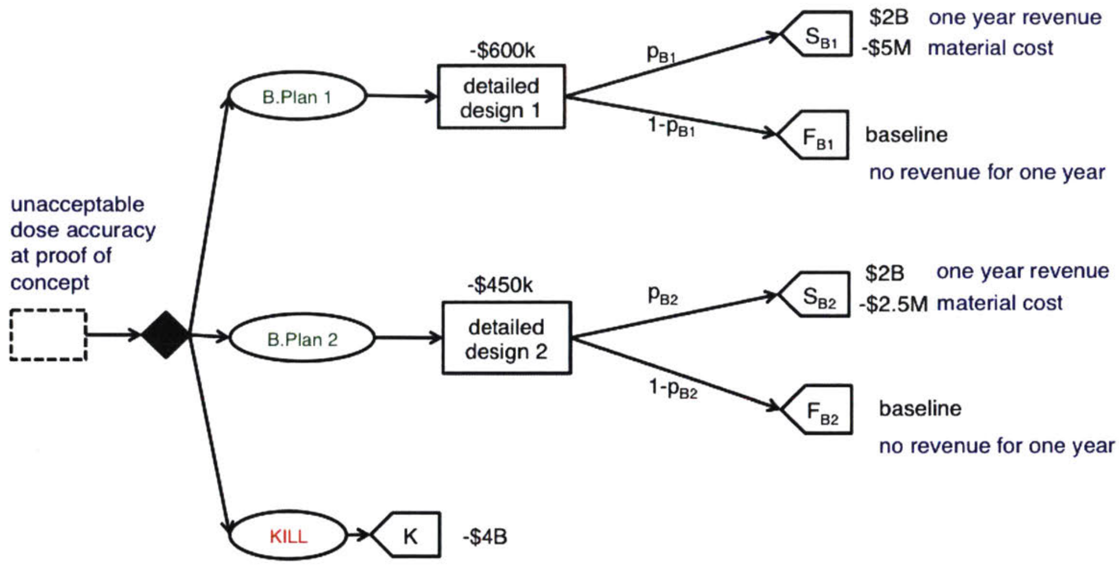

Figure 4-14: Decision tree for injectable drug device example.

dence (p_{B1}). This major difference was largely because the manufacturing manager had previously worked on a similar product, which had major development issues leading to high scrap rate and subsequent rework and schedule slip. On the other hand, the project manager had previously worked at an organization that had successfully used a similar manufacturing process, and he had confidence that this would be achievable on this product. In fact it was because of this low confidence by the manufacturing manager that the second back-up plan was even explored and considered as an alternative at the gate. Both functions had more confidence in the successful development of this option: project managers with 85% and the manufacturing team with 50% (p_{B2}). The model outputs for this case are presented below in figures 4-15a and 4-15b. We see in Figure 4-15a that as expected, the Kill option is entirely dominated by the two back-up plan options. In Figure 4-15b, we see that the boundary between the highest value option falls almost exactly as a 45 degree diagonal. This implies that the best option to choose is the one in which the team has greater confidence.

In agreement with the model, senior management decided to proceed with Back-up plan 2. The team is currently developing the manufacturing equipment and proceeding with detailed design.

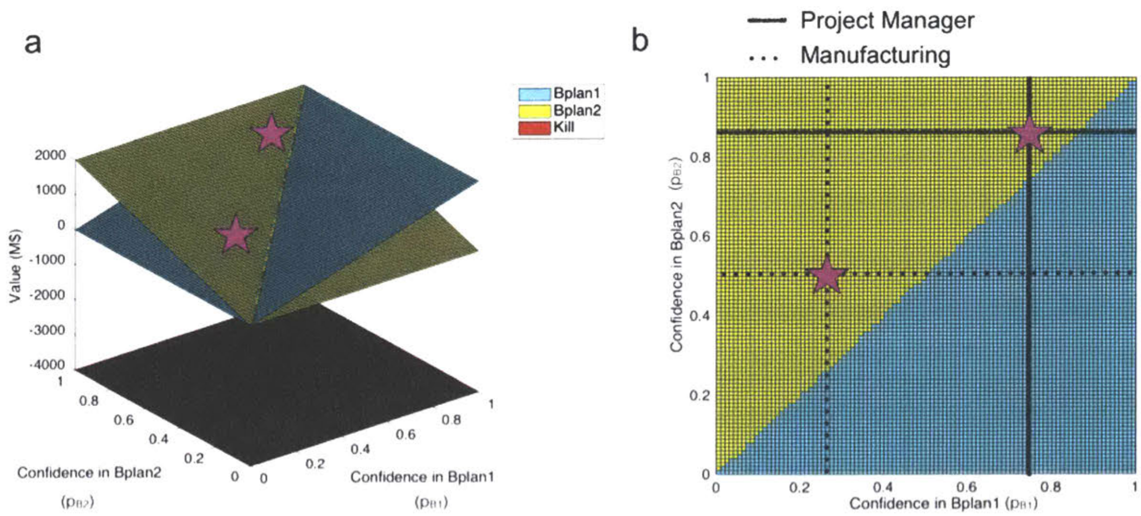

Figure 4-15: Model output for injectable drug delivery device case.

4.5 Insights from the Case Studies

These case study demonstrations revealed that analytical consideration of the realistic (expanded) set of gate options does not require complex implementation and is achievable. According to our case partners at these three organizations, useful decision support can be provided from a straightforward decision tree analysis and output graphic. We followed up with three of the four case sites to discuss the results with members of the gate authority who had not been part of the initial case generation.

4.5.1 Model value

Overall we received positive feedback. One technical expert saw value in this tool, explaining that "sometimes the cost side of it and the technical side of it is not kept together enough. I think what yours does is bring them both together. Very often you'll have a technical meeting, but cost implications ... another group would be looking at." We heard that this model has the potential to present an "objective view" of the decision, facilitating discussion and revelation of new insights between different functions and levels of management. For example, considering the second case study, and reflecting on the 50% confidence assessment made by the project management on the success of the waiver, one of the gate authority member stated:

> Why on earth did the PM say we should have a waiver then, if he really thought they had 50% shot at success, as a PM I wouldn't have done that. Not for this case. The most important thing is for the project to be delivered on time, and if you're carrying an item that has a 50% chance of failure in your plan, then that doesn't sound like a good project management decision. We never had that conversation.

He argued that facilitating the gate review, or the lead up to the gate review, with this type of model would have brought these concerns to light sooner. The same individual explained that:

> The benefit of doing this plot, of forcing people to make an assessment - whether it's right or not - is that it enables a conversation around the different perceptions or different perspectives. That's useful. All the other stuff [that we use for making the gate decision currently] like reports, decision information is on the technology itself, [is in] the details.

4.5.2 Assessment of confidences

We discovered that this analysis is based on data that are available and/or assessable by development teams, and thus could be readily implemented on projects. We learned that the financial estimates (costs and payoffs) are typically available, having already been assessed for planning and project approvals. Conversely, it is not the convention to quantitatively assess confidences of progress in the product development phases. As one technical expert summarized: "The key to it is being able to assess those probabilities. In the financial side it's not too difficult to estimate. But those probabilities are the tricky things." We were able to facilitate the confidence assessment in each of our four cases. We understand that it is not a naturally easy assessment to make; another risk expert explained that the current norm is to express probabilities in common language, not probabilities:

> When we provide our input to the gate meetings, we're expressing our level of confidence in the notes, or the brief that's prepared. We might

say - we're really concerned about the level of risk, we think there's a high chance of having a problem with delivery of the technology. There will be some words in there that express our view on the probability of success.

Still, some discussion centered on what could be done to make the assessment easier. Some case participants shared their belief that it would be easier to assess the confidence in groups; others preferred the additional information gained from first assessing individually, before sharing in a group.

4.5.3 Facilitating a discussion

We discovered that the model can represent different estimates of confidence, facilitating a discussion of heuristics, biases, and information gaps between the decision option championed by different functions in the organization at the gate. Our cases provide anecdotal evidence that there are major differences between the confidence assessments of project managers and technology experts, who both contribute to the gate decision. A decision-making process supported by our model can quantify the impact of different decision options and reveal when a difference of opinion is important to understand with more precision and when it is not. The design of the model output graphic is such that the financial estimates are static and the confidences are provided as variables on the axes - essentially providing an immediate and straightforward sensitivity analysis for the confidence estimates. This means that not only would a specific confidence assessment be useful, but one could complete this analysis with a point estimate of confidence and uncertainty interval, or even a range of confidence. As one gate authority member explained: "[The model output] is putting you straight into one decision but I can see with your chart that you get ranges, you can see how close to a range you are." There are well-documented decision-making biases that affect all of our decisions [7]. Decision makers are further biased by politics, past experience, decision myopia (related to narrow framing [95]) and performance incentives. We saw decision myopia and performance incentives play a factor in both the second oil and gas case and the injectable drug delivery case. This occurred because project staff is incentivized to manage the budget and schedule now, in the current phase, and tends to underestimate the impact on value of future phases. They therefore make decisions that optimize value in the current phase, but are risky or lower value in future phases. Perhaps this is because there is a hand-off of responsibility in a future phase, or perhaps it is simply a limitation on our ability to project ahead into time and make complex decisions. One risk expert from the gate decision team reflected on the second case decision, where the lower current capital cost, but also lower expected future value option was selected, and explained in his own words: "there's this balance of real money vs. potential money. So I make these savings today, the risk is later on, I may or may not be there. That definitely plays into that mindset."

4.5.4 Estimating financial data

It should be noted that this model considers only those decision factors that the team decides to quantify for inputs to the model. The decision's true value may be influenced by many qualitative factors, including: platform or portfolio effects, brand, competition, market uncertainty, and other difficult-to-quantify factors. We expect that this model can help facilitate the identification and discussion of these factors, even if they cannot be included in the quantitative analysis. In these four cases, we are confident that gate decision makers could have invested some additional time to assign value to hard-to-quantify factors like future platform products and brand value. We gain insight from even uncertain financial estimates, such as those from the injectable drug delivery case. Remember that we saw the optimal decision boundary between the two back-up plans as a 45 degree angled line down the middle of the design space. This means that the optimal decision should be the option in which the team has a higher confidence. The reason we see this feature is because the potential revenue payoff for either of the back-up plans is orders of magnitude larger (billions versus millions) than the capital and material costs that were also considered in this example. What this tells us is that rather than spending significant effort assessing the capital costs and scrap rates in more detail, from a value perspective, the team would be better served spending effort towards improving the team's confidence in a successful detailed design, and on-time delivery of the device. We could extend the model with the incorporation of sensitivity analysis of the financial estimates to the model output, thus allowing the decision makers to understand which factors have the most influence on the value maximizing output. A change to the financial estimates would look like a shift of the surface in the model output, and thus a shift of the boundaries between the optimal decision zones. As an example of this type of analysis, we returned to the automotive wireless charging example, and in particular the 90 day estimated length of a delay to rework the centre console design. With our model it is possible to identify the length of delay less than which it would be an optimal decision to delay, rather than revert to the back-up plan, given the 80% assessed confidence in the delay option. See the model output in Figure 4-16 below.

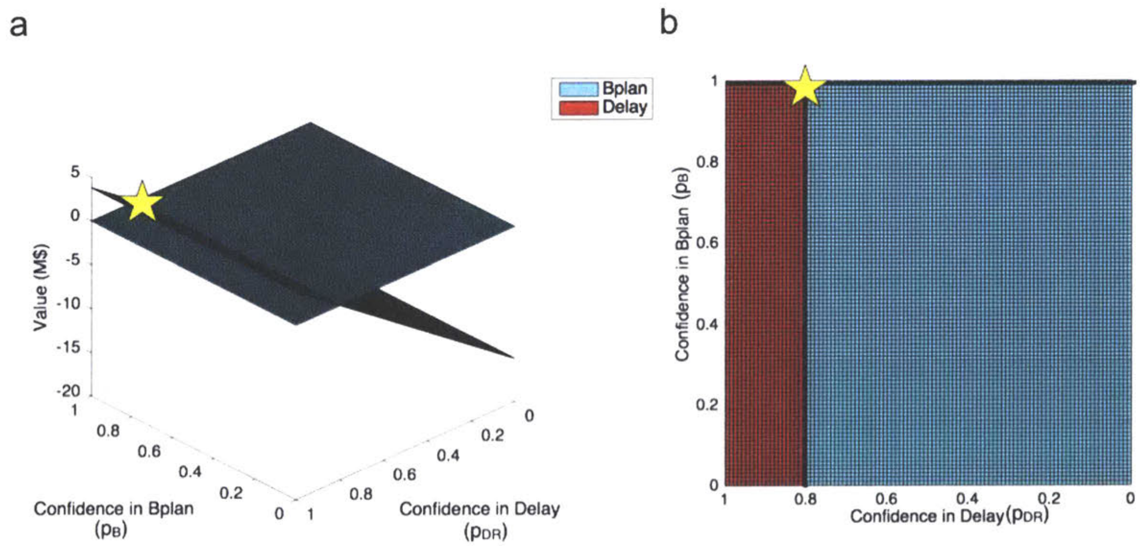

Figure 4-16: Model output graphic for the major automotive program, with an adjusted delay length from 90 to 2.7 days.

We found that any delay length less 2.7 days (65 hours), at one million dollars a day, would be value optimal over reverting to the back-up plan. In Figure 4-16a we see the delay plane has moved up towards the back-up plan plane to the point where they now intersect. In Figure 4-16b, we see that the intersection lies right at the 80% confidence estimate of delay success. We see that sensitivity analysis of the financial information and scenario exploration is facilitated with this model.

4.5.5 Research Implications

There are several extensions envisioned for this work. For example, some of these decision options are recursive, and the outcome of one choice will open up another set of choices. This is currently reflected in the model using expected value, but it may be more powerful to add the complexity of multi-phase decisions and contingency choices. This analysis will prove difficult to graphically illustrate for a decision support tool but still simple to compare expected values for each decision, and thus reveal interesting insights. The analysis presented here is risk neutral, however a simple extension to this model could explore and represent risk-averse or risk-seeking attitudes. This would allow for a study of whether some gate options are systematically better suited to different risk attitudes; for example, we would expect that a risk-seeking firm would be more likely to choose the waiver option. None of the cases observed considered the option of waiver with re-review. We chose to continue to include this option in our set of gate choices, given that we have collected anecdotal evidence to suggest the waiver and re-review is implemented by thoughtful decision makers. This option provides the benefit of the momentum into the next phase, with increased investment and work proceeding for the rest of the project, while accounting for the incomplete deliverable early enough to allow another project decision to recover value. Future work could expand this model to consider the multi-project context of large organizations, or the multi-deliverable nature of gate reviews. We currently view each deliverable in isolation but in reality, a delay caused by one incomplete deliverable may allow another incomplete deliverable time to catch up, too. On the other hand, the delay may adversely affect the development of another subsystem, due to expectations of suppliers or contractors.

4.6 Conclusions

We develop a realistic representation of gate decision options through the incorporation of options such as Waiver (with and without re-review), Back-up plan, and Delay - along with the standard Go and Kill options. This is accomplished through an expansion of the simple decision tree modelling approach currently used for the Go/Kill choice to analyse the expected value of the broader set of options available. Finally we demonstrate this new approach with application case studies from industry and reveal the insights gained from these examples. This rational and structured gate-decision model also provides an opportunity to explore the rational versus intuitive decision making that occurs in these critical gate decisions. Future work could address the decision maker myopia and other qualitative factors evidenced in this study.

Chapter 5

Conclusion

We have explored three different structured decision-making and assessment tools in this thesis. We hope that for you, like for us, these studies have inspired a number of ideas for future work. We conclude the thesis with a discussion of our ideas of future research, beyond those detailed at the end of each chapter. Some are further extensions to the tools, while others are studies involving the phenomena studied by the tools.

5.1 Testing archetypes

The multilevel DSM presented in the first essay could provide a repeatable means of comparing test suites across different complex systems. We can imagine a study that would compare the multilevel testing DSMs for radically new vs. incrementally innovative systems. Further, how does testing differ between systems built under the "design and own" model when compared to those developed through a "specify and lease" model? If we are looking to efficiently test a system, must we always perform the full set of tests? Can we combine tests or re-order to gain more information sooner?

5.2 Readiness-level mania

In our time spent studying the TRL, we could not help but notice how many [insert word] readiness level variants (XRLs) exist, at least in scale form. We have described the technology RL, integration RL and system RL in detail in this work. The manufacturing RL is relatively widely discussed [96]. How about the knowledge RL (KRL), human integration RL (HIRL), demand RL (DRL), or the market RL (MRL), just to name a few? We will leave it to you to find references to these online. Why are we so drawn to this simple 7- or 9-step decomposition of maturity? Perhaps these scales are simply appealing because they are structured methods that put repeatability and a common language around complex development concepts. We see problems emerge in the way these scales are used. It is tempting to average, add or otherwise mathematically manipulate these scales to achieve composite readiness measures. These operations are mathematically nonsensical; the scales are rarely designed to map 1:1, they are non-linear, and fundamentally a low RL in one dimension should not cancel with a high RL in another dimension. There is room for future work related to better understanding how these various RLs can be used together effectively.

5.3 Learn more about TRLs and project success

The most detailed information that we have about TRLs and project outcomes is from the Government Accountability Office and their yearly NASA and defense acquisition assessment reports and best practice papers [97, 98, 99]. We are interested in collecting a large set of TRL data from industries outside of defense or aerospace. In particular, the TRL of key technologies at the hand-off between technology development and product development strikes us as particularly interesting as a question of study. Further, this gate in particular gives an interesting context for further exploration of the selection and success of the expanded set of gate options. How often is a low TRL given a waiver at the entry to product development gate? And how often is that waiver successful? How often is the start of the project delayed so that the TRL can be raised? How often does the project continue successfully after such a delay?

5.4 Flawed decision making

The tools described in the essays of this thesis are each intended to augment and structure the information available to an engineering decision maker. Nevertheless, our studies have revealed that biases and heuristics exist, confirming what is known about human cognitive abilities. There is much to be explored on this topic. The TRL assessment, or gate review decision, or assessment of confidences all provide interesting contexts to examine differences in the way project members process and make sense of complex information. Do technical experts' and project managers' opinions differ in a systematic way? Can we plan incentives to combat the decision myopia that we see in gate review decision making? What are useful techniques to elicit reasonable predictions of confidence? We understand that humans are better at assessing the change in probability due to some factor change, rather than the probability itself. Could this advantage be incorporated into a confidence assessment technique? There is a great deal still to be understood and exploited regarding bias in decision making in the context of complex systems engineering.

5.5 Back up plans

We were surprised in our study of gate review options to see a dearth of discussion in the product development literature regarding the back-up plan. We consistently heard from smart engineers that having a Plan B was a basic part of good engineering, and whether that plan was made explicit or not, they typically had an alternative planned as a mitigation. We wonder why the back-up plan is not a more explicit part of planning. In fact, we need a clearer set of vocabulary to describe the back-up plan. Sometimes a back-up plan started out as a less-ideal alternative; sometimes a back-up plan is conceived of part-way through the project when a problem is encountered;

sometimes a back-up plan exists, but is not practically implementable because the same lead work was not done with suppliers, drawings, and integration that has been done with the plan A. These cases each involve different risk characteristics and thus would imply different decision-making, yet we label them all as back-up plans. On the other hand, we learned anecdotally from Doug Field, Senior VP Engineering at Tesla Motors that back-up plans are counter productive in high innovation settings like Tesla: "if you have a back-up plan, you'll resort to the back-up plan." What is the right balance to strike between being prepared for problems with a back-up plan mitigation, and encouraging resilient work to figure out a new solution? And does the context of the organization influence this balance? For example, Tesla has been known to let announced product launch dates slip, without great consequence to their brand or sales expectation. Could Ford do the same?

We hope that the ideas presented in this section provide fruitful future work, generating insights useful to future complex engineering systems developers.

Bibliography

[1] NASA. NASA Systems Engineering Handbook. Technical report, Washington, D.C., 2007.

[2] Office of the Chief Engineer NASA. NASA 7123.1B Systems Engineering Processes and Requirements. Technical report, 2013.

[3] Aaron Smith. How many jobs did the Mars landing create?, 2012.

[4] Allison Linn. Hundreds of suppliers, one Boeing 737 airplane, 2010.

[5] Chris Welch. The iPhone's camera is so good because 800 people are working on it, 2015.

[6] Herbert A Simon. *The Sciences of the Artificial.* MIT press, Cambridge, MA, 3 edition, 1996.

[7] David Arnott. Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, 16(1):55–78, 2006.

[8] Steven D. Eppinger and Tyson R. Browning. *Design Structure Matrix Methods and Applications.* MIT Press, Cambridge, MA, 2012.

[9] John C. Mankins. Technology Readiness Levels: A White Paper, 1995.

[10] R G Cooper. Stage-Gate Systems: A New Tool for Managing New Products. *Business Horizons*, 33(3):44–54, 1990.

[11] B.S. Blanchard and W.J. Fabrycky. *Systems Engineering and Analysis.* Prentice Hall, Englewood Cliffs, NJ, 1990.

[12] J.I.M. Halman and B.F.M. Braks. Project alliancing in the offshore industry. *International Journal of Project Management*, 17(2):71–76, 1999.

[13] Monica Giffin, Olivier L. De Weck, Gergana Bounova, Rene Keller, Claudia Eckert, and P. John Clarkson. Change Propagation Analysis in Complex Technical Systems. *Journal of Mechanical Design*, 131(8):081001, 2009.

[14] Yoram Reich. Learning in design: From characterizing dimensions to working systems. *Ai Edam*, 12(02):161–172, 1998.

[15] Janet K. Allen, Carolyn Seepersad, HaeJin Choi, and Farrokh Mistree. Robust Design for Multiscale and Multidisciplinary Applications. *Journal of Mechanical Design*, 128(4):832, 2006.

[16] X.F. Zha, R.D. Sriram, M.G. Fernandez, and F. Mistree. Knowlege-intensive collaborative decision support for design processes: a hybrid decision support model and agent. *Computers in Industry*, 59(9):905–922, 2008.

[17] Kemper Lewis. Making Sense of Elegant Complexity in Design. *Journal of Mechanical Design*, 134(12):120801–120801, 2012.

[18] Nikolaos Papakonstantinou, Seppo Sierla, David C. Jensen, and Irem Y. Tumer. Simulation of Interactions and Emergent Failure Behavior During Complex System Design. *Journal of Computing and Information Science in Engineering*, 12(3):031007, 2012.

[19] T.R. Browning. Applying the design structure matrix to system decomposition and integration problems: a review and new directions. *IEEE Transactions on Engineering Management*, 48(3):292–306, 2001.

[20] Mike Danilovic and Tyson R. Browning. Managing complex product development projects with design structure matrices and domain mapping matrices. *International Journal of Project Management*, 25(3):300–314, apr 2007.

[21] Michael D Griffin. How Do We Fix System Engineering. In *61st International Astronautical Congress*, pages 1–9, Prague, Czech Republic, 2010.

[22] N. Leveson, N. Dulac, K. Marais, and J. Carroll. Moving Beyond Normal Accidents and High Reliability Organizations: A Systems Approach to Safety in Complex Systems. *Organization Studies*, 30(2-3):227–249, 2009.

[23] DH Stamatis. *Failure mode and effect analysis: FMEA from theory to execution.* American Society for Quality, Milwaukee, 2nd edition, 2003.

[24] International Electrotechnical Commission. IEC 60812: Analysis techniques for system reliability - Procedures for failure mode and effect analysis (FMEA), 2006.

[25] International Electrotechnical Commission. IEC 61882: Hazard and operability studies (HAZOP studies) - Application guide, 2001.

[26] International Electrotechnical Commission. IEC 61025: Fault tree analysis (FTA), 2006.

[27] Nancy G. Leveson. *Engineering a Safer World.* The MIT Press, Cambridge, MA, 2011.

[28] Robert B. Stone, Irem Y. Tumer, and Michael Van Wie. The Function-Failure Design Method. *Journal of Mechanical Design*, 127(3):397, 2005.

[29] D. Braha and O.Z. Maimon. *A Mathematical Theory of Design: Foundations, Algorithms and Applications.* Kluwer Academic, Boston, 1998.

[30] Albert Albers, Andreas Braun, Eike Sadowski, David C. Wynn, David F. Wyatt, and P. John Clarkson. System Architecture Modeling in a Software Tool Based on the Contact and Channel Approach (C&C-A). *Journal of Mechanical Design*, 133(10):101006, 2011.

[31] Andrew Harold Tilstra, Carolyn Conner Seepersad, and Kristin L. Wood. A high-definition design structure matrix (HDDSM) for the quantitative assessment of product architecture. *Journal of Engineering Design*, 23(10-11):767–789, nov 2012.

[32] Thomas U Pimmler and Steven D Eppinger. Integration analysis of product decompositions. In *ASME Design Theory and Methodology Conference*, number September, pages 1–10, Minneapolis, MN, 1994.

[33] A Yassine. An introduction to modeling and analyzing complex product development processes using the design structure matrix (DSM) method. *Urbana*, pages 1–17, 2004.

[34] W.R. Ashby. *An Introduction to Cybernetics.* Chapman & Hall, New York, 1956.

[35] S. Beer. *Designing Freedom.* Wiley, New York, 1975.

[36] Y. Bar-Yam. When systems engineering fails-toward complex systems engineering. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance (Cat. No.03CH37483)*, volume 2, pages 2021–2028. Ieee, 2003.

[37] Eero P Simoncelli. Shiftable Multiscale Transform. *IEEE Transactions on*, 38(2):587–607, 1991.

[38] Y. Bar-Yam. Engineering complex systems: multiscale analysis of evolutionary engineering. In *Complex engineered systems*, pages 22–39. Springer, Berlin, 2006.

[39] E. Weinan, B. Engquist, X. Li, W. Ren, and E. Vanden-Eijnden. Heterogeneous multiscale methods: a review. *Communications in Computational Physics*, 2(3):367–450, 2007.

[40] C. Hartmann, W. Zhang, J. Latorre, and G. Pavliotis. Optimal control of multiscale systems: an approach using logarithmic tranformations. In *Proceedings of the international conference on scientific computation and differential equations, SciCADE*, Valladolid, Spain, 2013.

[41] D.E. Whitney, Qi Dong, J. Judson, and G. Mascoli. Introducing knowledge-based engineering into an interconnected product development process. In *Proceedings of the 1999 ASME Design Engineering Technical Conference*, 1999.

[42] David S Rosenblum and Elaine J Weyuker. Lessons learned from a regression testing case study. *Empirical Software Engineering*, 2(2):188–191, 1997.

[43] INCOSE. *Systems engineering handbook*. Number October. INCOSE, San Diego, 3.2.2 edition, 2011.

[44] N. F. Michelena and P. Y. Papalambros. A Network Reliability Approach to Optimal Decomposition of Design Problems. *Journal of Mechanical Design*, 117(September 1995):433, 1995.

[45] EG Anderson and Nitindra R. Joglekar. *The Innovation Butterfly: Managing Emergent Opportunities and Risks During Distributed Innovation*. Springer, New York, 2012.

[46] A. J. Groen and S. T. Walsh. Introduction to the Field of Emerging Technology Management. *Creativity and Innovation Management*, 22(1):1–5, 2013.

[47] Mike Hobday and Howard Rush. Technology management in complex product systems (CoPS) - ten questions answered. *International Journal of Technology Management*, 17(6):618, 1999.

[48] Michael Song and Mitzi M. Montoya-Weiss. The Effect of Perceived Technological Uncertainty on Japanese New Product Development. *The Academy of Management Journal*, 44(1):61, 2001.

[49] Armin P. Schulz, Don P. Clausing, Ernst Fricke, and Herbert Negele. Development and integration of winning technologies as key to competitive advantage. *Systems Engineering*, 3(4):180–211, 2000.

[50] Rudolf Smaling and Olivier De Weck. Assessing Risks and Opportunities of Technology Infusion in System Design. *Systems Engineering*, 10(1):1–25, 2006.

[51] Marco Iansiti. Technology integration: Managing technological evolution in a complex environment. *Research Policy*, 24(4):521–542, 1995.

[52] V. Krishnan and Shantanu Bhattacharya. Technology Selection and Commitment in New Product Development: The Role of Uncertainty and Design Flexibility. *Management Science*, 48(3):313–327, 2002.

[53] Nuno Gil, Marcela Miozzo, and Silvia Massini. The innovation potential of new infrastructure development: An empirical study of Heathrow airport's T5 project. *Research Policy*, 41(2):452–466, 2012.

[54] Steven R. Meier. Best Project Management and Systems Engineering Practices in the Preacquisition Phase for Federal Intelligence and Defense Agencies. *Project Management Journal*, 39(1):59–71, 2008.

[55] United States Government Accountability Office. Assessments of Selected Weapon Programs GAO-07-406SP. Technical Report March, 2007.

[56] Gregory F. Dubos, Joseph H. Saleh, and Robert Braun. Technology Readiness Level, Schedule Risk, and Slippage in Spacecraft Design. *Journal of Spacecraft and Rockets*, 45(4):836–842, 2008.

[57] Daniel R Katz, Shahram Sarkani, Thomas Mazzuchi, and Edmund H Conrow. The Relationship of Technology and Design Maturity to DoD Weapon System Cost Change and Schedule Change During Engineering and Manufacturing Development. *Systems Engineering*, 18(1):1–15, 2015.

[58] Stanley Sadin, Frederick Povinelli, and Robert Rosen. The NASA Technology Push Towards Future Space Mission Systems. *Acta Astronautica*, 20:73–77, 1989.

[59] James Moultrie. Understanding and classifying the role of design demonstrators in scientific exploration. *Technovation*, 43-44:1–16, 2015.

[60] GAO U.S. Accountability Office. Technology Readiness Assessment Guide: Best Practices for Evaluating the Readiness of Technology for Use in Acquisition Programs and Projects. Technical report, 2016.

[61] Det Norske Veritas. Forensic Examination of Deepwater Horizon Blowout Preventer. Technical Report March, 2011.

[62] US Department of Defense. Technology Readiness Assessment (TRA) Guidance, 2011.

[63] Homeland Security Institute. Department of Homeland Security Science and Technology Readiness Level Calculator (ver 1.1) Final Report and User's Manual. Technical report, 2009.

[64] Margaret A Frerking and Patricia M Beauchamp. JPL Technology Readiness Assessment Guideline. In *2016 IEEE Aerospace Conference*, 2016.

[65] John C. Mankins. Technology readiness assessments: A retrospective. *Acta Astronautica*, 65(9-10):1216–1223, nov 2009.

[66] Jordan J. Louviere, Terry N. Flynn, and A.A.J. Marley. *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge University Press, Cambridge, UK, 1 edition, 2015.

[67] Alistair Brandon-Jones, Michael Lewis, Rohit Verma, and Matthew C Walsman. Examining the characteristics and managerial challenges of professional services: An empirical study of management consultancy in the travel, tourism, and hospitality sector. *Journal of Operations Management*, 42-43:9–24, 2016.

[68] L. L. Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927.

[69] Katharina Tomaschek, Alison L. Olechowski, Steven D. Eppinger, and Nitin R. Joglekar. A Survey of Technology Readiness Level Users. In *INCOSE International Symposium (IS 2016)*, Edinburgh, UK, 2016.

[70] A.A.J. Marley and J.J. Louviere. Some probabilistic models of best, worst and best-worst choices. *Journal of Mathematical Psychology*, 49:464–480, 2005.

[71] Katharina Tomaschek. *An Empirical Investigation of the Challenges of Technology Readiness Levels Application.* PhD thesis, Swiss Federal Institute of Technology (ETH), 2015.

[72] Hernando Jimenez and Dimitri N. Mavris. Characterization of Technology Integration Based on Technology Readiness Levels. *Journal of Aircraft*, 51(1):291–302, jan 2014.

[73] Brian J. Sauser, Ryan Gove, Eric Forbes, and Jose E Ramirez-Marquez. Integration maturity metrics: Development of an integration readiness level. *Information Knowledge Systems Management*, 9:17–46, 2010.

[74] Jose Emmanuel Ramirez-Marquez and Brian J. Sauser. System Development Planning via System Maturity Optimization. *IEEE Transactions on Engineering Management*, 56(3):533–548, 2009.

[75] James W Bilbro. Systematic Assessment of the Program / Project Impacts of Technological Advancement and Insertion Revision A. 2007.

[76] Edouard Kujawski. Analysis and Critique of the System Readiness Level. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 43(4):979–987, jul 2013.

[77] Timothy K Brady. *Utilization of Dependency Structure Matrix Analysis to Assess Implementation of NASA's Complex Technical Projects.* Master of science in engineering and management thesis, MIT, 2002.

[78] Sirous Yasseri. Subsea system readiness level assessment. *Underwater Technology: The International Journal of the Society for Underwater*, 31(2):77–92, mar 2013.

[79] John C. Mankins. Approaches to strategic research and technology (R&T) analysis and road mapping. *Acta Astronautica*, 51(1-9):3–21, 2002.

[80] Robert Tierney, Wahid Hermina, and Steven Walsh. The pharmaceutical technology landscape: A new form of technology roadmapping. *Technological Forecasting and Social Change*, 80(2):194–211, 2013.

[81] BL Britt, MW Berry, M Browne, M Merrell, and J Kolpack. Document classification techniques for automated technology readiness level analysis. *Journal of the American Society for Information Science and Technology*, 59(4):675–680, 2008.

[82] Scott Cunningham. Big data and technology readiness levels. *IEEE Engineering Management Review*, 42(1):8–9, mar 2014.

[83] S.L. Cornford and L. Sarsfield. Quantitative methods for maturing and infusing advanced spacecraft technology. *2004 IEEE Aerospace Conference Proceedings*, pages 663–681, 2004.

[84] Don Clausing and Maurice Holmes. Technology Readiness. *Research Technology Management*, 53(4):52–59, 2010.

[85] P. John Clarkson, Caroline Simons, and Claudia Eckert. Predicting Change Propagation in Complex Design. *Journal of Mechanical Design*, 126(5):788, 2004.

[86] Steven D. Eppinger, Nitin R. Joglekar, Alison Olechowski, and Terence Teo. Improving the systems engineering process with multilevel analysis of interactions. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, 28(04):323–337, sep 2014.

[87] Robert Shishko, Donald H. Ebbeler, and George Fox. NASA technology assessment using real options valuation. *Systems Engineering*, 7(1):1–13, 2004.

[88] Robert G Cooper. Perspective: The Stage-Gate Idea-to-Launch Process - Update, what's new, and NexGen Systems. *Journal of Product Innovation Management*, 25:213–232, 2008.

[89] Robert G. Cooper. *Winning at New Products*. Basic Books, New York, NY, 4 edition, 2011.

[90] John K. Christiansen and Claus J. Varnes. Formal rules in product development: Sensemaking of structured approaches. *Journal of Product Innovation Management*, 26(5):502–519, 2009.

[91] Jeffrey B Schmidt and Roger J Calantone. Escalation of Commitment During New Product Development. *Journal of the Academy of Marketing Science*, 30(2):103–118, 2002.

[92] Judith Behrens and Holger Ernst. What Keeps Managers Away from a Losing Course of Action? Go/Stop Decisions in New Product Development. *Journal of Product Innovation Management*, 31(2):361–374, 2014.

[93] Arnd Huchzermeier and Christoph H. Loch. Project Management Under Risk: Using the Real Options Approach to Evaluate Flexibility in R & D. *Management Science*, 47(1):85–101, jan 2001.

[94] Kim E van Oorschot, Fred Langerak, and Kishore Sengupta. Escalation, Deescalation, or reformulation: Effective interventions in delayed NPD projects. *Journal of Product Innovation Management*, 28:848–867, 2011.

[95] D. Kahneman and D. Lovallo. Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking. *Management Science*, 39(1):17–31, jan 1993.

[96] OSD Manufacturing Technology Program. Manufacturing Readiness Level (MRL) Deskbook. Technical report, The Joint Service/Industry MRL Working Group, 2011.

[97] United States Government Accountability Office. Better Management of Technology Development Can Improve Weapon System Outcomes. (July), 1999.

[98] United States Government Accountability Office. NASA Assessments of Major Projects. Technical Report March, 2016.

[99] United States Government Accountability Office. Assessments of Selected Weapon Programs. Technical Report March, 2016.