1

On time, reliability, and spacecraft[1]

1.1 On time and reliability

Tempus edax rerum (time, devourer of all things). This exclamation by the Roman poet Ovid is meant as a reflection on the human condition and its ephemeral nature. But for an engineer, this phrase can also take a different, less profound but equally thought-provoking meaning: that things fail in time. Engineering artifacts degrade and fail in time; just how they do so, this particular aspect of their relationship with time, is the realm of reliability engineering.

1.1.1 Reliability: from the word to the engineering discipline

Reliability is a popular concept that has been celebrated for years as a commendable attribute of a person or an artifact (Saleh and Marais, 2006). The *Oxford English Dictionary* defines it as "the quality of being reliable, that may be relied upon; in which reliance or confidence may be put; trustworthy, safe, sure." Although many words and expressions in the English language seem to have been coined by or attributed to Shakespeare, it seems we owe the word *reliability* to another English poet who, along with William Wordsworth, founded the English Romantic Movement, namely, Samuel T. Coleridge (1772–1834). The first recorded usage of the word *reliability*

[1] This chapter was written in part in collaboration with Karen B. Marais, and it is based in part on an article published in *Reliability Engineering and System Safety* (Saleh and Marais, 2006).

Spacecraft Reliability and Multi-State Failures: A Statistcal Approach, First Edition.
Joseph Homer Saleh and Jean-François Castet.

dates back to 1816. In praise of his friend the poet Robert Southey, Coleridge wrote (Coleridge, 1983; our emphasis):

> He inflicts none of those small pains and discomforts which irregular men scatter about them and which in the aggregate so often become formidable obstacles both to happiness and utility; while on the contrary he bestows all the pleasures, and inspires all that ease of mind on those around him or connected with him, with perfect consistency, **and (if such a word might be framed) absolute reliability**.

From this modest, almost apologetic beginning in 1816, reliability grew into an omnipresent attribute – with qualitative and quantitative connotations – that pervades every aspect of the present-day technologically intensive world.

Beyond the etymology of the word, the discipline of reliability engineering emerged in the 1950s, its existence officially recognized in a report by the Advisory Group on Reliability of Electronic Equipment (AGREE):

> The authoritative announcement of the birth of reliability engineering was provided by the AGREE report on June 4, 1957. The report [. . .] provided all the [US] armed services with the assurance that reliability could be specified, allocated, and demonstrated; i.e., that a reliability engineering discipline existed.
>
> (*Coppola, 1984*)

AGREE was jointly established in 1952 between the US Department of Defense and the American electronics industry. Its mission was (1) to recommend measures that would result in more reliable equipment, (2) to help implement reliability programs in government and civilian agencies, and (3) to disseminate a better education of reliability (Coppola, 1984). The AGREE report in 1957 announced the birth of reliability engineering, but how did the discipline come about?

1.1.2 Brief (pre)history of reliability engineering: the enablers and the catalyst

The essential ingredients for reliability engineering are *probability* and *statistics*. These constitute the analytical foundation upon which rests this and many other engineering disciplines. Tradition has it that we owe the theory of probability to two Frenchmen, Blaise Pascal and Pierre de Fermat. They established the theory in 1654 in a famous exchange of letters spurred by a challenge posed to Pascal by a French nobleman who had an interest in gaming and gambling (Apostol, 1969). The theory was confined to games of chance until Laplace, in 1812, introduced a series of new techniques – mostly an extensive application of analysis to probability – into probability and statistics and expanded their scope of application to numerous practical problems, other than gaming (e.g., demographics, population estimation, life insurance).

In addition to probability and statistics as essential pillars that enabled the rise of reliability engineering, the idea and practice of *mass production* – the manufacture of goods in large quantities from standardized parts – is another fundamental ingredient in the development of reliability engineering. Interest in the quality of a product goes as far back in time as humans created artifacts (Duncan, 1974):

> As far back as the Middle Ages, the medieval guilds insisted on a long period of training for apprentices [. . .]. Such rules were in part aimed at the maintenance of quality.

Such craftsmanship, however, while apt to deal with quality issues during the design of a single or small number of artifacts, could not have satisfied the need for quality in a high-volume production environment. Mass production therefore heightened the need for a new discipline capable of dealing with quality issues in high-volume production settings. In response to this quality pressure, first came *statistical quality control* in the late 1920s and early 1930s with the groundbreaking contributions by a young engineer from Bell Laboratories, Walter Shewhart (1891–1967), then reliability engineering in the mid 1950s.

Mass production is traditionally associated with Henry Ford and his Model T car. However, although Ford popularized the concept around 1910, high-volume production of items from standardized parts, that is, mass production, had been achieved many years earlier. For example, during the American Civil War, the Springfield Armory was producing over 300 000 rifles per year for the Union Army in 1863, almost as many as the peak production of the Model T Ford. One key idea at the root of mass production is what came to be called the American System of Manufacturing, or the use of standardized, interchangeable parts. The system was developed by Eli Whitney in the United States in the late 1790s (Maier *et al*., 2003), although credit for the idea of interchangeable parts goes to a French gunsmith, Honoré le Blanc, who first suggested it in the mid eighteenth century (but did not go very far with it because other gunsmiths saw this idea as a threat to their livelihoods and opposed it).

At the onset of World War II, with statistics, in particular the theory of sampling, and mass production well established, reliability engineering was ripe to emerge. The catalyst came in the form of an electronic component, the vacuum tube (specifically the Audion or the triode, which was a major improvement on previous tubes), invented by an American, Lee de Forest, in 1906. The tube, which for all practical purposes initiated the electronic revolution, enabled a series of applications such as radio, television, radar, and others. How did the tube contribute to the birth of reliability engineering? Electronics played a critical role in World War II and contributed to the Allies winning the "wizard war":

> The vacuum tube, the active element that made the wizard war possible, was also the chief source of equipment failure. Tube replacements were required five times as often as all other equipments.
>
> (*Coppola, 1984*)

It is this experience with the vacuum tubes that prompted the US Department of Defense to initiate a series of studies for looking into these failures after the war. These efforts were eventually consolidated and gave birth to a new discipline, reliability engineering. In short, the catalyst that accelerated the coming of this new discipline was the unreliability of the vacuum tube.

By the time the existence of reliability engineering was acknowledged in the late 1950s, attitudes began to change since the unwelcoming response that Walter Shewhart's *Statistical Quality Control* work received, and the "deep-seated conviction of American production engineers [. . .] that laws of chance have no proper place among scientific production methods" (Duncan, 1974, citing Freeman, 1936), was slowly being replaced by a better predisposition toward statistics and probability in product design.

In a faint echo of Ovid's reflection on TIME, and highlighting the foundational idea of reliability engineering, C. Raymond Knight (1918–), who contributed to the emergence of this discipline, noted in 1991:

> It may seem strange today, but at that time [1950s] there was considerable resistance to recognizing the stochastic nature of the time to failure, and hence reliability.[2]
>
> (*Raymond Knight, 1991*)

After its establishment, reliability engineering evolved in several directions, on the one hand toward increased specialization in its statistical techniques, and on the other hand toward a physics of failure approach and what came to be called *structural reliability*, which was concerned with the structural integrity of buildings, bridges, and other constructions (Denson, 1998). In addition, reliability improvement programs began to emerge, along with the specification of quantitative reliability requirements, marking the beginning of the contractual aspect of reliability.

The story of reliability engineering intersects another major technological development, the emergence of the space industry. In a serendipitous accident of history, these two events, the official birth of reliability engineering and the beginning of the space age with the launch of the first active space system, occurred in the same year, 1957. This book is at the intersection of these two developments and it brings the former, reliability engineering, to bear on the latter, space systems.

1.2 On spacecraft and reliability: early studies

On October 4, 1957, a small beeping satellite, Sputnik, heralded the beginning of the space age. From this humble start, the space industry grew into an impressive

[2] At present, reliability is more formally defined as the probability that an item will perform a required function under stated conditions for a given period of time.

$100+ billion industry five decades later. Roughly speaking, around 6500 spacecraft were launched in the five decades after Sputnik. And although the launch rate has been highly variable (Hiriart and Saleh, 2010), a rough estimate would set it at present to around 80 to 100 spacecraft per year. Spacecraft today fulfill a myriad of functions, from defense and intelligence missions (early warning, reconnaissance, etc.), to science missions (Earth observation, interplanetary probes), and communication functions (direct-to-home, fixed satellite services, and mobile satellite services).

Spacecraft can cost several hundred millions of dollars to design and launch,[3] and as such reliability is essential for these systems. More generally, reliability is a critical design attribute for high-value systems operating in remote or inhospitable environments such as spacecraft or subsea installations. Since physical access to these assets is difficult or impossible, maintenance cannot be relied upon to compensate for substandard reliability (Rausand and Høyland, 2004). As a result, designing high reliability into these systems is an essential engineering and financial imperative.

For space systems, statistical analysis of flight data, in particular of actual on-orbit (field) anomaly and failure data, would provide particularly useful feedback to spacecraft designers. For example, such analyses can help guide spacecraft testing programs and provide an empirical basis for subsystem redundancy and reliability growth plans. Analyzing spacecraft failure behavior on orbit, and identifying their subsystems' actual reliability profiles, not their reliability requirements (how they actually degrade and fail on orbit, not how they should or are expected to), can help spacecraft manufacturers prioritize and hone in on problematic subsystems that would benefit most from reliability improvements. Reliability improvements can be achieved through redundancy, increased testing prior to launch, or better design and parts selection, and these efforts would result in a decreased likelihood of spacecraft experiencing failure events. In addition, identifying whether specific spacecraft subsystems experience "infant mortality," for example, would provide a clear opportunity for spacecraft manufacturers and equipment providers to develop burn-in procedures for weeding out early failures in such subsystems. Statistical analysis of on-orbit failure and spacecraft reliability can also provide important and actionable information to stakeholders other than spacecraft manufacturers. For example, satellite operators may be particularly interested in the reliability profiles of their on-orbit assets, for planning and risk mitigation purposes, and insurers evidently rely on such analysis and information to set up their policy and insurance premiums.

The importance of statistical analysis of on-orbit failure data was recognized early in the advent of the space age. The following subsections provide a brief overview of past spacecraft reliability studies.

[3] Except for microsatellites, which are typically in the $10–50 million range, and ongoing efforts are seeking to significantly reduce this price tag. Whether useful functions can be performed on orbit below this range remains to be seen.

1.2.1 Overview of early spacecraft reliability and on-orbit failure studies

A few years after the launch of the first satellites, statistical analyses of spacecraft reliability and on-orbit failures began to appear.

One of the earliest reliability studies, according to Leventhal *et al*. (1969), was published in 1962, and it analyzed the failure behavior of 16 satellites launched before November 1961 (ARINC, 1962). Over the years, similar analyses were conducted with larger sample sizes or spacecraft populations. For example, Bean and Bloomquist (1968) analyzed the failure behavior of 225 satellites; Timmins and Heuser (1971) and Timmins (1974; 1975) analyzed the failure behavior of 57 satellites; and Hecht and Hecht (1985) and Hecht and Fiorentino (1987) analyzed the failure behavior of some 300 satellites. The present work analyzes the anomaly and failure behavior of 1584 Earth-orbiting satellites launched between January 1990 and October 2008. The choice and impact of the sample size on the statistical results, and their relevance, are discussed in Chapter 2.

Early spacecraft reliability studies assumed an exponential distribution and constant failure rate. This assumption, however, was challenged by Timmins and Heuser (1971) who showed that, for their small sample of 57 spacecraft launched from NASA Goddard Space Flight Center, the failure rate was not constant but higher in the early days on orbit:

> The number of failures per spacecraft were abnormally high for the first 30 days in space. The number of first-day failures departed even more from the longer trend.

This finding of spacecraft "infant mortality" and a decreasing failure rate was repeated in subsequent studies (Timmins, 1974; 1975), and led Baker and Baker (1980) to comment that "those spacecraft that last, last on and on," which in effect reflects for these authors the absence of wear-out failures in spacecraft.

Hecht and Hecht (1985) analyzed a different population of spacecraft than the one used in the previous four studies (the 57 NASA spacecraft). Their sample consisted of some 300 spacecraft launched between 1960 and 1984, and covered 96 different space programs. Their analysis also found a decreasing failure rate in their spacecraft sample, and they took issue with the constant failure rate models proposed in the military reliability handbook, MIL-HDBK-217, as unrealistic for system reliability predictions. MIL-HDBK-217 was first developed in 1961 and revised several times afterward. Similar conclusions were advanced by Krasich (1995) and Sperber (1990; 1994) who noted a qualitative agreement in prior studies "that as the mission goes on, risk per unit time to surviving spacecraft decreases."

Some studies explored causal hypotheses for this actuarial result, the decreasing failure rate of spacecraft. Norris and Timmins (1976) for example stated that "a plausible explanation for this decreasing trend is that the data sample includes a wide variety of components, and as the high risk component fail, the remaining units are the ones with lower failure rates." Baker and Baker (1980) excluded the space

environment as a possible cause of this trend by noting that "space itself is not a harsh environment for spacecraft; for if it were, the hazard rate would increase as a function of time as cumulative exposure precipitates failures." Hecht and Fiorentino (1987) argued for the existence of decreasing failure rate as follows:

> in terms of spacecraft reliability, that the equipment has survived under the environmental stresses experienced during a period of m years on orbit does not preclude the occurrence of a phenomenon during year m+1 that produces a greater stress and hence lends failure. However, the likelihood that greater stresses will be encountered decreases over successive intervals, and that leads to the decreasing failure rate.

The quality of these arguments pertaining to the cause of spacecraft infant mortality is questionable.

1.2.2 Beyond the failure rate emphasis in spacecraft reliability studies

Sperber (1994) suggested that "the causes of the [on-orbit failures and] anomalies are not random overstress or wear-out, but are perhaps weakness in design or execution uncovered in the mission." His comment echoes an earlier finding by Bean and Bloomquist (1968) that, for the sample they had, that is, 225 spacecraft launched prior to 1968, the "most common cause of spacecraft anomalies is inadequate design, representing nearly 60% of all incidents with assignable causes."

More recent studies revolved around specific spacecraft subsystems. For example, Cho (2005) and Landis *et al*. (2006) focused on failures in spacecraft power subsystems, Brandhorst and Rodiek (2008) on solar array failures, and Roberston and Stoneking (2003) on failures in attitude control subsystems. Sperber (2002) and Tafazoli (2009) analyzed not just a single subsystem's failures but the comparative contribution of various subsystems to spacecraft on-orbit failures. And instead of spacecraft subsystems, Bedingfield *et al*. (1996) focused on spacecraft failures due only to the natural space environment.

1.3 Book organization

In Chapter 2, a statistical analysis of spacecraft failure data is conducted and nonparametric spacecraft reliability results are derived.

In Chapter 3, parametric analysis of spacecraft reliability is conducted, and single Weibull as well as mixture distribution models are derived using the maximum likelihood estimation (MLE) method.

In Chapter 4, the previously analyzed failure data is specialized first by spacecraft mass category and then orbit type. Nonparametric analysis and parametric reliability models are then derived for these different types of spacecraft.

In Chapter 5, the statistical failure analysis is extended to spacecraft subsystems, that is, the analysis is narrowed down from system-level to subsystem-level

failures, and reliability results, both nonparametric and parametric, are derived for spacecraft subsystems.

Chapter 6 is a turning point in the book. The previous chapters deal with the reliability of spacecraft and spacecraft subsystems. Only two states are considered up to this chapter, *operational* and *failed*, and the (sub)systems are analyzed and modeled as being in one of these two states. In reality, engineering artifacts can experience failure events of varying severity, and thus transition from fully operational to various states of partial degradation and failure. Chapter 6 extends the previous analyses of reliability, in its traditional binary-state understanding, to account for spacecraft anomalies and failures of various severities. Partial failures constitute a significant portion of the anomalous events that spacecraft experience on orbit, and as such their analysis provides additional and important information for understanding the spacecraft and subsystems' failure behavior on orbit. Chapter 6 can be characterized as an exploratory data analysis of failure distributions on orbit, as well as the time to anomaly and failure of spacecraft subsystems. The chapter serves as an easy transition between the formal binary-state understanding of reliability conducted in the previous chapters and the formal multi-state failure analysis in the following chapters.

Chapter 7 provides a formal multi-state failure analysis of spacecraft subsystems. And Chapter 8 extends the previous analyses to include considerations of survivability of spacecraft and space-based networks. In addition, Chapter 8 introduces an important tool for the modeling and analysis of stochastic processes, namely, stochastic Petri nets (SPNs), and develops SPN models for the analysis of spacecraft survivability, building on the detailed models of subsystems' multi-state failures developed in Chapter 7.

The analysis and results in each chapter build to some extent on those in previous chapters. As a result, cross-references between chapters are frequent. An effort has been made to facilitate readability by repeating some of the essential material for each chapter. The reader who wishes to go through the whole book in one sitting can easily skip through the overlapping parts.

Two appendices are included in this book, and they address specialized topics under the broad theme of this work. Appendix A focuses solely on communication satellites in geosynchronous orbit (GEO). These satellites represent an important segment of the space industry, and, as such, a dedicated appendix for their reliability analysis is provided here. In addition, a health scorecard is developed, summarizing for each subsystem its track record of on-orbit anomalies and failures. Appendix B focuses solely on the electrical power subsystem (EPS) on board spacecraft, and it analyzes the differences in failure behavior of the EPS in low Earth orbit (LEO) and geosynchronous orbit (GEO).