International Scholarly Research Network
ISRN Applied Mathematics
Volume 2012, Article ID 850686, 36 pages
doi:10.5402/2012/850686

Review Article

System Reliability at the Crossroads

Vitali Volovoi

School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

Correspondence should be addressed to Vitali Volovoi, vitali@gatech.edu

Received 30 August 2012; Accepted 10 October 2012

Academic Editors: J. R. Fernandez and Y. Tsompanakis

This paper surveys the current state of research related to the modeling and prediction of failures of engineering systems. It is argued that while greater understanding of the physics of failure has led to significant progress at the component level, there are significant challenges remaining at the system level. System reliability, a field of applied mathematics that addresses the latter challenges, is at a juncture where fundamental changes are likely. On the one hand, the traditional part of the field entered a phase of diminishing returns, largely having followed the trajectory of the Cold-War era technology development: golden years of rapid growth in the 1950s and 1960s, followed by maturation and slowing down in the ensuing decades. On the other hand, the convergence of several technologies related to data collection and processing, combined with important changes in engineering business and government priorities, has created the potential for a perfect storm that can revive and fundamentally transform the field; however, for this transformation to occur, some serious obstacles need to be overcome. The paper examines these obstacles along with several key areas of research that can provide enabling tools for this transformation.

1. Background

1.1. Introduction: Avoiding Failures as an Inherent Part of Engineering

It can be argued that studying the failures of engineering systems is as old as engineering systems themselves. By its very definition, successful design implies that failures are avoided: the airplane designed by S. P. Langley failed, while that made by the Wright brothers did not. As a result, any engineering discipline relies on practices that avert failures. Aging and associated degradation of the system's properties might introduce additional challenges, as it might take years before a technological flaw manifests itself. For example, the design of the I-35W bridge over the Mississippi river in Minneapolis would have ensured that the bridge would not collapse under normal operating conditions, if it were not for the design flaw related to the sizing of Gusset plates. The bridge, which was opened for traffic in 1967 and

collapsed forty years later, was built using design principles specified in the construction codes of the time that should have been sufficient to prevent failures. In the case of I-35W bridge, the influence of corrosion on integrity of bridges is well known, so appropriate safety factors are used in construction, combined with the periodic inspections for estimation of the corrosion impact. Steel bridges have been constructed for more than a century (the first steel bridge was built in 1870 in Kymijoki, Finland, and the Brooklyn bridge was built in 1883); the structural risks (including those due to aging) are now well understood, most of the failures having occurred before the corresponding technologies become established.

More generally, the process of the maturation of certain technologies or design implies a transition from "radical" design (with relatively low expectations about its success) to "normal" design, when enough knowledge has been accumulated to develop specific procedures and requirements that effectively split the potential risks of failure into three categories: the risks that are sufficiently improbable that they may be ignored, those that are implicitly dealt with by the system design, and finally those that are are explicitly dealt with by the design (the latter usually involves quantitative analysis) [1, 2]. While distinction between the radical and normal designs is usually well understood by the engineers involved in the system design, quite often this distinction is lost in translation by the time the message is conveyed to the decision makers. As a result, existing technologies too often are pushed beyond the limits established by historical practice, without the full realization of the consequences. (The history of the Swedish battleship Vasa can serve as Exhibit A of such a risky project. The ship that was supposed to symbolize the military might of the Swedish kingdom sunk within minutes of starting its maiden voyage [3]. The Swedish king Gustavus Adolphus provided enough pressure from the top to override the concerns of the ship builders over the Vasa's stability (the ship failed a stability test that included a squad of sailors running from side to side of the deck). Such pressure is painfully familiar to the investigators of modern disasters, including the explosion of the Challenger and the Deepwater Horizon disaster [4, 5]. One can recall that the very name "Shuttle" was supposed to convey a sense of routine operations for a system that is arguably the most complex nondistributed engineering system ever designed (in other words, it is the engineering equivalent of a blue whale, rather than clonal colonies of Aspens connected to a single root). This forced "normalization" of the design was important in order to justify the Shuttle *raison d'être* as a cheap and reliable means for launching satellites [6, 7].) Nevertheless, as the collapse of the I-35 bridge demonstrates, design flaws are still possible, and indeed they can became apparent more than forty years after the bridge became operative.

In the case of new technologies, the time scale of degradation might exceed the age of the technology itself, making sound design decisions more challenging. For example, polyimide coating for electric wires (introduced by DuPont under the trade name Kapton in the 1960s) was widely adopted in the aerospace industry due to its light weight and good insulation properties. However, by early eighties, an increasing number of wire shorts (some with significant consequences) forced the U.S. Navy to ban this material. It turned out that as polyimide ages, cracks develop, and if a small short occurs due to those cracks, the resulting spark triggers a chemical reaction that converts the insulator into conductor, leading to what is called arc-tracking (when an increasingly powerful spark propagates along the wires). The aging of the insulation material is accelerated by humidity, which explains why the U.S. Navy faced this problem first. Civil aircraft continued to use polyimide wiring for two more decades, until several high-profile accidents (including TWA Flight 800 in 1996, where the probable cause of the explosion of the fuel tank was attributed to a wire short [8]) led to phasing out the use of this material in wiring for aerospace applications. It must

be noted that each Space Shuttle has still contained about 150 miles of Kapton wiring until its retirement, leading to NASA's concern about the risks of wire shorts. In particular, one considered consequence of such a wire short could be an uncommanded firing of one of the Shuttle Reaction Jet Drivers (RJDs) while the Shuttle is docked at the International Space Station (ISS). It was calculated that, as the ISS was expanded and its weight increased, such an uncommanded firing of an RJD would lead to catastrophic consequences for both the shuttle and the ISS before the diagnostic software would be able to shut down the uncommanded RJD [9].

Trial and error is a vital part of technical innovation [10], but its efficiency is greatly weakened by delayed feedback, when decades might pass between the beginning of the "trial" and manifestation of the "error," thus greatly obscuring any causal relationship between the two. The broader implication of this connection between the success of any complex system (not only a technological, but a biological or a sociological system as well) and the strength of the feedback is a fascinating topic that has recently attracted attention in popular economic literature [5], as well as from complexity scientists [11].

The normal wear and tear of engineering systems during their life cycle can be contrasted with external shocks on the engineering structures that go beyond the standard design practices of the time. For example, the fact that the Lighthouse of Alexandria collapsed during the earthquakes of 303 and 323 A.D. does not render the design a failure and does not negate its status as one of the seven wonders of the ancient world. The difference between normal operating conditions and external shocks is not clear-cut and can be considered as evolving, so that the design of modern engineering systems increasingly includes protection against any credible risks to the system during its life time. For example, modern construction codes in earthquake-prone areas include safety requirements that are based on earthquakes of magnitudes that are deemed to be sufficiently likely for the area. Similarly, the skyscrapers should withstand winds that are representative of winds likely to occur every 50 years [12]. The distinction can be still made with respect to safety versus normal functionality, as the buildings are not required to continue their functionality after the earthquake; instead, they are designed to protect (and not harm) the occupants during the earthquake. Since external shocks (such as hurricanes and earthquakes) occur infrequently, their influence on the successful design and operation of engineering systems is also delayed and weakened as a result.

1.2. Reliability as a Distinct Engineering Field

Several recent good surveys of the history of reliability as a distinct engineering field are available, including [13] that covers the early years, and [14] that provides an overview on the recent developments. There are several journals exclusively devoted to this field, and a number of more general-purpose journals regularly publish papers on this topic. Similarly there are several major conferences, including the Reliability and Maintenance Symposium (RAMS) in the United States, and the European Safety and Reliability Conference (ESREL). Due to the sheer size of the research in the field, it is inevitable that the selection of this survey is far from exhaustive and to a certain degree reflects the viewpoint of the author.

Conceptually, reliability engineering includes consideration of the following aspects of system behavior.

(i) Explicit recognition that not all entities that comprise an engineering system will be functioning as intended. This is a critical distinction of reliability engineering,

as opposed to "classical" engineering where specific entities (e.g., components) are introduced into the system in order to deliver certain functionality with an implicit expectation that the functionality will indeed be delivered. More generally, this can be viewed as the recognition of variability of the properties and therefore the performance of individual constituents of an engineering system and/or the environment in which the system in question operates. Indeed, in the absence of this variability, all parameters of those constituents assume their nominal values, and proper functionality of the system falls under the realm of expertise of classical engineering. When deviations from those nominal behaviors reach certain thresholds, errors or component failures might occur. Finding those thresholds is far from trivial, as they depend on many factors, including (potentially very complex) dynamics of the system.

(ii) Quantification of the likelihood of individual errors/component failures (and potentially their correlations).

(iii) Determination of the impact of individual errors/component failures on system-level objectives (e.g., delivering the functionality of the system, ensuring that no external harm is done).

(iv) Developing strategies for handling those errors/component failures and finding design alternatives aimed at reducing the chances of system failures or their effects.

(v) Conducting trade-offs studies, which provide the guidelines for selecting among the proposed design alternatives.

The roots of reliability concepts can be traced to the early twentieth century when mass-manufactured engineering products, such as cars, came onto the scene. However, the foundations of reliability were established after the Second World War with the advent of computer technologies that were originally vacuum-tube-based. As a result of the high frequency of failure of vacuum tubes and other components, a fundamental understanding of the design and operation of reliable systems with large numbers of less reliable components was reached [15, 16]. Due to the fact that these systems were often electronic, the distinct dichotomy of approaches to system reliability still evident today can be traced to those developments.

On the one hand, hardware reliability issues are inevitably related to the reliability of components and their relationship with the reliability of the system. At the component level, this led to the statistical characterization of components' time to failure; at the system level, it led to the application of Boolean algebra methods, such as Fault Tree Analysis (FTA) and Reliability Block Diagrams (RBDs) [17, 18], as well as theory related to stochastic processes of repairable systems [19] that incorporate the statistical characterization of components' failures and evaluate the resulting impact on the system.

On the other hand, software failures are not necessarily related to components *per se*; instead, the fundamental building blocks that can potentially cause system failures are errors [20], which are generally characterized as the wrong states for some parts of the system which have a potential to cause system failure. In this setting, faults are characterized as adjudged or hypothesized causes of the errors. Importantly, errors cause system failures only if they are not "caught" prior to reaching the system interface, where those errors can lead to altering system functionality.

This distinction related to the importance of component decomposition in system modeling between hardware- and software-induced failures is not as critical as might seem at

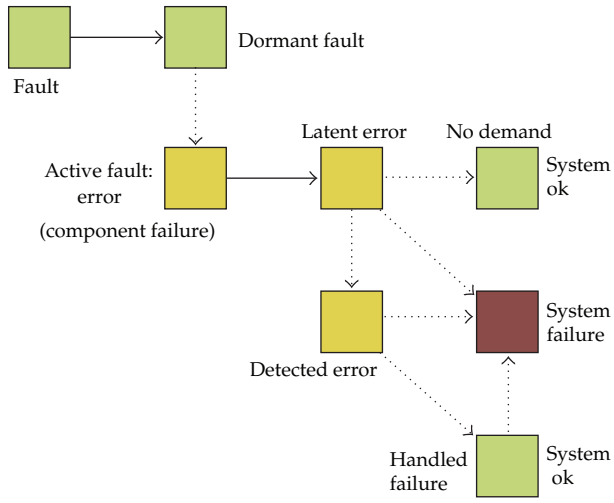

Figure 1: Abstract failure progression for failures due to internal faults. Dashed lines indicate time delay (as opposed to solid lines that indicate the absence of such delay).

first glance, due to the modular structure of modern software, as espoused by the principles of object-oriented programming [21]. As a result, individual software modules can be construed as the software equivalent of components, and so both hardware and software are potentially suited for system reliability methods that rely on inferring the system's properties from its components. Here "component" (even for hardware) is understood in the functional rather than physical sense. Indeed, even for hardware, a functional approach to failure investigation is beneficial for large complex systems, by separating the functionality of an entity from its physical implementation (e.g., functional view of Failure Mode and Effect Analysis (FMEA); see, e.g., [22]).

A general diagram (see Figure 1) can be drawn to represent an internal fault progression in a system. The fault can always be interpreted as being dormant in the beginning ($t = 0$). This point of reference does not need to coincide with the beginning of the system operation t_b: the former, generally speaking, precedes the latter $t_b > 0$. If the fault stays dormant during the duration of the operation of the system s, the system is obviously not affected by this fault. On the other hand, if the fault becomes active at some point ($t_s < t_b + s$), it causes an error (or component failure in the case of hardware). This error or component failure can be considered as latent until it is detected and possibly handled by the system control logic (e.g., by means of redundancy); otherwise, the error propagates to the interface of the system and becomes a system failure. The timing is extremely important in this context (e.g., a backup component may provide the functionality of the failed component for only a limited amount of time, as the back-up power generator). As discussed below, the modeling of this timing requires dynamic system reliability tools.

Fundamental differences among the software, hardware, and "human-ware" portions of complex systems do exist, however. To discuss these differences, it is useful to adopt terminology that is domain-neutral. To have a common language for failure description is critical, as modern complex systems consist of a combination of hardware and software, with human operators playing an important role as well, and with the boundaries among those three domains constantly shifting. The terminology provided below is perhaps the closest

to the one used in the context of failure modeling that involves human and organizational factors [23]. This is not entirely surprising, given the fact that consideration of the human factors in the failure of engineering systems lagged behind developments in software and hardware, so the insights from the other two domains could be utilized. In order to understand the behavior of an entity that is a part of the system (i.e., a component or a module), one needs to characterize the sources of the variability of its performance and their implications for the system performance. Those sources can be grouped into three categories.

(i) Internal sources of variation (due to the variability of internal properties of the component in isolation). The interpretation is most familiar for hardware entities, in terms of both entity-to-entity variability (e.g., due to manufacturing tolerances), or variability in time for the same entity (e.g., stochastic behavior due to aging) (see the discussion below on component reliability). Similarly, human operators and organizations certainly exhibit large entity-to-entity variability, while the time variability for a given entity is somewhat more obscured given the fact that a human (or a group of humans) represents an open system that constantly adapts to the environment. As a result, it is difficult to separate the internal variability from that which was externally induced. In contrast, software does not exhibit any internal variability: the same version of the software is identical in all systems, and it does not change with time (software updates can be considered as a result of external actions). While it is feasible to consider software that internally relies on pseudorandom number generation, certification of such nondeterministic algorithms for any safety-critical systems presents an interesting challenge for the future, and in any case the requirements for stringent predictability of the outputs can be expected even for those algorithms (e.g., finding an optimum value using stochastic search methods would rely on certain convergence criteria).

(ii) Environmental, exogenous, or external sources of variation (the source is outside of the system). While the shocks described earlier fall under this category, and so this variability is relevant for the hardware component, the functionality of a hardware component usually has a fairly simple relationship with respect to this variability. In fact, most of the time the desired output is constant for a given range of environmental conditions (e.g., structural components that are supposed to maintain their integrity under a range of loads, temperatures, etc.). Even in the case of mechanisms, the mapping between expected inputs and outputs is low-dimensional in terms of the "signal" processed (e.g., valve), while the rest of the environmental parameters are effectively considered as "noise" and should be filtered out. One of the challenges for the hardware component is that those "noise" parameters are not always well understood, as the hardware represents an open system (so, e.g., an unexpected degradation mechanism can interfere with the performance of the component). Human operators also represent open systems and therefore are susceptible to external influences in terms of filtering out the noise, but the dimensionality of the signal is significantly larger than that of the hardware. Finally, the software has a very well-defined interface, so the "noise" is zero. The challenge of predicting software performance stems from the complexity and high dimensionality of the signal. (It is important to distinguish the noise at the interface from the internally processed noise: for example, an input to the software can be a periodically sampled value of some sensor, and the output might be a filtered or denoised value. In this setting the noise in the received input is

still a signal (as the whole purpose of the software is explicitly designed to handle this noise), so the interface noise is still zero.) The dimensionality of signal inputs for human operators is an interesting question, as it requires a distinction between the "sharp end" operators that make tactical decisions on a short-time scale, and strategic decision makers (effectively the "designers" of the system in the broad sense, e.g., the developers of safety procedures, etc.). The actions of the former are based on the skills developed during training. The associated dimensionality of inputs is certainly less than that of software, something that has some interesting implications for the role of the human operators at the "sharp end" (see discussion later in the paper). In contrast, the dimensionality of the inputs for the latter type of operators is large, which makes it very difficult to characterize, unless it falls under the category of some external set of rules (e.g., regulations that are imposed to ensure safety).

(iii) Intercomponent coupling (in [23] this is referred to as upstream-downstream coupling to emphasize the potential for the presence of loops). While the characterization of these sources across the hardware, software, and "human-ware" domains is similar to the external sources, the important difference is potential for feedback and feedforward loops. This distinction is mostly a matter of convenience: as the boundary between the system under consideration and its environment is not well defined in the first place, it certainly makes sense to use the presence of loops as a criterion for including those entities into the system.

The rest of the paper is organized as follows: first, the challenges associated with modeling system reliability are discussed and illustrated with a simple example. Next, the current approaches for ensuring system reliability are reviewed. This is followed by a section describing the current convergence of several trends and identifying several promising directions of research that hold the potential to improve the safety and reliability of complex systems. Finally, conclusions are offered in the last section.

2. System Reliability Challenges

There are several fundamental challenges that system reliability faces, and they will be briefly reviewed next.

2.1. Errors of Commission

Individual components can contribute to system failure not only by failing to perform their intended functions (errors of omission) but also by performing unanticipated actions (errors of commission). This significantly increases the effective state space describing the system: instead of a binary choice for a component (it either functions or it does not), one needs to consider other dimensions of the component's state. More importantly, foreseeing those dimensions of the component's state which are potentially detrimental or hazardous to the system is far from trivial. As the design progresses, those other dimensions should be incorporated into the functional requirements for the component. For example, the primary functionality of the foam on the fuel tank of the Shuttle is to insulate the fuel and keep it at a sufficiently low temperature. When foam pieces break off the fuel tank during the shuttle launch, that functionality is not compromised; however (as we now know [24]), those pieces

can cause damage to other parts of the shuttle. So, a more complete set of requirements for the insulating foam would include the maximum permitted size of the foam pieces that break off during the launch.

2.2. Prioritization of Failure Modes Based on Their Likelihood

System failures do not have to be caused by a single component failure, but instead could be the result of an interaction of several deviations by components from their nominal states (while each deviation in isolation can be benign). In this context, a partial degradation of the component's performance becomes important, further increasing the underlying dimensionality of the state space (as those degradation levels must be distinguished). The resulting number of possible failure modes in a modern complex system is so large that it is imperative to develop an efficient means of prioritization of those modes. Risk matrices that used to prioritize failure modes usually have two dimensions: the risk and consequences of a given scenario [25] (in practice, there is a third dimension that is distinct from the first two and is related to the effectiveness of potential mitigation actions). Probabilistic risk quantification is most commonly used, but recently there has been a significant pushback by the software community that argues that probabilistic methods are not applicable to the software domain and advocate the use of formal verification techniques instead [2]. This creates conceptual difficulty at the system level when the combined effects of software and hardware need to be assessed. Incorporating software failures into the assessment of system failures is increasingly important as systems become "smarter" and more reliant on software.

2.3. Wide Array of Domain-Specific Intricate Failure Mechanisms

As discussed in the introduction, understanding failures is impossible without a thorough understanding of the specific domains involved (e.g., structures, controls, the human-machine interface). As a result, there is a Byzantine patchwork of reliability tools that are used by the experts to evaluate the reliability of components and systems.

In very general terms, two opposite approaches to reliability prediction can be identified: on the one hand, there is a domain-specific physics-based analysis that relies on extensive understanding of specific failure mechanisms for the components used as well as environment characterization (e.g., fatigue crack propagation in metals). On the other hand, there is a purely data-driven approach where field or test data is statistically analyzed, and predictions about reliability are made based on obtained distributions of the time to failure. In all practical cases, the data-driven approach actually includes some physics-based considerations, but they are rarely stated explicitly. For example, the selected parametric distributions (e.g., Weibull or Lognormal) at some point were motivated by the physics of failure, or historically performed well, which implies that the new designed system is deemed sufficiently similar from the physics of failure perspective to merit the use of the same type of distribution. As discussed below, the type of distribution can dramatically influence the quality of predictions.

Accelerated testing and system-level analysis both combine physics-based and data-driven approaches. The former relies on statistical representation of time to failure under a controlled environment and then utilizes physics-based considerations to predict the timing of failures in the field [26–28]. The latter aims at evaluating the reliability parameters of the system based on statistical information about component failures (obtained either from past

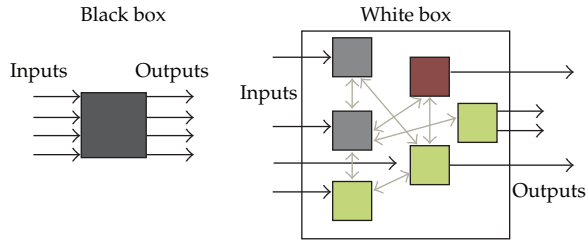

Figure 2: Black-box versus white-box models (reproduced with permission from [33]).

experience or based on physics-of-failure models) combined with the information about the interrelationships among the components. In the case of repairable systems, those system reliability parameters are not limited to time to failure, but might also include availability, the expected number of failures, and so forth. One can identify three levels of fidelity of modeling interrelationships among components in order of complexity:

(i) simple logical "and/or" (e.g., fault trees),

(ii) discrete event representation where the timing and order of events are taken into account, but the state space is discrete (e.g., Markov chains [29] and stochastic Petri nets (SPN) [30]),

(iii) models where both time and spatial description are continuous (e.g., multiphysics or agent-based simulation [31]).

The timing (and order) of events can have a fundamental impact on the likelihood of failure. For example, if a piece of foam had separated from a tank a fraction of a second later than it actually did, it might have been harmless. Accounting for the timing of events precludes the use of static tools (e.g., fault trees) and requires dynamic analysis (e.g., Markov chains, SPNs, agent-based simulation). In addition to the contrast between physics-based versus data-driven modeling, one can also note that the reliability and safety of engineering systems have been approached from the opposite direction in at least two distinct dimensions. On the one hand, there is white-box versus black-box dichotomy [32], which is based on whether the failure process of an entity is modeled with or without the explicit recognition of individual constituents (components) that comprise the entity. Here "component" refers to an elementary building block of a white-box (system) model, which can correspond to a lower-level entity if models are constructed hierarchically, or to the lowest level of the hierarchy, as determined by practical considerations (e.g., individual modules, such as line-replaceable units or LRU). White-box models explicitly model those effects by describing interrelationships among the entities that comprise the system (see Figure 2).

In contrast, black-box models do not require explicit modeling of the described effects, and they can be useful in analyzing the behavior of existing systems; however, they are of limited use in predicting the behavior of new systems or the impact of new features (e.g., condition-based maintenance) that are introduced in existing systems.

On the other hand, there is a nonrepairable versus repairable dichotomy [34], with the former approach dealing with a single failure event of an entity, and the latter addressing repeated failure events, which assumes the possibility of partial or full recovery from failures. While a nonrepairable entity is characterized by its lifetime distribution (e.g., its cumulative distribution function of time to failure), repairable entity behavior is described by a point process, and so must be characterized differently, for example, using the rate of occurrence

of failures (ROCOF), or the expected number of failures for a given time period. Any permutation of those choices translates into an appropriate set of tools: for example, selecting the black-box direction can lead to accelerated testing techniques for nonrepairable entities [26, 28] and to modeling repairable entities by means of stochastic processes.

To make things more exciting, the white-box approach entails selecting either the repairable or nonrepairable option both at the system (output) and component (input) levels. Boolean algebra methods (e.g., fault trees and reliability block diagrams) assume that both systems and the components comprising those systems are nonrepairable. This symmetry between the inputs and outputs characterization (and associated simplicity and clarity) is perhaps one of the reasons for the popularity of those tools. In contrast, the use of superimposed processes [34, 35] implies that both inputs and outputs are repairable entities. In this context, the state-space models (e.g., Markov chains and SPNs) can be classified as selecting repairable outputs and nonrepairable inputs within the white-box (system) approach.

2.4. A Simple Example

To illustrate the differences in terms of various modeling options in system reliability, let us consider a triple-redundant system, S_1, consisting of three identical components that each follow exponential distribution with the failure rate $\lambda_c = 10^{-4}$ 1/hour. The convenience of the use of exponential distribution that is fully defined by a single parameter is difficult to overestimate in the context of reliability analysis. Effectively, the use of exponential distribution implies that the system neither improves nor deteriorates with time. Let us further assume that our time horizon (life of the system) is $T = 10^4$ hours. If this system is nonrepairable, one can immediately observe that the probability of failure for each component is

$$P_{cf} = 1 - \mathrm{Exp}[-\lambda_c T] = 1 - \mathrm{Exp}[-1] \approx 0.6321. \tag{2.1}$$

Failure of the system requires that all three components fail simultaneously, so

$$P_{sf} = P_{cf}^3 \approx 0.256. \tag{2.2}$$

One can also find an equivalent constant failure rate for system S_2 that consists of a single component and would produce the same probability of failure

$$P_{S2} = P_{S1} = 1 - \mathrm{Exp}\left[-\lambda_{\mathrm{eq}} T\right] \Longrightarrow \lambda_{\mathrm{eq}} = -\frac{1}{T}\log[1 - P_{S1}] \approx 2.91 \times 10^{-5}. \tag{2.3}$$

However, the properties of systems S_1 and S_2 are fundamentally different, since S_1 exhibits an increasing failure rate as opposed to a constant failure (hazard) rate for S_2 (see Figure 3). Here hazard (failure) rate at time t, unlike the $P_f(t)$ time-dependent probability density function, takes into account only the entities that are still operational at that time t [19]:

$$h(t) = \frac{1}{1 - P_f(t)} \frac{dP_f(t)}{dt}. \tag{2.4}$$

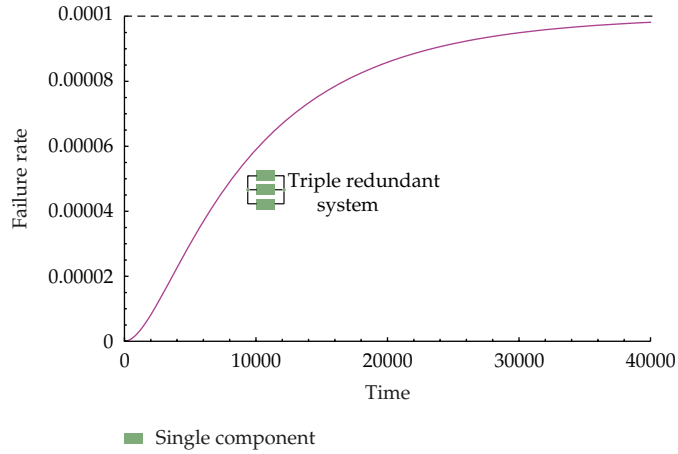

Figure 3: Failure (hazard) rate for a triple-redundant nonrepairable system.

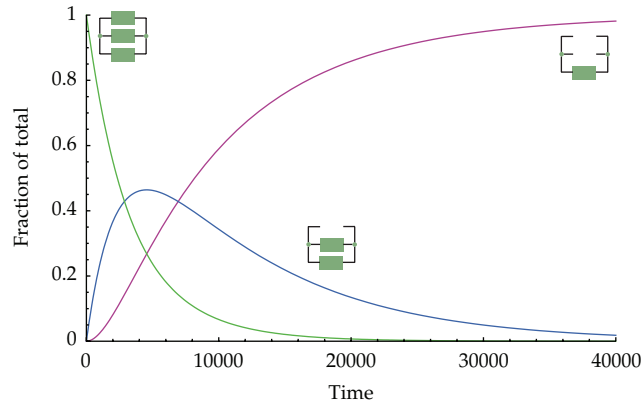

Figure 4: Relative likelihoods of different states (right) for a triple-redundant nonrepairable system.

In the considered example, the hazard rate can be easily calculated in closed form:

$$h(t) = \frac{3\lambda \exp[-\lambda t]\left(1 - \exp[-\lambda t]\right)^2}{1 - \left(1 - \exp[-\lambda t]\right)^3}. \tag{2.5}$$

Since we assume that there are no repairs, as time progresses, the hazard (failure) rate will increase, indicating progressively better odds that redundancy is degraded. In fact, at some point the redundancy will be fully degraded (see Figure 4), and as a result the failure of the system will be converging to that of a single component (see Figure 3). Let us recall that the constant failure rate implies that there is no benefit in repairing or inspecting the system S_2. On the other hand, one can observe that the situation with the triple-redundant system is fundamentally different since the benefits of repairs are substantial. Let us assume that the system undergoes periodic inspections during which failed components are immediately repaired as new, or replaced (i.e., the system is restored to its initial state). The resulting

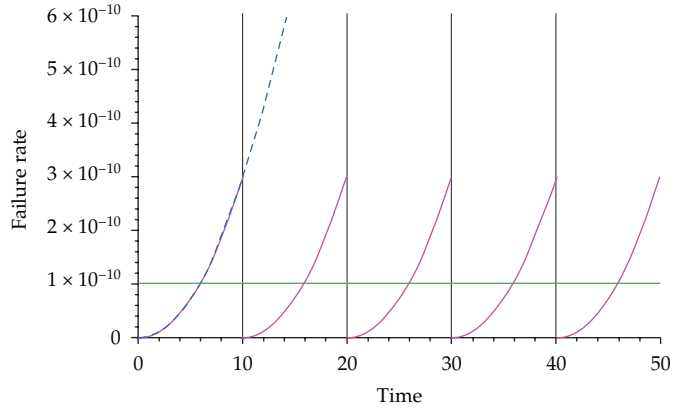

Figure 5: Failure rate for a system with renewals every 10 hours: the blue dashed curve corresponds to the failure rate without repairs, and the horizontal green line represents the equivalent (i.e., average) failure rate.

process is usually referred to as a renewal process [19]. (A comprehensive treatment of the renewal processes in the context of maintenance policies is given in [36].) The effect on the resulting hazard rate is obvious: the shorter the inspection interval, the more reliable the resulting system. The calculation is fairly straightforward since we can divide the life of a system into a number of identical segments (see Figure 5, for the case when inspections occur every $\tau = 10$ hours):

$$n = \frac{T}{\tau} = 1000. \tag{2.6}$$

The total probability of the system failure during its lifetime can be readily evaluated:

$$P_{sf} = 1 - \left(1 - \left(1 - \mathrm{Exp}[-\lambda_c \tau]^3\right)\right)^n \approx 9.985 \times 10^{-7}. \tag{2.7}$$

In fact, in the limit, if we make more and more frequent inspections, the probability of failure will tend to zero (assuming that the failures are independent). It is easier to understand this conceptually if a system is considered where a failed component is instantaneously replaced upon failure. Of course, in practice repairs are not instantaneous, and there is also the possibility of common-cause simultaneous failures that violate the assumption of failure independence. So far the calculations have not required any dynamic capabilities, but they do demonstrate the importance of repairs and inspection in redundancies, as well as the significance of variable hazard rate.

Let us now consider a (practical) situation where repairs are conducted upon failure (with a delay) rather than at fixed predetermined time intervals. In addition, there are usually distinct costs associated with inspection and replacement/repair, so the expected number of the latter needs to be estimated. Those are essentially dynamic effects. Dealing with repairs is one of the most important situations (but not the only situation) where the system changes its configuration at points in time that are not known a priori, and the result depends on when those changes have occurred, necessitating dynamic modeling. Thus, dynamic modeling in system reliability is briefly discussed next.

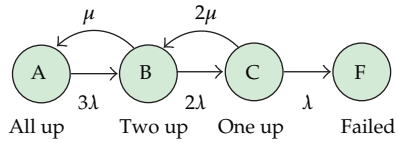

Figure 6: Markov model of a triple-redundant system.

Markov chains (usually referred to a discrete time domain) or processes (where time is continuous) are characterized by a "memory-less" property: the probability of a transition from one system state to another does not depend on the past states of the system [29]. It can be conveniently represented by a state-space graph (referred to as machines or automata in computer science) such as the one depicted in Figure 6, where a simple model for a repairable triple-redundant system is shown. At any given point in time a system is in one of its possible states with a certain probability that can be interpreted as a physical quantity "flowing" from one state to another in accordance with the specified transition rates. The governing system of ordinary differential equations for those probabilities is called Chapman-Kolmogorov equations and can be written in the following matrix form:

$$\frac{dP(t)}{dt} = Q \cdot P(t), \qquad P(t) = \begin{pmatrix} P_{\mathrm{A}}(t) \\ P_{\mathrm{B}}(t) \\ P_{\mathrm{C}}(t) \\ P_{F}(t) \end{pmatrix}, \qquad Q(t) = \begin{pmatrix} -3\lambda & \mu & 0 & 0 \\ 3\lambda & -2\lambda - \mu & 2\mu & 0 \\ 0 & 2\lambda & \lambda - 2\mu & 0 \\ 0 & 0 & \lambda & 0 \end{pmatrix}. \tag{2.8}$$

In the context of system reliability, Markov models are typically homogeneous with respect to time, implying that transition rates between system states are constant. This time independence for transition rates drastically simplifies the analysis of Markov models, as it renders constant the coefficients of the governing system of equations. Two types of analysis can be conducted using the Markov framework: transient and steady-state. The latter is appropriate only for processes that achieve a steady state, such as a renewal processes. The solution to the steady-state problem is obtained by setting the derivative term in (2.8) to zero, and solving the resulting system of algebraic equations (complemented by the condition that the sum of the probabilities of being in each state add up to one). Not only is the steady-state solution easier to find, but also it might provide a more compact description of the underlying process. Indeed, let us note that the system depicted in Figure 6 does not achieve a steady state, since there are no outbound transitions from the failed state F. Such a state is called "absorbing," and the corresponding column in matrix Q consists of zeros (see the right-most column in (2.8)). As a result, the probability of failure constantly increases. From the pragmatic perspective of designing reliable and safe systems, one needs to distinguish between two fundamentally different scenarios.

(1) The failure is latent (not detectable at the moment). The main quantity of interest is the probability of failure that will increase with time. This quantity is important for safety-critical components that are not engaged during normal use (which in this context is referred to as the probability of failure on demand).

(2) The failure is immediately detected. In this situation, the main quantity of interest might be the hazard rate (2.5), where the designer would like to assess the chances that the system will fail *knowing* that failure has not yet occurred. For a car driver

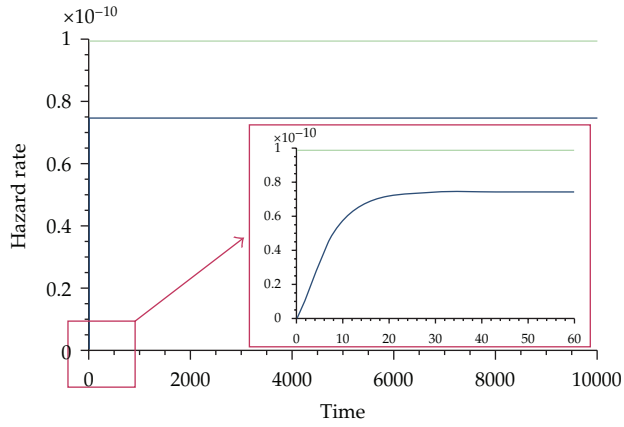

Figure 7: Probability of failure for triple-redundant repairable system (results of transient Markov analysis).

moving at the speed of 80 miles per hour in the left lane of a highway, who might worry about the reliability of the car's engine, that would be a relevant metric. Figure 7 shows this quantity as a function of time for the considered triple-redundant system.

The advantages of failure detection are obvious as failure accumulation is avoided. Modern engineering systems are increasingly equipped with sensors and sensor data processing capabilities, and the trend toward a more detectable nature of failures in the future is clear. This explains the check of air-bag functionality every time a modern car starts. It must be noted that such a check assures only that no air-bag failures occurred prior to the driver starting the car—the failure will not be detected if it takes place en route. Furthermore, the driver might not notice (or might ignore) the warning light, or the warning light itself can fail. One can see that even this simple scenario requires some educated design decisions: is it worthwhile to have a continuous check of the air bag? Should the warning light be bigger or the information be transmitted to the appropriate automechanic directly? How reliable should the warning system be?

Returning to the considered example of a triple-redundant system with immediate detection, one can observe in Figure 7 that the hazard rate is constant after a very short initial transitional period. This constant hazard rate can be calculated by introducing a fictitious transition from the failed state to the original (undegraded) state. It can be shown that the resulting Markov chain model reaches a steady state, while the detailed balance for the failed state allows the calculation of the equivalent hazard rate:

$$\lambda_{\text{eq}} = \frac{\nu P_f}{1 - P_f}. \tag{2.9}$$

It can be further shown that the result does not depend on the rate ν. In fact, by considering limit $\nu \to \infty$ we can conclude that state F is vanishing and consider instead the model Figure 8(c), so that $\lambda_{\text{eq}} = \lambda P_C$ (this can be observed by a simple algebraic manipulation of matrix Q). Another method for calculating the equivalent hazard rate is based on

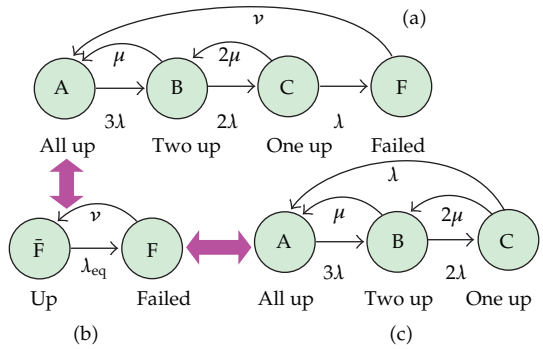

Figure 8: Equivalent steady-state Markov model of a triple-redundant system.

the fundamental Perron-Frobenius theorem of linear algebra [37], which states that, for a matrix with positive entries, there is a unique largest real eigenvalue with strictly positive components of the eigenvector. While matrix Q is not positive, we can add appropriate diagonal terms (that will not obscure the calculation of the eigenproblem) and then apply the Perron-Frobenius theorem (or, more precisely, its extension to the so-called primitive matrices [38]). The result is that, for matrix Q, we know the two largest eigenvalues: the first one is zero, and the second one is a $-k$ (negative and real). Therefore, as time goes to infinity, those two dominate the solution of the Kolmogorov-Chapman equation for the failed state. As a result, the hazard rate is simply k (the absolute value of the smallest negative eigenvalue of Q). The discrete-time analog of this result was obtained in [39]. While the considered example is very simple, it can be readily generalized.

Importantly, from the design perspective we can conclude that the triple-redundant system will behave like a single component, but the equivalent failure rate depends on the specified repair policy. Steady-state solutions for Markov state-space models possess an important property that greatly facilitates hierarchical model construction, thus providing a means of dealing with system complexity: an equivalent transition rate is fully sufficient for representing a group of components: for example, once the equivalent failure rate is found for the triple-redundant system, it can be used in a higher-level model as if it were a single component.

This simple example demonstrates both challenges and several important strategies that can help avoiding complex system failures that are described next in more general terms.

3. Existing Approaches for Making Systems Reliable

3.1. *Simple Is Beautiful*

The complexity of a system is related to the amount of information needed to describe the system (following the general definition of Kolmogorov complexity expressed as entropy [40]), and, specifically, to the size of the state space representing distinct states of the system. The simplest proxy for this parameter is the number of components that comprise the system. Effectively, the use of "part count" as a measure of system reliability implies that the system is designed so that the failure of any of its components results in the failure of the whole system. Under these assumptions, the reliability of the system is simply the product of the reliability of the individual components [17]. As discussed below, redundancy

alters this relationship. More sophisticated measures of complexity take into account not only the number of entities and the size of the state space describing those entities, but also some measures of the amount of interrelationships among those entities (couplings). Specifically, graph-theoretical considerations can be useful by representing the system entities as nodes and connecting some pairs of nodes with either unidirectional or bidirectional links to represent the couplings. In this setting, several measures of complexity related to couplings can be introduced, including a measure based on branching diversity for graphs that can be represented as trees (no cycles) or a collection of trees (forests) [41]. Perhaps the most commonly used among these measures is cyclomatic complexity (related to the number of linearly independent loops) [42].

We can note the axiomatic design description that explicitly accounts for complexity [43]; therein, complexity is related to information defined specifically as the probability of success in achieving a set of specific functions or functional requirements. Extending this concept to explicit inclusion of functional requirements related to system safety and reliability effectively leads to risk-based design. Risk-based design has been extensively studied in the context of various complex engineering systems, including ship design [44], water supply systems [45], packaging for hazardous chemicals [46], and earthquake-resistant design of nuclear plants [47]. The popularity of applications of risk-based design clearly indicates the practical need for an explicit inclusion of risk-related goals into design set requirements. However, this takes place at a fairly detailed level of design, where either the design architecture is fixed, or several alternatives are evaluated within well-defined risk models. Similarly, reliability-based optimization implies the existence of risk-related objectives that are a known function of design parameters and usually obtained based on physics-based structural reliability models [48].

3.2. Modularity: A Measure of Coupling

From the safety perspective, as discussed in the seminal (and controversial) work by Perrow [49], a high degree of complexity and level of coupling inevitably lead to accidents in complex systems by inducing a conflict between centralized and decentralized modes of control. Coupling is often described as the number of links characterizing the dependence of an entity upon other units within a system. From the safety perspective, coupling can also be measured in terms of the time required for the disturbances to propagate along those links, since the propagation time is related to the time available for mitigation purposes. In some (but not all) circumstances it can be assumed that those two definitions are related, as a large number of links are expected to speed up the disturbance propagation.

Modularity and the related principles of encapsulation and information hiding in object-oriented programming provide an effective means of reducing complexity by deliberately designing the system's architecture in a hierarchical fashion and minimizing coupling. However, efficiency and other practical considerations often lead to a situation where a single functionality is supported by a very large number of components, and this common purpose, as well as reliance on common resources, provides inevitable coupling mechanisms. As a result, for such "open" systems it is impossible to group the components into modules that can be independently analyzed.

At the same time, there is compelling evidence that in both natural and engineering domains complex systems are unlikely to be fully coupled, as modular architecture provides clear advantages in developing desirable systems properties. The evolutionary advantage of

the so-called nearly decomposable systems has been demonstrated for biological systems [50], while similar processes were identified in the history of steam engine development [51]. These concepts are also explicitly employed in the design principles of computer systems [52] (including structured design [53]). In this context the importance of the so-called "weak links" (Another recent example of the utility of weak links is related to the resistance to disturbances of financial networks [54]: while a fully connected network is the most resistant to small disturbances, as a certain threshold of disturbances is crossed, the dynamics of the system undergoes a "phase transition," and the fully connected network loses its resistance to disturbances due to the fact that complete interconnectivity provides an effective mechanism for propagation of shocks throughout the system. An intriguing question arises as to which network configuration provides the best resistance of those larger disturbances. For a single shock, such a configuration consists of weakly connected modules, each consisting of two connected nodes. Extending this result to multiple shocks remains an open research question, but it is reasonable to expect that the effectiveness of modular (near decomposable) structures in shock resistance will persist.) has been explored in various domains, including biological systems [55]. The implications of the modular structure are twofold: on the one hand, it is wise to avoid tightly coupled systems at the design phase, and it is also quite logical to take full advantage of the modular structure of the systems in modeling system failures.

While in some situations a fully decoupled modeling of each unit is possible, coupling mechanisms can significantly impact the results. In order to address this issue, mean-field modeling of coupling provides a feasible direction [56]: instead of modeling pair-wise coupling, an aggregate representation of the effect of multiple couplings is used. For this strategy to succeed, the time distribution associated with the combined coupling must be represented accurately, which implies a smart selection of the type of parametric distribution and its parameters if parametric distributions are used. Couplings that involve competing risks (or more precisely, a race of several competing events, as those events do not need to be negative: e.g., when the first available repair crew is utilized) are the most challenging in terms of compact representation, because the mean (expected) rate is not sufficient to characterize the distribution. Usually, such additional characterization is provided by a measure of variability (e.g., variance, or standard deviation). However, in the context of competing risks, the so-called winning ratio (the fraction of races where the event of interest occurs first, given that both events occur within the interval of interest) is shown to be a more important parameter for determining the accuracy of the distribution representation. For a given interval, this winning ratio increases if events are relatively more likely to occur in the beginning of the interval. The Weibull distribution provides the needed flexibility for matching any desired winning ratio by adjusting the shape parameter (the smaller the shape parameter, the larger the winning ratio).

The resulting strategy involves the evaluation of individual ratios that correspond to the pair-wise couplings, and assembling them into an equivalent combined ratio using an analytical formula. The final step consists of finding parameters of Weibull distributions that match the equivalent combined winning ratio along with the mean rate. An order-of-magnitude improvement in accuracy is observed when using this procedure as compared to matching the mean rate alone [56]. However, those are only first steps on the road of accounting for system-level effects in a compact fashion by building accurate component models.

3.3. Defense in Depth

Redundancy is a fundamental principle of design for reliability [57], recognized since the time of Von Braun and implemented under various names: "no single-point failure" (in aerospace), "damage tolerance" (in structures), and "defense-in-depth" (nuclear plants) [4]. First, we note that redundancy increases system complexity and has the potential to introduce new failure modes (Galileo's Dialogue Concerning Two New Sciences [58], published in 1638, provides a classical example of this pitfall in context: a large marble column was laid out and supported by two pieces of wooden beam at each end; fearing that the column might break under its own weight in the middle, a mechanic decided to add an additional support in the middle of the column, only to find out that one of the end beam supports decayed and sunk, so that the column was effectively supported only by the other two supports, and the column broke. In fact, instead of reducing the maximum bending moment, and therefore maximum normal stress in the column by 86%, as one would expect from solving the "ideal problem," the stress actually remains unchanged. This misalignment problem would not appear if only two supports were used (redundancy is avoided), and given the fact that the deformation of the marble column is minimal, misalignment is practically inevitable, so in the other scenario the middle support would be simply useless (and the maximum stress unchanged again). It is interesting to note that Gere and Timoshenkos classical text book on Mechanics of Materials [59] contains a problem that asks for finding the optimal location of two supporting columns, the solution to which probably constitutes an alternative solution to Galileo's mechanic by shifting two columns symmetrically toward the middle by $L(\sqrt{2}-1)/2 \approx 0.207L$, thus reducing the stress by 83%, which is pretty close to the unattainable "ideal" solution with three supports, using only two supports, and is quite practical!) and also alters the relationship between reliability and operational costs. If the reliability of the system is driven by component reliability, then lower reliability implies more frequent maintenance (and increased demand for spare parts), leading to increased operational cost. However, this relationship between reliability and maintenance costs can be reversed if redundancy is used to improve system reliability, as component failures do not result in system failures, yet require maintenance actions. Redundancy provides the most common means of ensuring high degree of reliability in safety-critical systems. Effective estimation of the reliability of redundant repairable systems requires an understanding of maintenance policies as well as of common cause failures (when the failures of redundant components are not independent).

If a system requires continuous operation, the so-called "hot" or "warm" spare configurations are usually used, where the spare components are in operation all the time (not only when they are needed) and are therefore subject to failure as well. As a result, for such configurations the demand for maintenance actions increases, which leads to pragmatic maintenance policies that allow for temporary operation of the system with degraded redundancy as long as the overall system reliability meets the requirements. In aeronautical applications this is referred to as limited-time dispatch, while in electronics "deferred maintenance" name is used [39, 60–62]. This implies that not only is the functionality of complex systems viewed as service in the external relationship to the customer, but also the reliability of the system as a whole, as well as of its individual components, should be viewed as a package that includes the intrinsic properties of an entity, and also a service (maintenance) policy that is assigned to this entity.

The drawbacks of redundancy are reduced in the case of network systems, where link redundancy and the so-called mirrored redundancy provide clearer benefits. The latter type

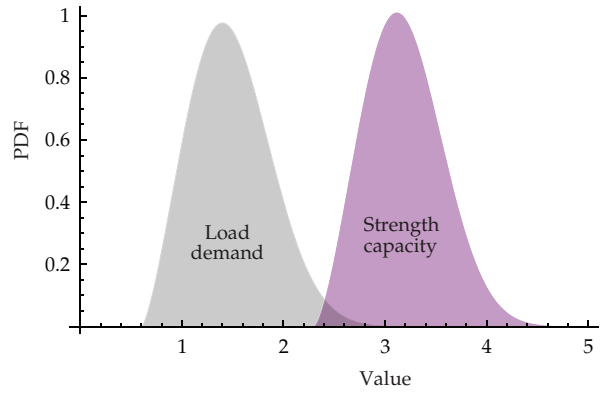

Figure 9: Load (demand)-strength (supply) relationship.

of redundancy corresponds to situations where the content (e.g., packets of information that need to be delivered) is actually duplicated [63].

3.4. Component Reliability

Revisiting the issue of variability is appropriate here. Plotting the probability distribution of both the load, $l(x)$, and the strength, $s(x)$ (also known in other applications as demand and available resources or supply, resp.) provides visual representation of the chances of failure (see Figure 9). The intersection of the two probability density functions is indicative of the possibility of failure, although, contrary to common belief, this area does not numerically represent the probability of failure. (More, precisely, if the strength and the loads are random variables $\mathcal{S}$ and $\mathcal{L}$, respectively, then using cumulative distribution for the strength $S(x) = \int_0^x s(y)dy$ we can express the chances of failure as $P_f = P(\mathcal{S} < \mathcal{L}) = \int_0^\infty S(y)l(y)dy$. In contrast, the intersection area is given by $\int_0^\infty \min[s(y), l(y)]dy$ and does not have any physical meaning, except that if one quantity is zero, the other must be zero as well.)

One of the simplest methods of increasing component reliability is derating [57], where the components are rated for a higher stress environment than that they experience in operation. Effectively, derating implies increased safety margins when the mean corresponding strength (capacity) is further separated from the load (demand). One can also interpret derating as internal redundancy (as the component has a spare capacity that provides protection against variability). Another means of reducing the probability of failure is by decreasing the variance of each of the distributions, in particular that of the strength, which provides one of the motivations for statistical quality control [64] and similar concepts. Indeed, improving manufacturing tolerances and reducing other variability during the manufacturing process leads directly to the reduction of the variability of the strength distribution, and indirectly (e.g., via the influence on the surrounding components) on the load variability. In structural applications, the distance between the nominal values of strength and load can be associated with safety factors [65]. Qualitatively, it is clear that for a structural component that is subject to degradation, both distributions will move toward each other as time progresses, leading to increased probability of failure. While the leftward movement of the strength curve is attributed to the degradation of the component itself (at the local level, some portion of the component), the rightward movement of the load curve

is caused by the load redistribution due to the degradation of the component environment (or at the local level, the degradation of the adjacent portions of the same component). Quantification of the dynamic relationship between the load and the strength is significantly more challenging and generally not well understood.

3.5. Use of Parametric Distribution as Information Compression

As an example, let us consider stress-rupture failure of composite overwrapped pressure vessels (COPVs) that are used to store fuel in space vehicles and gas in other applications. COPVs are designed to contain gas under the pressure, p_0, which is a fraction of the ultimate pressure that would destroy COPVs. Manufacturing processes for COPVs have significant variability (with one of the main sources being the strength variability of the carbon or graphite fibers used in the composite). In order to screen the vessels for the weak ones, a proof test is used: for a short period of time (minutes), COPVs are tested under pressure that is significantly higher than p_0. This leads to the question "to proof or not to proof": on the one hand, the damaged vessels are eliminated; on the other hand, those that survive might have been weakened by the proof test (*cf.* Heisenberg uncertainty principle). Is the resulting population actually better (i.e., has greater reliability) than the original population? The question can be related to the shape of the time-to-failure distribution: if the population failure rate decreases with time, then proofing makes sense (as the effective age of the system is increased by proofing). It can be observed that the heterogeneity of the population leads to a decrease in failure rate with time (as only the stronger members of the population survive). At the same time, effective redundancy acts in the opposite direction (as redundancy degrades due to random failures at the component level which do not cause system failure). Similarly, degradation mechanisms at the component level also lead to the failure rate increasing with time. Therefore one needs to understand which of the opposing trends dominates.

State-space based models for evaluating system safety and reliability rely on compact representation of state transitions (e.g., failures and repairs, or recoveries from intermittent faults). Parametric distributions are preferred from the compactness perspective, assuming that their accuracy is assured. The question of selecting appropriate distribution might seem obsolete in the modern world where nonparametric representation can be easily stored on a computer; however, selection of the distribution effectively implies infusing the statistical process with physics-based knowledge and significantly reduces the need for experimental data about the system. Successful application of parametric distributions is closely related to taking advantage of the underlying general physical processes, just as the blind use of parametric distributions can lead to serious modeling flaws. The central limit theorem assertion, that the sum of large numbers of independent random variables follows normal distribution, is the best-known case, but in the context of reliability the importance of several types of distributions is similarly clear, as briefly discussed next.

3.5.1. Exponential Distribution

A failure transition with the constant rate λ follows an exponential distribution, whose cumulative form is given by $F_e(t) = 1 - e^{-\lambda t}$. λ is the inverse of the mean time to failure. State-space models with constant transition rates are particularly convenient: first, each transition is fully characterized by a single parameter, λ, and second, the resulting process is Markov (i.e., the chances of transitioning to a new state are fully determined by the current state),

which significantly simplifies the calculations. In the context of repairable systems, steady-state results often depend only on the mean parameters of the distribution, justifying the use of exponential distribution even if the underlying distributions are different (see, e.g., the extensions of the Palm-Khinchin theorem, especially in logistics [66]). An additional argument for using exponential distribution is based on what can be characterized as the central limit theorem for repairable systems: in a complex repairable system with multiple components, failures form a homogeneous Poisson process [19]. This is true, however, if there is no coordination among component failures. In practice, for many systems with clear aging or degradation patterns (e.g., gas turbine engines), major inspections and overhauls impose an overall structure, and within each maintenance cycle the failure rate can vary.

3.5.2. Weibull Distribution

$F_w(t) = 1 - e^{-(t/\theta)^\beta}$ are often used due to their flexibility of representing rates that can be either increasing or decreasing with time. The former correspond to the shape parameter $\beta > 1$ (e.g., failures in deteriorating systems), while the latter correspond to the shape parameter $\beta < 1$. Conveniently, for $\beta = 1$, an Weibull distribution becomes an exponential distribution with the scale parameter θ representing the inverse of the transition rate. An additional reason for using Weibull distribution in system reliability is its relationship to the "weakest link" mode of failure. The Fisher-Tippett-Gnedenko theorem [67, 68] states that for a large number of identically distributed functions, the competing risk (i.e., the minimum of failure times) will converge to one of the three families of extreme value distributions (Weibull, Gumbel, or Fréchet).

3.5.3. Lognormal Distribution

The lognormal model of time to failure is justified when a process moves towards failure based on the cumulative effect of many small "multiplicative" shocks. Specifically, if at any instant in time a degradation process undergoes a small increase in the total amount of degradation that is proportional to the current total amount of degradation, then the time to failure (i.e., reaching a critical amount of degradation) is expected to follow a lognormal distribution [69].

Other distributions can also be applicable, including Gamma distribution [70] and Birnbaum-Saunders distribution, which is often used to model fatigue life [71]. In addition, time shifts can be introduced to a distribution.

4. Promising Directions

4.1. Convergence of Several Trends

Despite a certain maturity of the reliability field with its associated signs of stagnation (including the top-heavy age distribution of both practitioners and academicians), there are several recent developments that hold a promise of fundamentally transforming the field in the near future, possibly creating a perfect scenario for an upcoming renaissance in reliability.

(i) Analytics. At the time of writing this paper, data mining is out and analytics is in, as a quick look at a Google trend search demonstrates. However, regardless of

the most fashionable term de jour, the associated capabilities of collecting, storing, and analyzing large amounts of data to improve business decisions are clearly here to stay. Reliability as a field is yet to fully benefit from this development, but it will. These benefits include obvious improvements to and automation of failure data collection, which provide better input for reliability models and allow verification of the quality of reliability predictions. In addition, one of the promising directions lies at the intersection of analytics and failure modeling and anomaly detection, where large-scale experiments demonstrated the feasibility and usefulness of automatic processing of operational data [72, 73].

(ii) Better and cheaper sensor technologies provide a wealth of new information that can be used to identify the failure precursors in structures [74] and other critical components [75].

(iii) Servitization. The emergence of servitization as a global trend in manufacturing has been advocated since the late 1980s, when manufacturers either added services or integrated services to their core products [76]. From that perspective, operation and maintenance support is an important service to the consumer that provides a natural step up the value chain. Recent empirical studies of servitization indicate that, for large manufacturing companies, engaging in service negatively correlates with higher profits [77]. The correlation should not be confused with causation, as it is quite likely that a willingness to engage in services is indicative of the relative maturation of the particular field, and of saturation of traditional sources of growth. More generally, servitization can be considered a part of a general new (service) dominant logic of marketing that deemphasizes the importance of material things (operand resources, which are subject to external acts and operations and are transformed as a result to create desired effects) and emphasizes the growing importance of operant resources, which perform operations on operands and on each other [78]. There are several distinguishing aspects of the logic and two of them are of particular relevance in the context of increased importance of support for complex systems:

(a) the key role of knowledge in value creation (*cf.* the emergence of the knowledge era [79]), which in the context of operation and support pertains to the know-how about inner workings of the systems (e.g., the technical data package, historical failure data, etc.). This often favors original equipment manufacturers (OEMs), and, in the modern environment, system integrators as well, as they are the ones who have the "big picture" of the maintenance processes;

(b) the prevalence of long-lasting relationships instead of single transactions. The important thing in this context is the cost of switching to a competitor: as long as the benefits of switching do not outweigh those costs, the customer has incentives to stay with the current service provider, thus ensuring a long-term stream of revenues. From this perspective, Performance Based Contracts (PBCs) provide a mechanism to assess those costs and benefits and facilitate the transfer to the service provider of the financial risk associated with the operation and support of the system.

This shift toward providing services rather than products can have a large impact on the field of reliability. Traditionally, the financial burden of operating a system falls onto

the buyer, creating a moral hazard problem for the OEMs: spares and associated services often provide a high-margin source of revenue. The cost of consumables used in system operation is relatively well known at the point of sale and therefore can contribute to the competitiveness of the product (e.g., fuel-efficient cars that become more popular as the price of petroleum goes up). In contrast, predicting maintenance costs, especially those related to corrective maintenance (i.e., correcting system failure), requires substantial historical data on the reliability of the system. For mass-produced consumer products, such historical data accumulates sufficiently quickly for reliability to often have a significant impact on the buyer's choice (e.g., the American preference for Japanese cars over domestically produced ones during the 1980s). As a result, the market provides the necessary incentives for improving reliability and lowering the maintenance costs (as exhibited by the ability of American car manufacturers to close the reliability gap with their Japanese competitors in recent years). However, more complex engineering systems are either tailor-made or developed in relatively small quantities, often precluding the accumulation of sufficient historical evidence to impact buyer choices. Proliferation of PBCs fundamentally changes this balance and strongly encourages the OEMs to improve reliability of the supported systems. While there is empirical evidence [80] that even for existing systems PBCs lead to reliability improvements, even bigger increases in reliability can be realized when the direct incentives brought by PBCs impact the design phase.

The availability of enabling technologies combined with the emergence of powerful incentives provides fertile ground for improvements to the reliability of complex systems in the near future. Next, promising research directions that can capitalize on these new trends are discussed.

4.2. Software Reliability

Since, strictly speaking, software is not a subject to any variability for a fixed set of inputs, the "only" problem is to ensure proper behavior for any credible set of inputs. However, the effective dimensionality of inputs is so large that even for relatively simple cases the exhaustive test of all possible combinations of inputs is not practical. Traditionally, reliability at the system level is assessed by probabilistic tools (such as FTA or RBD) that require probabilistic estimates of component failure. While a probabilistic framework provides a natural means for quantification of hardware failures, its applicability to software failures is far from straightforward. Software response is fundamentally deterministic, so in a certain sense it either works or it does not. Theoretically, there are two ways that can provide a probabilistic estimate of software errors.

(i) Uncertainty propagation method: analogous to uncertainty propagation used in structural reliability [81] and other physics-based methods (such as computational fluid dynamics). Therein, probabilistic representation of the uncertainty about the inputs to the model is mapped into the probabilistic representation about the outputs, and the corresponding mapping is deterministic. The method would consist of the following steps: conduct a (random) relatively diverse subset of tests on a software module; estimate the relative exposure to the diversity of inputs during the testing phase in comparison to the operation phase; and finally infer the possible number of errors during operation based on the number of errors uncovered during the test phase. Unfortunately, an objective comparison of the profiles of inputs for the test and operation is challenging at best. In fact,

the more errors are uncovered by testing for the "known unknowns" (e.g., using fault insertion techniques), the more biased the test sample becomes; the errors in the test sample become less and less likely, as the known unknowns are resolved and the corresponding errors are eliminated; at the same time the "unknown unknowns" (i.e., the failure modes due to the unforeseen combination of inputs) are likely to remain uncovered. In a certain way, the problem is analogous to an overfitting problem [82] when an overly flexible approximation model can fit the training data very well, while the errors for the real data (that was not used in the model construction) can be quite poor indeed.

(ii) Software classification based on historical data: for safety-critical software there are detailed process-based specifications [83] describing the requirements that commensurate with the consequences of the software failures, and assigning appropriate "design assurance levels" (DAL) from A for catastrophic consequences, to E for no consequences. The mapping between DAL and error rates can be estimated based on historical data (possibly taking into account some other parameters, such as software complexity, measured either in the lines of code, or some other metrics). However, such classification is unlikely to provide enough resolution to be practically useful, as there are so many parameters that need to be taken into consideration, and there is very little empirical data that would support such error rate estimates.

As a result, neither of the two approaches appears to be satisfactory at the moment, and an alternative paradigm of using formal methods to make safety claims about safety-critical software is advocated by many experts in the field [2, 84]. The field of formal verification techniques is slowly but surely transitioning from an obscure obsession of few computer scientists into a mainstream practical area of research and development [85]. Two important stepping stones for this transition can be identified [2].

(i) A transition from the so-called "weak" (or descriptive) formal methods that focus only on specifications, to strong (operational) formal methods that rely on tool-based semantics analysis.

(ii) A shift from "heavyweight" formal methods that require special tools and highly complex skills to "lightweight," fully automated analysis embedded in the standard development environment for software engineers.

4.3. Modeling an Automation-Human Interface

Initially, the impact of interaction on reliability between humans and machines was viewed from the point of view of human errors. Effectively, those errors were classified into categories, including errors of omission (the human operator fails to perform an action when needed) and errors of commission (the operator performs an action when it was not needed/expected). The likelihood of each type of error can be estimated based on the complexity of the task at hand as well as on the environmental conditions (the so-called performance-shaping factors, or PSFs) [86]. From this perspective, the presence of errors is fundamental and the influence of the environment is secondary [23].

As the field evolved, more emphasis on the importance of the environment led to a fuller appreciation of organizational issues [4], as well as to the realization that focusing on human errors does not provide a complete picture of the role humans play in operating

complex engineering systems. In particular, the positive contribution of humans to safety has been recognized as well [87]. More generally, the variability of human behavior came to the attention of experts in human factors. Instead of perceiving humans as entities making perfectly rational decisions that are occasionally interspersed with errors, an alternative interpretation of human behavior has been developed. It recognizes that human behavior relies on approximations in making decisions [88] and that often the quality of these approximations are related to the time available to make the decisions. This process is referred to as the Efficiency Thoroughness Trade-Off (ETTO) [89] by Hollnagel. Importantly, these approximations are not random, but are systematic, as are the systematic biases studied by Kahneman and his colleagues [90]. From this perspective engineering systems exhibit complex dynamic behavior where small variability in human response can result either in normal operation, or in failures by means of positive feedback (*cf.* second cybernetics [91]). This sensitivity to the initial conditions of complex dynamic systems is well known [11] and will be discussed next. Importantly, finite state machines (FSM) [92, 93] can be used to represent the most common operation modes of the system, providing a state-space framework that can be related to system reliability models, thus leveraging the insights obtained in studying complex systems in general.

4.4. Networks and Complexity Science

A significant body of work has accumulated in recent years in the area of large-scale networks, and specifically, their robustness to failure. The key aspect of this work is its reliance on understanding coarse-grained (macro)models that focus on aggregate metrics characterizing the system without the need for precise description of each individual entity that comprises the system. Effectively, the quest is for finding the network equivalent of thermodynamics laws that would help to predict network behavior, so perhaps it is not surprising that the researchers involved in this quest are mainly physicists [94]. Unfortunately, most of this work has been mainly ignored by mainstream reliability researchers (with some notable exceptions [95]). In stark contrast to classical thermodynamics, where averaging provides the mechanism for establishing macroproperties (e.g., the averaging of the kinetic energy of individual molecules leads to temperature), the coarse-grained characterization of systems failure dynamics relies on finding extreme values among the collection of individual entities that comprise the system, leading to so-called self-organized criticality [96, 97] (*cf.* the weakest link principle; see the discussion of Weibull distributions above).

The prevalence of certain network topologies in nature, including those that follow a power law distribution of links (so-called scale-free networks), is one of the most exciting discoveries made in the late last century [94]. The long list of networks that fall under this category includes Internet, neural networks, power grids, and various transportation networks (although the worldwide air transportation network possesses some peculiarities [98]). The relative susceptibility of scale-free networks with respect to various failures in comparison to deliberately designed topologies such as "highly optimized topologies" (HOTs) [99] provides important insights into the nature of failures for networked systems. In particular, it is important to note the lack of robustness of HOT with respect to conditions that are significantly different from the ones that the networks were originally designed for.

The initial focus of studies dealing with network failures was on static random failures (e.g., nodal removal). Gradually, dynamic failure scenarios attracted more attention, including studies of capacity constraints and propagating failures as a result of the shared

load [100], applications to power grids [101], similar phenomena in aviation [102], and congestion in networks [103]. Recently, the characterization of networks explicitly based on their dynamic properties has been introduced as well [104, 105]. Specifically, a consistent ordering of states related to their efficiency in serving as sinks or sources of disturbances in the networks has been developed. This ordering relies on tail distribution of the hitting times associated with each state.

While it is clear that network models (where nodes and links are distinguished) are relevant to the understanding of failures in complex systems, even simpler architecture that effectively consists of nodes only (with links implied by geometric proximity) can also provide sufficient flexibility while being simple enough to establish relevant statistical properties of failure patterns. Cellular Automata (CA) [106, 107] and especially sandpile models [108] and associated concepts of self-criticality can prove to be quite useful in this regard. One of the promising directions of the research in this field is to facilitate classification of failure dynamics, and in particular identify distinct patterns of failure propagation as functions of the input parameters and "tipping points," as well as the most efficient ways of delaying the occurrences of those tipping points, or preventing them altogether. However, traditional CA models assume that the behavior of individual cells is purely local, while in real failure modes, global variations of the load (shock models [109]) are of significant importance. In addition, memory of past states (accumulated damage in a given cell) might be critical for understanding failure progression (*cf.* recent research on the use of cell memory [110]).

In recent years there have been multiple applications of CA to provide detailed domain-specific models, including the durability of concrete in aggressive environments [111], multipit corrosion [112], wind damage in forest planning [113], rock failures [114], and creep rupture [115]. Among the relevant general research in CA, connections to self-organized critical behavior models used to model landslides, forest fires, and earthquakes [116] must be noted, as well as models that extend the notions of damage in CA, such as the introduction of damaging agents [117]. Those and similar resources can be used to map the properties of the CA to specific domains. To this end, relevant detailed damage propagation models (not CA-based) can be utilized as well, including the work on semicrystalline polymer fiber [118] and models of composite damage propagation [119]. In general, CA is mostly concerned with steady-state patterns (e.g., in terms of the failure propagation, only the averaged property, for example, expected transient time, is usually assessed). In contrast, specific shapes of distributions of time to failure need to be studied more, thus relating the study to existing statistical reliability models [120, 121].

4.5. Dynamic Reliability Modeling

4.5.1. Why Static System Reliability Modeling Is Not Enough

Modern complex systems are capable of reacting to the changes to both external environments and the internal states of systems. The system's reaction to those changes may be autonomous or it may rely on external supervision. The associated control logic may consist of a simple policy of replacement of the system or its component at specified time, but it might also include reconfiguring the system (switching to the backup component, changing the regime of operation, and trying again if the previous attempt to deliver the functionality has not succeed). If those changes have impact on the reliability of the system, they need

to be captured in the reliability analysis. The more complex the control logic is, the more challenging the corresponding reliability modeling is.

For example, in order to evaluate the impact of condition-based maintenance policy, all possible outcomes and their likelihoods need to be accounted for. Even for a single entity (component) considered in isolation, this implies simultaneous assessment of the influence of several important factors, including the rates of missed defects and false alarms, as well as the delay time between the detection of the condition that requires a maintenance action, and the action itself. As the timing of changes to the system is not known *a priori*, static system reliability tools that rely on Boolean logic (such as fault trees, or reliability block diagrams [17]) are not directly applicable. This has been recognized as a drawback of existing methods of probabilistic risk assessment [122]. Indeed, fault trees rely on Boolean (static) logic and evaluate the probability of the occurrence of relevant events regardless of their relative timing. As a result, when the timing of events and other couplings are important, the likelihood of the resulting events for populating fault trees must be obtained from external sources, usually by means of physics-based simulations.

Physics-based simulations (including agent-based simulations) are increasingly realistic in capturing particular physical phenomena, but the depth of the analysis comes at the expense of its breadth, and so they are limited to only a few relevant interactions. For example, in evaluating the efficiency of sense-and-avoid systems [123], a so-called inner loop includes a dynamic simulation based on Monte Carlo, while the outer loop relies on fault trees. This two-tier approach to system safety modeling relies on the fact that events that impact the behavior of an inner loop are independent from the events that comprise other branches of the system fault tree. Quite often this is not the case, and so the inner loop has to be expanded across several layers of protection from accident. This can lead to very large Monte Carlo simulations [124] that not only involve a large number of entities, but also require large numbers of samples to capture rare events.

4.5.2. Discrete Space Continuous Time Models

Models with finite state space and continuous time avoid some of the drawbacks of static reliability tools (as they can take into account the timing of individual events), and these models often provide an efficient level of resolution. Importantly, analytical results can be obtained for a sufficient subclass of the models (including Markov processes). Broadly speaking, these models can be broken into two types as follows.

Global Models

Each state represents the whole system, explicitly accounting for the possible permutations of the individual states of the system constituents (components). These models scale poorly with the number of components that constitute the system. For example, if a system consists of n distinct components, and each component can be in two states (say, fully operational or with detectable damage), system representation requires 2^n states. Markov chains fall under this category (when state transition rate is fully determined by the current state and therefore is independent of the past), but useful extensions include semi-Markov models, where the time spent in the current state (the so-called holding time) can impact the transition rate.

Local Models

Components' states are explicitly modeled, as well as the intercomponent dependencies, while the system state is inferred from the component states. Continuing with the same example of a system with n components, one can observe that only $2n$ states are required. Stochastic Petri nets (SPNs) [30] provide a graphical formalism for describing such local models. Therein, the so-called tokens (depicted as small circles) are introduced to denote individual components of the system; they can move between the "places" (depicted as large hollow circles) representing possible states of those components. SPNs are a subset of the so-called nonautonomous Petri Nets [125–128] and are of particular relevance to the modeling of system reliability. SPNs introduce delays between the enabling and firing of a transition; those delays are transitions attributes that can be deterministic or can be sampled from a given distribution (stochastic). SPNs are often used as a modeling preprocessor, so the model is internally converted to Markov state space and solved using standard Markov methods [129]. However, a discrete event (e.g., Monte Carlo) simulation can be used to solve SPNs directly [130]. Although the resulting solution is not as fast or precise, with modern computers the importance of this disadvantage is constantly decreasing. On the other hand, the flexibility gained allows arbitrary distributions to be considered.

If, as in the so-called colored Petri nets [131], tokens can have unique identities (labels), an alternative interpretation of firing facilitates the preservation of the information about the system's past states: rather than considering removing a token from the transition's input place and depositing a (different) token to the output place as two disjoint actions, these two actions may be united into a single action of moving the same token from an input place to the output place. Memory (continuously changing labels) can be assigned to tokens (the result is "aging tokens" [132]). Such tokens can move freely throughout the Petri net without losing their memory. Firing delays for timed transitions can be interpreted by associating backward clocks: the clock starts when the transition is enabled, and once the clock reaches zero, the firing takes place. In standard SPNs, this clock is associated solely with the transition, and if more than one token is present in the input place, the token to be fired is selected at random. With aging tokens, a clock can be associated with a token-transition pair, which allows several clocks to run simultaneously for a single transition, and often results in a more compact model. Effective system modeling using SPN involves its decomposition into a set of relevant entities, where each entity does not necessarily represent a physical component of the system, as it might, for example, describe a phase of operation, or environmental condition. Due to their flexibility in terms of modeling both discrete logic and continuous states, SPNs have been considered particularly useful in the context of modeling air transportation systems [133]. An enhanced modeling of continuous parameters using aging tokens provides a unique opportunity for modeling various scenarios.

If the behavior of each component is independent of the states of other components, the dynamic system model is superfluous, as the relevant system properties can be obtained by means of static Boolean operations. The global models do not require any special means to model component interdependencies, since the components are not explicitly represented. As a result of dependency, some of the global states can be absent, reducing the state-space size. Symmetry considerations can also greatly facilitate the state space compression. In contrast, the local models do require special means for modeling interdependence among the components. For example, inhibitors in SPN (arcs originating at a place and terminated at a transition with a hollow circle) are a means to disable the transition if there are sufficient

tokens in the originating place, that is, if disabling the transition of the component due to a certain state of another component or several of them).

4.5.3. Continuous State and Time Models

Quite often the coupling among various entities of the system cannot be reduced to boolean operations and timing of individual events, and several continuous parameters fundamentally impact systems behavior. It can be argued that such systems exhibit a high level of coupling and are therefore inherently prone to failures [49]; however, the existence of such systems necessitates their failure modeling, while the need for control loops might (or might not) justify the systems existence. While some analytical frameworks exist to model such systems, most notably piecewise-deterministic Markov processes [134], most of the time the complexity of the modeling requires resorting to simulations [135].

To this end, the agent-based models provide a means to shift attention from individual systems and entities to their inherent interactions and the environment in which they operate. Many systems are intrinsically, or have further evolved into, large and complex architectures of interoperable parts and players—examples of which can be found in many domains, from complex ecosystems [136] or virtual societies [137, 138] to the global economy [139, 140], or airlines' economic strategies [141] to the system-of-systems concept [142, 143]. Some characteristics of these types of networks are the presence of open boundaries evolving in time, internal heterogeneity, and high quantitative dimensionality. As a consequence, unified or centralized approaches may not be appropriate, as they are better suited to describe closed and well-structured systems. As an alternative, agent-based techniques exploit the idea of distribution by focusing on system constituents and their behavioral rules at the microscopic level, thus allowing the network's dynamics and the components' integration to emerge at the macroscopic level. Improved system-level robustness, adaptability, and self-organization are some of the resulting features that make agents appealing to engineering integration and management of complex infrastructures. In order to accomplish its objective, an agent interacts with other agents and the environment by exhibiting a host of qualities such as reactiveness, proactiveness, sociability, learning, in-time evolution, and others. Interactions and heterogeneity within a system generate the need for communication protocols and schemes to optimally resolve conflicts and/or enhance interagent coordination, for which various solutions have been proposed in the literature.

The National Airspace System (NAS) is characterized by the interaction of various heterogeneous entities, such as aircraft, control towers, or various personnel, all of those entities being spatially distributed. As a result, this domain provides a fertile ground for complex system modeling. The use of agent-based simulations to assess systemic risks was advocated by Blom and his co-workers [144] where a particular runway incursion scenario was investigated, in which an aircraft taxis toward the crossing of an active runway while its crew has inappropriate situation awareness. Besides the intrinsic complexity of such a system, an interesting point raised at the simulation phase is the difference in time scale among the various entities: a physics-based model may require a fine time step to guarantee adequate accuracy as it is continuous in time, while a discrete-event model will need to be updated less frequently. As observed by Lee et al. [145], this issue of different time granularity works against the possibility of asynchronous simulation and forces synchronization, especially in the presence of stochastic events for which events times are not known *a priori*. As an alternative, in order to guarantee consistency of results, asynchronous simulation with partial

resynchronization is suggested, where information and data updates are predicted and occur when necessary.

Disruptions or unforeseen events can cause a series of cascading effects which call for time-critical decisions. Decisions may, however, be hard to agree upon when many competing players are involved. An attempt at modeling such circumstances is offered by Campbell et al. [146] who employed the agent-based model IMPACT (intelligent agent-based model for policy analysis of collaborative traffic flow management) to simulate the decision-making process involving airlines and traffic control authorities in response to weather-due schedule changes. Harper et al. [147] have also conducted similar studies with a focus on the human element in the context of decision making. Pilots, airline dispatchers, and traffic controllers are all modeled using the same agent structure, made up of three units: air-traffic situation assessor, collaborative decision-making element, and plan executor, respectively, in charge of collecting and processing current data, resolving traffic issues, and performing plan changes. The SAMPLE (Situation Assessment Model of Pilot-in-the-Loop Evaluation) agent-based architecture for modeling human behavior has been integrated in the FACET (Future Air Traffic Management Concepts Evaluation Tool) environment, and principled negotiation has been employed as a means to provide coordination and resolve conflicts between aircraft [147], where a solution is sought by providing communal advantages for all the interested parties.

In summary, the following nested hierarchy of models can facilitate the comprehensive assessment of the reliability of complex systems by exploiting information compression and keeping the overall modeling complexity manageable.

(i) Most detailed level: physics-based (including agent-based) simulations that rely on continuous space- and time-state representation and might include human-, software-, and hardware-in-the-loop simulations of specific scenarios.

(ii) Intermediate level: stochastic Petri nets or analogous discrete-event simulations that capture the timing of events, but provide discrete state-space representation,

(iii) Static level: static evaluation of failure scenarios, relying on Boolean algebra (Fault Trees, Reliability Block Diagrams), or providing qualitative description of the failure state space (Failure Mode and Effect Analysis). At this level both spatial and time states are explicitly enumerated and discrete,

(iv) Coarse-grained level: in the presence of large numbers of interacting entities, aggregate properties of the system are assessed based on (usually asymptotic) considerations related to self-organized criticality and similar concepts.

5. Conclusions

Modern engineering systems exhibit complex dynamic behavior, and their failures cannot be adequately described by traditional reliability tools. Accounting for the interactions among individual modules of a system requires understanding the specifics of hardware, software, and "human-ware" domains, and the ability to use a common framework for all these domains. Modern business trends emphasize selling the functionalities of engineering systems as services rather than products and, as a result, provide new incentives for lowering life-cycle costs and increasing system reliability. Sophisticated sensors combined with analytics and associated capabilities for collecting, storing, and processing large amounts of data provide practical mechanisms for the implementation of sound business strategies that

respond to those incentives. As a result, increasingly sophisticated maintenance strategies for the design and operation of complex engineering systems are being developed [148] which rely on a better understanding of failure modes of the system. At the same time, the field of reliability and the safety of complex systems have not reached the maturity of "normal" design; as discussed in the context of software reliability, the successful transition from radical to normal design requires a dedicated ecosystem or community of professionals (both academicians and practitioners) [2]. At the moment such a united community does not exist, although fragments of this ecosystem emerge in different domains (including complexity, computer and material science, solid mechanics, controls, human factors, and econometrics). In this paper an attempt has been made to outline the boundaries of this emerging ecosystem. Importantly, the skills of this new community are likely to be distinct from the skills in the reliability field of the twentieth century. In this information age, failures, large and small, will be increasingly scrutinized; thus, the vital task of this emerging community to learn from past failures and avoid future failures cannot be overestimated.

References

[1] W. G. Vincenti, *What Engineers Know and How They Know It: Analytical Studies from Aeronautical History*, The Johns Hopkins University Press, 1993.

[2] M. Hinchey, M. Jackson, P. Cousot, B. Cook, J. P. Bowen, and T. Margaria, "Software engineering and formal methods," *Communications of the ACM*, vol. 51, no. 9, pp. 54–59, 2008.

[3] F. M. Hocker, *Vasa: A Swedish Warship*, Medströms Bokförlag, Stockholm, Sweden, 2011.

[4] J. Reason, *Managing the Risks of Organizational Accidents*, Ashgate, Brookfield, 1997.

[5] T. Harford, *Adapt: Why Success Always Starts with Failure*, Farrar, Straus and Giroux, New York, NY, USA, 2011.

[6] D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*, University of Chicago Press, Chicago, Ill, USA, 1996.

[7] S. G. Vick, *Degrees of Belief: Subjective Probability and Engineering Judgment*, American Society of Civil Engineers, 2002.

[8] "In-flight breakup over the atlantic ocean Trans World Airlines flight 800 Boeing 747-131, n93119 near East Moriches, New York," Aircraft Accident Report PB2000-910403, NTSB/AAR-00/03, DCA96MA070, National Transportation Safety Board, 2000.

[9] "Space shuttle orbiter reaction jet driver (RJD): independent technical assessment/inspection (ITA/I) report," Tech. Rep. NASA/TM-2005-213750, NASA Engineering and Safety Center, 2005.

[10] H. Petroski, "History and failure," *American Scientist*, vol. 80, no. 6, pp. 523–526, 1992.

[11] P. Érdi, *Complexity Explained*, Springer, Berlin, Germany, 2008.

[12] "Minimum design loads for buildings and other structures," no. ASCE-7-05, American Society of Civil Engineers, 2006.

[13] J. H. Saleh and K. Marais, "Highlights from the early (and pre-) history of reliability engineering," *Reliability Engineering and System Safety*, vol. 91, no. 2, pp. 249–256, 2006.

[14] E. Zio, "Reliability engineering: old problems and new challenges," *Reliability Engineering and System Safety*, vol. 94, no. 2, pp. 125–141, 2009.

[15] E. F. Moore and C. E. Shannon, "Reliable circuits using less reliable relays," *Journal of the Franklin Institute*, vol. 262, no. 3, pp. 191–208, 1956.

[16] J. von Neumann, "Probabilistic logics and the synthesis of reliable organisms from unreliable components," in *Annals of Mathematics Studies*, no. 34, pp. 43–98, Princeton University Press, Princeton, NJ, USA, 1956.

[17] M. Rausand and A. Høyland, *System Reliability Theory. Models, Statistical Methods, and Applications*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 2nd edition, 2004.

[18] A. Birolini, *Reliability Engineering: Theory and Practice*, Springer, Berlin, Germany, 5th edition, 2007.

[19] R. E. Barlow and F. Proschan, *Mathematical Theory of Reliability*, John Wiley & Sons, New York, NY, USA, 1965.
[20] A. Avižienis, J. C. Laprie, and B. Randell, "Fundamental concepts of dependability," Tech. Rep. 1145, Laboratoire d'Analyse et d'Architecture des Systèmes (LAAS), Centre National de la Recherche Scientifique (CNRS), 2001.
[21] B. P. Zeigler, *Object Oriented Simulation with Hierarchical Modular Models*, Academic Press, New York, NY, USA, 1990.
[22] "Failure modes, effects and criticality analysis (FMECA) for command, control, communications, computer, intelligence, surveillance, and reconnaissance (C4ISR)," Tech. Rep. TM 5-698-4, Department of the Army, 2006.
[23] E. Hollnagel, *FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*, Ashgate Publishing, Farnham, UK, 2012.
[24] "The Columbia Accident Investigation Board (CAIB) final report," 2003.
[25] "Probabilistic risk assessment procedures guide for nasa managers and practitioners, Second edition," Tech. Rep. SP-2011-3421, NASA, 2011.
[26] W. B. Nelson, *Accelerated Testing: Statistical Models, Test Plans, and Data Analysis*, Wiley Series in Probability and Statistics, Wiley-Interscience, New York, NY, USA, 1990.
[27] W. Q. Meeker and L. A. Escobar, *Statistical Methods for Reliability Data*, Wiley Series in Probability and Statistics, John Wiley & Sons, New York, NY, USA, 1998.
[28] V. Bagdonavicius and M. Nikulin, *Accelerated Life Models: Modeling and Statistical Analysis*, Monographs on Statistics & Applied Probability, Chapman and Hall/CRC, Boca Raton, Fla, USA, 2001.
[29] S. M. Ross, *Stochastic Processes*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1996.
[30] P. J. Haas, *Stochastic Petri Nets. Modelling, Stability, Simulation*, Springer Series in Operations Research, Springer, New York, NY, USA, 2002.
[31] G. C. Fraccone, V. Volovoi, A. E. Colón, and M. Blake, "Novel air traffic procedures: investigation of off-nominal scenarios and potential hazards," *Journal of Aircraft*, vol. 48, no. 1, pp. 127–140, 2011.
[32] W. R. Blischke and D. N. P. Murthy, *Reliability. Modeling, Prediction, and Optimization*, John Wiley and Sons, New York, NY, USA, 2000.
[33] V. Volovoi, "Quantification of system-level business effects of IVHM," in *IVHM the Business Case*, I. Jennions, Ed., chapter 5, SAE, 2012.
[34] W. A. Thompson Jr., "On the foundations of reliability," *Technometrics*, vol. 23, no. 1, pp. 1–13, 1981.
[35] S. Blumenthal, "New approximations for the event count distribution for superimposed renewal processes at the time origin with application to the reliability of new series systems," *Operations Research*, vol. 41, no. 2, p. 409, 1993.
[36] T. Nakagawa, *Maintenance Theory of Reliability*, Springer, London, UK, 2005.
[37] C. Meyer, *Matrix Analysis and Applied Linear Algebra*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa, USA, 2000.
[38] M. Kijima, *Markov Processes for Stochastic Modeling*, Stochastic Modeling Series, Chapman & Hall, London, UK, 1997.
[39] M. Boussemart, T. Bickard, and N. Limnios, "Markov decision processes with a constraint on the asymptotic failure rate," *Methodology and Computing in Applied Probability*, vol. 3, no. 2, pp. 199–214, 2001.
[40] D. Braha, A. A. Minai, and Y. Bar-Yam, Eds., *Complex Engineered Systems: Science Meets Technology*, Springer, Berlin, Germany, 2006.
[41] B. A. Huberman and T. Hogg, "Complexity and adaptation," *Physica D*, vol. 22, no. 1–3, pp. 376–384, 1986.
[42] T. J. McCabe, "A complexity measure," *Institute of Electrical and Electronics Engineers*, vol. SE-2, no. 4, pp. 308–320, 1976.
[43] N. P. Suh, "Theory of complexity, periodicity and the design axioms," *Research in Engineering Design*, vol. 11, no. 2, pp. 116–131, 1999.
[44] E. Vanem and R. Skjong, "Designing for safety in passenger ships utilizing advanced evacuation analyses—a risk based approach," *Safety Science*, vol. 44, no. 2, pp. 111–135, 2006.
[45] P. Latinopoulos, N. Mylopoulos, and Y. Mylopoulos, "Risk-based decision analysis in the design of water supply projects," *Water Resources Management*, vol. 11, no. 4, pp. 263–281, 1997.
[46] R. A. Freeman and J. W. Kleffner, "Risk based package design," *Process Safety Progress*, vol. 16, no. 1, pp. 14–17, 1997.
[47] R. P. Kennedy, "Risk based seismic design criteria," *Nuclear Engineering and Design*, vol. 192, no. 2, pp. 117–135, 1999.

[48] R. V. G. Seung-Kyum Choi and R. A. Canfield, *Reliability-Based Structural Design*, Springer, London, UK, 2007.
[49] C. Perrow, *Normal Accidents: Living with High-Risk Technologies*, Princeton University Press, Princeton, NJ, USA, 1999.
[50] H. A. Simon, "Near decomposability and the speed of evolution," *Industrial and Corporate Change*, vol. 11, no. 3, pp. 587–599, 2002.
[51] K. Frenken and A. Nuvolari, "The early development of the steam engine: an evolutionary interpretation using complexity theory," *Industrial and Corporate Change*, vol. 13, no. 2, pp. 419–450, 2004.
[52] P.-J. Courtois, *Decomposability: Queueing and Computer System Applications*, ACM Monograph Series, Academic Press, New York, NY, USA, 1977.
[53] W. P. Stevens, G. J. Myers, and L. L. Constantine, "Structured design," *IBM Systems Journal*, vol. 13, no. 2, pp. 115–139, 1974.
[54] A. Tahbaz-Salehi, D. Acemoglu, and A. Ozdaglar, "Systemic risk and stability in financial networks," in *Proceedings of the Annual Meeting Phoenix (INFORMS '12)*, INFORMS, Phoenix, Ariz, USA, 2012.
[55] P. Csermely, *Weak Links: The Universal Key to the Stability of Networks and Complex Systems*, The Frontiers Collection, Springer, 2006.
[56] V. Volovoi and R. V. Vega, "On compact modeling of coupling effects in maintenance processes of complex systems," *International Journal of Engineering Science*, vol. 51, pp. 193–210, 2012.
[57] D. Crowe and A. Feinberg, Eds., *Design for Reliability*, CRC Press, Boca Raton, Fla, USA, 2001.
[58] G. Galilei, *Dialogues Concerning Two New Sciences*, MacMillan Company, New York, NY, USA, 1914.
[59] J. M. Gere and S. P. Timoshenko, *Mechanics of Materials*, Pws Publication, Boston, Mass, USA, 4th edition, 1997.
[60] M. Boussemart, N. Limnios, and J. C. Fillion, "Non-ergodic Markov decision processes with a constraint on the asymptotic failure rate: general class of policies," *Stochastic Models*, vol. 18, no. 1, pp. 173–191, 2002.
[61] M. Boussemart and N. Limnios, "Markov decision processes with asymptotic average failure rate constraint," *Communications in Statistics*, vol. 33, no. 7, pp. 1689–1714, 2004.
[62] D. Prescott and J. D. Andrews, "A comparison of modelling approaches for the time-limited dispatch (TLD) of aircraft," *Proceedings of the Institution of Mechanical Engineers O*, vol. 220, no. 1, pp. 9–20, 2006.
[63] C. Perrow, *The Next Catastrophe: The Next Catastrophe: Reducing Our Vulnerabilities to Natural, Industrial, and Terrorist Disasters*, Princeton University Press, Princeton, NJ, USA, 2007.
[64] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, New York, NY, USA, 1931.
[65] H. O. Madsen, S. Krenk, and N. C. Lind, *Methods of Structural Safety*, Dover Publications, Mineola, NY, USA, 2006.
[66] M. J. Carrillo, "Extensions of Palm's theorem: a review," *Management Science*, vol. 37, no. 6, pp. 739–744, 1991.
[67] R. Fisher and L. H. C. Tippett, "Limiting forms of the frequency distribution of the largest and smallest member of a sample," *Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 180–190, 1928.
[68] B. Gnedenko, "Sur la distribution limite du terme maximum d'une série aléatoire," *Annals of Mathematics*, vol. 44, pp. 423–453, 1943.
[69] A. N. Kolmogorov, "On the log-normal distribution of particles sizes during break-up process," *The Proceedings of the USSR Academy of Sciences*, vol. 31, no. 2, pp. 99–101, 1941.
[70] J. M. van Noortwijk, "A survey of the application of gamma processes in maintenance," *Reliability Engineering and System Safety*, vol. 94, no. 1, pp. 2–21, 2009.
[71] Z. W. Birnbaum and S. C. Saunders, "A new family of life distributions," *Journal of Applied Probability*, vol. 6, pp. 319–327, 1969.
[72] A. N. Srivastava and J. Han, *Machine Learning and Knowledge Discovery for Engineering Systems Health Management*, Chapman & Hall/CRC, 2011.
[73] E. Chu, D. Gorinevsky, and S. Boyd, "Scalable statistical monitoring of fleet data," in *Proceedings IFAC World Congress*, pp. 13227–13232, Milan, Italy, August 2011.
[74] F. K. Chang, Ed., *Structural Health Monitoring. The Demands and Challenges*, SRC Press, Boca Raton, Fla, USA, 2002.

[75] A. Tiwari, P. Ballal, and F. L. Lewis, "Energy-efficient wireless sensor network design and implementation for condition-based maintenance," *ACM Transactions on Sensor Networks*, vol. 3, no. 1, Article ID 1210670, 2007.

[76] S. Vandermerwe and J. Rada, "Servitization of business: adding value by adding services," *European Management Journal*, vol. 6, no. 4, pp. 314–324, 1988.

[77] A. Neely, "Exploring the financial consequences of the servitization of manufacturing," *Operations Management Research*, vol. 1, no. 2, pp. 103–118, 2008.

[78] S. L. Vargo and R. F. Lusch, "Evolving to a new dominant logic for marketing," *Journal of Marketing*, vol. 68, no. 1, pp. 1–17, 2004.

[79] M. Uhl-Bien, R. Marion, and B. McKelvey, "Complexity Leadership Theory: shifting leadership from the industrial age to the knowledge era," *Leadership Quarterly*, vol. 18, no. 4, pp. 298–318, 2007.

[80] J. A. Guajardo, M. A. Cohen, S. H. Kim, and S. Netessine, "Impact of performance-based contracting on product reliability: an empirical analysis," *Management Science*, vol. 58, no. 5, pp. 961–979, 2012.

[81] R. E. Melchers, *Structural Reliability: Analysis and Prediction*, Ellis Horwood Series in Civil Engineering, Ellis Horwood, Chichester, UK, 1987.

[82] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, Springer, New York, NY, USA, 2nd edition, 2009.

[83] "Software Considerations in Airborne Systems and Equipment Certification," Tech. Rep. DO-178C, Radio Technical Commission for Aeronautics (RTCA), 2011.

[84] D. A. Peled, *Software Reliability Methods*, Springer, 2001.

[85] Formal Methods Supplement to DO-178C and DO-278A, no. DO-333, Radio Technical Commission for Aeronautics (RTCA), 2011.

[86] A. Swain and H. Guttman, Handbook of Human Reliability Analysis with Emphasis on Nuclear Power Plant Applications, no. NUREG/CR-1278, US Nuclear Regulatory Commission, 1987.

[87] J. Reason, *The Human Contribution. Unsafe Acts, Accidents and Heroic Recoveries*, Ashgate Publishing, Great Britain, UK, 2008.

[88] H. A. Simon, "Bounded rationality and organizational learning," *Organization Science*, vol. 2, no. 1, pp. 125–134, 1991.

[89] E. Hollnagel, *The ETTO Principle: Efficiency-Thoroughness Trade-Off*, Ashgate Publishing, Great Britain, UK, 2009.

[90] D. Kahneman, P. Slovic, and A. Tversky, Eds., *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, New York, NY, USA, 1982.

[91] M. Maruyama, "The second cybernetics: deviation-amplifying mutual causal processes," *American Scientist*, vol. 44, pp. 164–179, 1963.

[92] D. Javaux and E. Olivier, "Assessing and understanding pilots knowledge of mode transitions on the A340-200/300," in *Proceedings of the International Conference on Human-Computer Interaction in Aeronautics (HCI-Aero '00)*, pp. 81–86, 2000.

[93] D. Javaux, "A method for predicting errors when interacting with finite state systems. How implicit learning shapes the user's knowledge of a system," *Reliability Engineering and System Safety*, vol. 75, no. 2, pp. 147–165, 2002.

[94] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports A*, vol. 424, no. 4-5, pp. 175–308, 2006.

[95] E. Zio, "From complexity science to reliability efficiency: a new way of looking at complex network systems and critical infrastructures," *International Journal of Critical Infrastructures*, vol. 3, no. 3-4, pp. 488–508, 2007.

[96] M. Paczuski, S. Maslov, and P. Bak, "Avalanche dynamics in evolution, growth, and depinning models," *Physical Review E*, vol. 53, no. 1, pp. 414–443, 1996.

[97] S. I. Zaitsev, "Robin Hood as self-organized criticality," *Physica A*, vol. 189, no. 3-4, pp. 411–416, 1992.

[98] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, "The worldwide air transportation network: anomalous centrality, community structure, and cities' global roles," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 22, pp. 7794–7799, 2005.

[99] J. M. Carlson and J. Doyle, "Highly optimized tolerance: robustness and design in complex systems," *Physical Review Letters*, vol. 84, no. 11, pp. 2529–2532, 2000.

[100] A. E. Motter, "Cascade control and defense in complex networks," *Physical Review Letters*, vol. 93, no. 9, Article ID 98701, p. 1, 2004.
[101] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the North American power grid," *European Physical Journal B*, vol. 46, no. 1, pp. 101–107, 2005.
[102] L. Lacasa, M. Cea, and M. Zanin, "Jamming transition in air transportation networks," *Physica A*, vol. 388, no. 18, pp. 3948–3954, 2009.
[103] D. De Martino, L. Dall'Asta, G. Bianconi, and M. Marsili, "Congestion phenomena on complex networks," *Physical Review E*, vol. 79, no. 1, Article ID 015101, 2009.
[104] V. S. Afraimovich and L. A. Bunimovich, "Which hole is leaking the most: a topological approach to study open systems," *Nonlinearity*, vol. 23, no. 3, pp. 643–656, 2010.
[105] Y. Bakhtin and L. Bunimovich, "The optimal sink and the best source in a Markov chain," *Journal of Statistical Physics*, vol. 143, no. 5, pp. 943–954, 2011.
[106] J. von Neumann, "The general and logical theory of automata," in *Cerebral Mechanisms in Behavior—The Hixon Symposium*, pp. 32–41, John Wiley & Sons, New York, NY, USA, 1951.
[107] S. Wolfram, "Statistical mechanics of cellular automata," *Reviews of Modern Physics*, vol. 55, no. 3, pp. 601–644, 1983.
[108] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," *Physical Review A*, vol. 38, no. 1, pp. 364–374, 1988.
[109] T. Nakagawa, *Shock and Damage Models in Reliability Theory*, Springer, London, UK, 2007.
[110] R. Alonso-Sanz and J. P. Cárdenas, "Effect of memory on Boolean Networks with disordered dynamics: the K = 4 case," *International Journal of Modern Physics C*, vol. 18, no. 8, pp. 1313–1327, 2007.
[111] F. Biondini, F. Bontempi, D. M. Frangopol, and P. G. Malerba, "Cellular automata approach to durability analysis of concrete structures in aggressive environments," *Journal of Structural Engineering*, vol. 130, no. 11, pp. 1724–1737, 2004.
[112] R. M. Pidaparti, L. Fang, and M. J. Palakal, "Computational simulation of multi-pit corrosion process in materials," *Computational Materials Science*, vol. 41, no. 3, pp. 255–265, 2008.
[113] H. Zeng, T. Pukkala, H. Peltola, and S. Kellomäki, "Optimization of irregular-grid cellular automata and application in risk management of wind damage in forest planning," *Canadian Journal of Forest Research*, vol. 40, no. 6, pp. 1064–1075, 2010.
[114] D. V. Alekseev and G. A. Kazunina, "Simulation of damage accumulation kinetics with a probabilistic cellular automaton," *Physics of the Solid State*, vol. 48, no. 2, pp. 272–278, 2006.
[115] M. Chrzanowski and K. Nowak, "Cellular automata in damage mechanics: creep rupture case," *Archives of Mechanics*, vol. 59, no. 4-5, pp. 329–339, 2007.
[116] D. L. Turcotte and B. D. Malamud, "Landslides, forest fires, and earthquakes: examples of self-organized critical behavior," *Physica A*, vol. 340, no. 4, pp. 580–589, 2004.
[117] C. H. A. Ferraz and H. J. Herrmann, "The strange man in random networks of automata," *Physica A*, vol. 387, no. 23, pp. 5689–5695, 2008.
[118] H. M. Taylor, "A model for the failure process of semicrystalline polymer materials under static fatigue," *Probability in the Engineering and Informational Sciences*, vol. 1, no. 2, pp. 133–162, 1987.
[119] S. Mahesh and S. L. Phoenix, "Lifetime distributions for unidirectional fibrous composites under creep-rupture loading," *International Journal of Fracture*, vol. 127, no. 4, pp. 303–360, 2004.
[120] Z. W. Birnbaum and S. C. Saunders, "A probabilistic interpretation of Miner's rule," *SIAM Journal on Applied Mathematics*, vol. 16, no. 3, pp. 637–652, 1968.
[121] A. J. Lemoine and M. L. Wenocur, "On failure modeling," *Naval Research Logistics Quarterly*, vol. 32, no. 3, pp. 497–508, 1985.
[122] P. E. Labeau, C. Smidts, and S. Swaminathan, "Dynamic reliability: towards an integrated platform for probabilistic risk assessment," *Reliability Engineering and System Safety*, vol. 68, no. 3, pp. 219–254, 2000.
[123] Sense and avoid for Unmanned Aircraft Systems, "Final report of FAA sponsored sense and avoid workshop," Tech. Rep., Federal Aviation Administration, 2009.
[124] D. M. Blum, D. Thipphavong, T. L. Rentas, Y. He, X. Wang, and M. E. Pate-Cornell, "Safety analysis of the advanced airspace concept using monte carlo simulation," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference*, Ontario, Canada, 2010.
[125] R. David and H. Alla, *Discrete, Continuous, and Hybrid Petri Nets*, Springer, Berlin, Germany, 2005.
[126] F. J. W. Symons, *Modelling and analysis of communication protocols using numerical petri nets [Ph.D. thesis]*, Department of Electrical Engineering Science University of Essex, Essex, UK, 1978.

[127] S. Natkin, *Les réseaux de petri stochastiques et leur application a l'evaluation des systemes informatiques [Ph.D. thesis]*, Conservatoure National des Arts et Metier, Paris, France, 1980.
[128] M. K. Molloy, *On the integration of delay and throughput measures in distributed processing models [Ph.D. thesis]*, Department of Computer Science, University of California, Los Angeles, Calif, USA, 1981.
[129] S. K. Trivedi, *Probability and Statistics with Reliability, Queuing and Computer Science Applications*, John Wiley and Sons, 2nd edition, 2002.
[130] Y. Dutuit, E. Châtelet, J. P. Signoret, and P. Thomas, "Dependability modelling and evaluation by using stochastic Petri nets: application to two test cases," *Reliability Engineering and System Safety*, vol. 55, no. 2, pp. 117–124, 1997.
[131] K. Jensen, *Coloured Petri Nets. Basic Concepts, Analysis Methods and Practical Use*, vol. 1, Springer, Berlin, Germany, 1993.
[132] V. Volovoi, "Modeling of system reliability Petri nets with aging tokens," *Reliability Engineering and System Safety*, vol. 84, no. 2, pp. 149–161, 2004.
[133] H. A. P. Blom, B. K. Obbink, and G. J. Bakker, "Safety risk simulation of an airborne self separation concept of operation," in *Proceedings of the 7th AIAA Aviation Technology, Integration, and Operations Conference*, pp. 331–339, Belfast, Northern Ireland, September 2007.
[134] M. H. A. Davis, "Piecewise-deterministic Markov processes: a general class of nondiffusion stochastic models," *Journal of the Royal Statistical Society B*, vol. 46, no. 3, pp. 353–388, 1984.
[135] F. Brissaud, C. Smidts, A. Barros, and C. Bérenguer, "Dynamic reliability of digital-based transmitters," *Reliability Engineering and System Safety*, vol. 96, no. 7, pp. 793–813, 2011.
[136] E. Bonabeau and C. Meyer, "Swarm intelligence. A whole new way to think about business," *Harvard Business Review*, vol. 79, no. 5, pp. 107–114, 2001.
[137] J. Epstein and R. Axtell, *Growing Artificial Societies: Social Science from Bottom Up*, Brookings Institution Press, Washington, DC, USA, 1951.
[138] M. T. Parker, "What is Ascape and why should you care?" *Journal of Artificial Societies and Social Simulation*, vol. 4, no. 1, 2001.
[139] Icosystem Corporation, 2009, http://www.icosystem.com/.
[140] E. Bonabeau, "Agent-based modeling: methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 3, pp. 7280–7287, 2002.
[141] W. Niedringhaus, "An agent-based model of the airline industry," Tech. Rep., The Mitre Corporation, McLean, Va, USA, 2000.
[142] D. DeLaurentis, J. H. Lewe, and D. Schrage, "Abstraction and modeling hypothesis for future transportation architectures," in *Proceedings of the AIAA/ICAS International Air and Space Symposium and Exposition: The Next 100 Years (AIAA '03)*, pp. 2003–2514, Dayton, OH, USA, 2003.
[143] D. DeLaurentis and R. K. Callaway, "A system-of-systems perspective for public policy decisions," *Review of Policy Research*, vol. 21, no. 6, pp. 829–837, 2004.
[144] S. H. Stroeve, H. A. P. Blom, and G. J. (Bert) Bakker, "Systemic accident risk assessment in air traffic by Monte Carlo simulation," *Safety Science*, vol. 47, no. 2, pp. 238–249, 2009.
[145] S. Lee, A. Pritchett, and D. Goldsman, "Hybrid agent-based simulation for analyzing the National Airspace System," in *Proceedings of the 33nd Winter Simulation Conference*, pp. 1029–1036, Arlington, Va, USA, December 2001.
[146] K. Campbell, W. Cooper, D. Greenbaum, and L. Wojcik, "Modeling distributed human decision-making in traffic flow management operations," in *Proceedings of the 3rd USA/Europe Air Traffic Management Research and Development Seminar*, Naples, Italy, 2000.
[147] K. Harper, S. Guarino, A. White, M. Hanson, K. Bilimoria, and D. Mulfinger, "An agent-based approach to aircraft conflict resolution with constraints," in *Proceedings of the AIAA Guidance, Navigation, and Control Conference and Exhibit (AIAA '02)*, pp. 2002–4552, Monterey, Calif, USA, 2002.
[148] I. K. Jennions, Ed., *Integrated Vehicle Health Management: Perspectives on an Emerging Field*, Society of Automotive Engineers (SAE), 2011.