Abstract: The objective of this collaborative-research award is to develop integrated quality and reliability models and analysis tools that provide fundamental insights for the successful development and commercialization of evolving technologies, such as Micro-Electro-Mechanical Systems (MEMS) and biomedical implant devices. This annual report summarizes the accomplished work since September 1, 2013, which contains six papers that have been published or submitted. For the past one year, we mainly work on task 2, task 3 and task 4. Probabilistic models have been investigated in Task 2 for predicting reliability of devices that experience multiple dependent failure processes. For Task 3, reliability models have been developed for complex systems with multiple independent or dependent components. Case studies are provided to demonstrate each model. In Task 4, multiple user objectives (e.g., quality, reliability, yield, cost, etc.) will be simultaneously considered by the formulation and solution of multi-objective optimization problems.

1. Introduction: The research objective of this award is to develop integrated quality and reliability models and analysis tools that provide fundamental insights for the successful development and commercialization of evolving technologies, such as Micro-Electro-Mechanical Systems (MEMS) and biomedical implant devices. For many new and evolving technologies, their continued successful development depends on the concurrent modeling and optimization of process variability and product life, which are inherently linked for many of these technologies. As shown in Figure 1, Task 1 is to develop an integrated quantitative methodology to jointly optimize system quality and reliability. Based on this integrated framework, probabilistic models are being investigated in Task 2 for predicting reliability of devices that experience multiple failure processes due to simultaneous exposure to degradation and shock loads. Competing risk models with dependent failure processes are being developed and extended. In Task 3, reliability models are being developed for complex systems with multiple independent or dependent components. In Task 4 (yet to be initiated), multiple user objectives (e.g., quality, reliability, yield, cost, etc.) will be simultaneously considered by the formulation and solution of multi-objective optimization problems. Finally, the developed models will be validated in Task 5 for MEMS and biomedical implant devices through collaborations with industrial and government partners.

The results of this research will provide fundamental insights that can be transformed to many newer design and manufacturing problems, such as nano-technology that also has unique manufacturing challenges. The integrated methodology can provide timely and effective tools for decision-makers in manufacturing to economically optimize operational decisions for improving reliability, quality and productivity.

This annual report summarizes our accomplished work in six published or submitted articles. The remaining report is organized as follows. Sections 2 summarizes three papers resulting from Task 2: reliability analysis for multiple failure processes. In Section 3, reliability analysis and condition-based maintenance for complex multi-components systems in Task 3 is described. Sections 4 introduces a two-stage stochastic cost-reliability optimization model for multi-component systems subjected to uncertain stress.


2. Reliability Analysis of Multiple Failure Processes:

The issue of multiple failure processes is of particular interest to researchers because it is a critical problem that systems have experienced in the field, and one that could limit further developments and advancements. Task 2 of our project focuses on the reliability of multiple failure processes, which is presented in this section covering three of our recent articles. Section 2.1 presents the results for reliability analysis considering different shock effects on stochastic degradation in dependent failure processes [1]. Section 2.2 studied component reliability of competing risks with generalized mixed shock model [2]. Section 2.3 conducts a condition-based maintenance strategy for repairable deteriorating systems subject to generalized mixed shock model [3].

2.1 Reliability Modeling with Dependent Failure Processes Considering Different Shock Effects on Stochastic Degradation: In this section, we study the reliability of a system experiencing two dependent competing failure processes, in which shocks are categorized into different shock zones. These two failure processes, a stochastic degradation process and a random shock process, are dependent because arriving shocks can cause instantaneous damage on the degradation process. In existing studies, every shock causes an abrupt damage on degradation. However, this may not be the case when shock loads are small and within the tolerance of system resistance. In this model, only shock loads that are larger than a certain level are considered to cause abrupt damages on degradation, which makes this new model realistic and challenging. Shocks are divided into three zones based on their magnitudes: safety zone, damage zone, and fatal zone. The abrupt damages are further modeled using an explicit function of shock load exceedances (differences between load magnitudes and a given threshold). Due to the complexity in modeling these two dependent stochastic failure processes, no closed form of the reliability function can be derived. Monte Carlo importance sampling is used to estimate the system reliability. Finally, two application examples with sensitivity analyses are presented to demonstrate models.

2.1.1 Reliability Modeling:

As shown in Figure 2, we consider a system that experiences two dependent competing failure processes: soft failure due to degradation and hard failure due to random shocks. These two failure processes are dependent, because they are subject to the same random shock process. The system fails when the overall stochastic degradation exceeds the soft failure threshold level \( H \), or when the magnitude of a shock is larger than the hard failure threshold \( W_r \).

![Figure 1: The framework of the research plan](image)

![Figure 2: Relationship of two failure processes simulated for microengine example: (a) Continuous degradation, (b) Overall degradation process, (c) Random shock process](image)
describe the continuous degradation, \( X(t) \), with the probability density function (pdf):

\[
f_{X(t)}(x,t) = \frac{2}{\Gamma(\alpha)} \frac{a^\alpha}{\Gamma(\tau)} \exp(-ax) x^{\alpha-1}, \quad x \geq 0, \quad t \geq 0
\]

where \( a > 0 \) is the scale parameter, and \( v(t) > 0 \) is the shape parameter. The gamma process with positive scale and shape parameters is a continuous-time stochastic process with the following properties:

1. \( X(0) = 0 \) with probability one.
2. \( X(t) - X(0) \sim \text{Ga}(v(t),a) \) for all \( t \geq 0 \), and
3. \( X(t) \) has independent increments.

The expectation and variance of the gamma process are:

\[
E(X(t)) = \frac{v(t)}{a}, \quad \text{Var}(X(t)) = \frac{v(t)}{a^2}.
\]

For some degradation measures such as fatigue crack growth, the expected degradation at time \( t \) often follows a power law empirically [5]

\[
E(X(t)) = \frac{v(t)}{a} = \frac{\tau}{\alpha} t^{\alpha-1} u, \quad \text{where} \quad \alpha, \theta, \text{and} \ u \ \text{are positive constants. When} \ \theta = 1, \ \text{it reduces to a linear degradation path, e.g., wear amount on rubbing surfaces.}
\]

Let \( T_i \) denote the first time the continuous degradation reaches the level \( l \). Then the distribution of a gamma degradation process is

\[
P(T_i \leq t) = P(X(t) \geq l) = \int_{l}^{\infty} f_{X(t)}(x,t)dx = \frac{\Gamma(v(t),lu)}{\Gamma(v(t))}
\]

where \( \Gamma(a,x) = \int_{x}^{\infty} e^{-t} t^{a-1} dt \) is the incomplete gamma function for \( x \geq 0 \) and \( a > 0 \). Gamma function \( \Gamma(a) \) is obtained when we set \( x=0 \), namely \( \Gamma(a) = \Gamma(a,0) \).

### 2.1.1.2 Truncated Generalized Pareto Distribution for Shock Damages

The instantaneous damage size on degradation \( Y_i \) depends on the magnitude of the shock load \( W_i \). We explicitly model \( Y_i \) as a linear function of the exceedances over the threshold, \( D_i \):

\[
Y_i = bD_i = b(W_i - W_0), \quad \text{for} \ W_i > W_0
\]

In the peaks-over-threshold (PoT) method, the estimates of extreme values are based on all values that exceed a threshold, which takes advantage of all useful information leading to higher accuracy.

A generalized Pareto distribution is typically used in the PoT method. We assume that shock loads \( W_i \) follow a generalized Pareto distribution with scale parameter \( \sigma (\sigma > 0) \) and shape parameter \( c \). The pdf of \( W_i \) takes the form [5]

\[
f_{W_i}(w) = \begin{cases} \frac{1}{\sigma} \left( 1 - \frac{cw_i}{\sigma} \right)^{-1-c}, & c < 0, \\ \frac{1}{\sigma} \exp \left( -\frac{w_i}{\sigma} \right), & c = 0. \end{cases}
\]

Therefore, we can get cdf, and the pdf that \( W_i \) larger than \( W_0 \) follows a truncated generalized Pareto distribution:

\[
g_{W_i}(w) = \frac{f_{W_i}(w)}{F(W_i)}, \quad \text{for} \ W_i > W_0.
\]

Because of the linear relationship between \( W_i \) and \( Y_i \) for \( W_i > W_0 \) in Eq. (5), \( Y_i \) also follows a truncated generalized Pareto distribution, and its pdf is:

\[
f_{Y_i}(y) = \frac{1}{b} g_{W_i} \left( \frac{y + bW_i}{b} \right) = \frac{1}{F(W_i) b \sigma} \exp \left( -\frac{y + bW_i}{b \sigma} \right)^{1-c}, \quad c < 0.
\]

We can also obtain cdf accordingly.

#### 2.1.1.3 Decomposition of Homogeneous Poisson Process for Shock Arrivals

We consider random shocks that arrive according to a homogeneous Poisson process (HPP) with a rate \( \lambda \). \( N_i(t) \) and \( N_2(t) \) denote the numbers of shocks arrived in the damage zone and the fatal zone by time \( t \), respectively. For an arriving shock, the probability it falls into the damage zone is \( p_1 = P(W_i < W_1 < W_2) = F_{W_1}(W_2) - F_{W_1}(W_0) \), and the probability it is in the fatal zone is \( p_2 = P(W_i > W_2) = 1 - F_{W_1}(W_2) \). Based on the decomposition of Poisson process (Kao, 1997), we have that the arrival of shocks in the damage zone follows a HPP with a rate \( \lambda p_1 \), and the arrival of shocks in the fatal zone follows a HPP with a rate \( \lambda p_2 \). Accordingly, the arrival of shocks in the safety zone follows a HPP with a rate \( \lambda (1-p_1-p_2) \). We also prove that \( N_1(t) \) and \( N_2(t) \) are independent of each other.

#### 2.1.2 Reliability Analysis of Dependent Failure Processes

In order to keep functioning, the system should experience no fatal shocks and the degradation level should be within its threshold. The reliability of the system by time \( t \) can be derive. If we denote the conditional pdf of \( X(t) \) given time \( t \) and \( N_1(t) = n \) to be \( f_{X_i}(x,|t,n) \), we have

A generalized Pareto distribution is typically used in the PoT method. We assume that shock loads \( W_i \) follow a generalized Pareto distribution with scale parameter \( \sigma (\sigma > 0) \) and shape parameter \( c \). The pdf of \( W_i \) takes the form:

\[
R(t) = \sum_{n=0}^{\infty} \sum_{i=0}^{n} f_{X_i}(x,|t,n) dx e^{-p_1} (\lambda p_1)^{n} e^{-p_2} (\lambda p_2)^0 n! / n! = \sum_{n=0}^{\infty} \sum_{i=0}^{n} f_{X_i}(x,|t,n) dx e^{-p_1(n+p_2)} (\lambda p_2)^0 n! / n!.
\]

where \( p_1 = P(W_0 < W_1 < W_2) = F_{W_1}(W_2) - F_{W_1}(W_0) \) and \( p_2 = P(W_0 > W_2) = 1 - F_{W_1}(W_2) \). By introducing a new set of variables \( Z_n \), \( i=1, \ldots, n \), we can get \( n+1 \) equations with \( n+1 \) unknown variables. Because the transformation is one-to-one, we can solve for \( X(t) \) and \( Y_i \), \( i=1, \ldots, n \), in terms of \( X_i \) and \( Z_i \), \( i=1, \ldots, n \).
Given the number of shocks in damage zone, \( N_i = n \), the joint pdf of \( X \) and \( Z_i \), \( i = 1, \ldots, n \), is

\[
f_{h_X, h_Z}(x, z_1, \ldots, z_n) = f_{X|Z}(x-z_1-z_2-\ldots-z_n, z_1, \ldots, z_n)
\]

where \( J \) is the Jacobian determinant,

\[
J = \begin{vmatrix}
\frac{\partial X}{\partial x} & \frac{\partial X}{\partial z_1} & \cdots & \frac{\partial X}{\partial z_n} \\
\frac{\partial X}{\partial x} & \frac{\partial X}{\partial z_1} & \cdots & \frac{\partial X}{\partial z_n} \\
\cdots & \cdots & \cdots & \cdots \\
\frac{\partial X}{\partial x} & \frac{\partial X}{\partial z_1} & \cdots & \frac{\partial X}{\partial z_n}
\end{vmatrix} = \begin{vmatrix} -1 & 1 & \cdots & 1 \\
1 & -1 & \cdots & 1 \\
\cdots & \cdots & \cdots & \cdots \\
1 & -1 & \cdots & 1
\end{vmatrix} = 1
\]

There is no closed form of this reliability function in Eq. (9). We can use numerical analysis method to find the solution, and Monte Carlo simulation is one of the most effective and efficient numerical analysis methods to solve this problem.

### 2.1.3 Simulation of Reliability Function

We use one of the most commonly used Monte Carlo methods, variance reduction through importance sampling, to estimate the result. For an integral \( A = \int f(x)dx \), the importance sampled Monte Carlo estimate can be written as follows [6]:

\[
\hat{A}_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i), \quad X_i \sim h(x)
\]

where \( h(\cdot) \) is an importance function that mimics the behavior of \( f(\cdot) \) over \( D \), and it is either integrable analytically or can be easily integrated numerically. Also \( h(\cdot) \) should be normalized to have \( \int_0^N h(x)dx = 1 \). The sampling procedure is then altered to generate points distributed according to \( h(\cdot) \) instead of points that are uniformly distributed.

To solve our reliability function in Eq. (13), we first use Monte Carlo importance sampling to estimate \( \hat{f}_{X_i}(x_i|t,n) \) in Eq. (13), where \( Z_i = 1, \ldots, n \), follows the truncated generalized Pareto distribution on \((0, b\sigma/c - W_0)\). By applying Eq. (14) to Eq. (16), the importance sampled Monte Carlo estimate of \( \hat{f}_{X_i}(x_i|t,n) \) is given as:

\[
\hat{f}_{X_i}(x_i|t,n) = \frac{1}{N} \sum_{i=1}^{N} f_{X|Z}(x_i-z_1-z_2-\ldots-z_n)
\]

Next, the integral of the estimated \( \hat{f}_{X_i}(x_i|t,n) \) can be further simplified by exchanging the order of summation and integral:

\[
\int_0^N \hat{f}_{X_i}(x_i|t,n)dx_i = \int_0^N \left( \frac{1}{N} \sum_{i=1}^{N} f_{X|Z}(x_i-z_1-z_2-\ldots-z_n) \right) dx_i
\]

Thus, we can estimate the reliability function in Eq. (9) by substituting the integral of the estimated \( \hat{f}_{X_i}(x_i|t,n) \) in Eq. (16).

### 2.1.4 Simulation of Reliability Function Case Study

To demonstrate our reliability model, two application examples are given in this section: microengine developed at Sandia National Laboratories, and stents implanted in human body.

#### Case 1: Micro-Electro-Mechanical System (MEMS) Devices

The microengine experiences two competing failure processes: soft failures due to the wear degradation and debris from external shocks, and hard failures due to hub fracture.

Parameters in the model and their values are given in Table 1. The stochastic degradation process and random shock process are simulated in Matlab R2010a and shown in Figure 1. Using Monte Carlo simulation, we obtain the plot of the importance sampled Monte Carlo estimate of \( \hat{f}_{X_i}(x_i|t,n) \) in Eq. (15).

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H )</td>
<td>0.00125 ( \text{um} )</td>
</tr>
<tr>
<td>( J )</td>
<td>2.5x10^-4</td>
</tr>
<tr>
<td>( a )</td>
<td>1.2x10^4</td>
</tr>
<tr>
<td>( \theta )</td>
<td>1.0</td>
</tr>
<tr>
<td>( \alpha )</td>
<td>1.02x10^-4</td>
</tr>
<tr>
<td>( c )</td>
<td>0.055</td>
</tr>
<tr>
<td>( b )</td>
<td>0.0004 ( \text{um}/\text{Gpa} )</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.33</td>
</tr>
<tr>
<td>( W_0 )</td>
<td>0.2 ( \text{Gpa} )</td>
</tr>
<tr>
<td>( W_f )</td>
<td>1.2 ( \text{Gpa} )</td>
</tr>
</tbody>
</table>

The estimated \( \hat{f}_{X_i}(x_i|t,n) \) is plotted in Figure 3 at \( t=50,000 \) revolutions under different numbers of exceedance occurrences, \( n \). As shown in the figure, the plot shifts to the right when \( n \) increases, implying that the system is more prone to soft failure when more shock load exceedances arrive. The safety zone threshold \( W_0 \), the soft failure threshold \( H \), and the shock arrival rate, \( \lambda \), have a great impact on the system reliability, and sensitivity analysis is conducted. Figure 4 shows the safety zone threshold \( W_0 \) effect on the system reliability.
Case 2: Stent Devices

Experimental study of stent failure processes indicates that the crack growth of stents is not only the effect of fatigue stress, but also the result of single-event overloads in a form of sudden step increase of crack for certain patients who have excessive activities. The parameters and their values for the stents example are listed in Table 2.

The estimated \( f_{X_s}(x_s, t, n) \) is plotted in Figure 5 at \( t=2\times10^8 \) under different numbers of exceedance occurrences, \( n \). As shown in the figure, the plot shifts to the right when \( n \) increases, meaning that soft failure has a higher chance to occur when more shock load exceedances arrive, similar to Figure 2 in MEMS devices case. Sensitivity analysis are conducted on parameters of interest: the safety zone threshold \( W_0 \), soft failure threshold \( H \), and shock arrival rate \( \lambda \). Figure 6 shows the soft failure threshold \( H \) effect on the system reliability.

<table>
<thead>
<tr>
<th>Table 2. Parameter values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameters</td>
</tr>
<tr>
<td>( H )</td>
</tr>
<tr>
<td>( \lambda )</td>
</tr>
<tr>
<td>( u )</td>
</tr>
<tr>
<td>( \theta )</td>
</tr>
<tr>
<td>( \alpha )</td>
</tr>
<tr>
<td>( c )</td>
</tr>
<tr>
<td>( b )</td>
</tr>
<tr>
<td>( \sigma )</td>
</tr>
<tr>
<td>( W_0 )</td>
</tr>
<tr>
<td>( W_f )</td>
</tr>
</tbody>
</table>

2.2 Reliability Assessment of Competing Risks with Generalized Mixed Shock Model: In this section, we investigates reliability for systems subject to dependent competing risks considering the impact from a new generalized mixed shock model. Two dependent competing risks are soft failure due to a degradation process, and hard failure due to random shocks. The shock process contains fatal shocks that can cause hard failure instantaneously, and nonfatal shocks that impact the system in three different ways concurrently: 1) damage the unit by immediately increasing the degradation level, 2) speed up the deterioration by accelerating the degradation rate, and 3) weaken the unit strength by reducing the hard failure threshold. While the first impact from nonfatal shocks comes from each individual shock, the other two impacts are realized when the condition for a new generalized mixed shock model is satisfied. New generalized mixed shock model includes three classic shock patterns: extreme shock model, \( \delta \)-shock model, and run shock model. According to the proposed generalized mixed shock model, the degradation rate and the hard failure threshold can simultaneously shift multiple times, whenever the condition for one of these three shock patterns is satisfied.

We aim to extend the previous models [7-10] by incorporating multiple sources of dependence between competing risks into a rich reliability model. In the literature, there has been extensive research devoted to three typical shock models: extreme shock model, \( \delta \)-shock model, and run shock model. In the extreme
shock model, a system fails as soon as the magnitude of any shock is greater than a threshold [11]. In the δ-shock model, a system is considered failed when the time lag of two successive shocks is less than a threshold δ [12]. In the run shock model, a system breaks down after a run of n consecutive shocks above a critical level [13].

2.2.1 Reliability Modeling with Generalized Mixed Shock Model

![Diagram](image)

Figure 7: Two dependent competing failure processes: (a) soft failure, (b) hard failure

Figure 7 depicts two dependent failure processes: soft failure due to a degradation process and hard failure due to random shocks. A soft failure occurs when the overall degradation level exceeds a predetermined threshold \( H \). The continuous degradation by time \( t \), \( X(t) \), is monotonically increasing, which is assumed to be a linear path, expressed as \( X(t) = \varphi + \beta t \), where \( \varphi \) is the initial degradation level and \( \beta \) is degradation rate. This linear path function can be applied to a wide range of degradation models.

Hard failure can be caused by fatal shocks that have a magnitude larger than \( D_j \) (\( j = 1, 2, 3 \)). Nonfatal shocks impact the systems in three different ways: sudden increment in the degradation level \( Y \), and the simultaneous transitions in the hard failure threshold \( D \) and the degradation rate \( \beta \), whenever the conditions of the generalized mixed shock model is satisfied. For example, the first transition occurs when the magnitude of the shock arrived at time \( t_2 \) is greater than the critical level for the extreme shock model \( (W_2 > D_2) \), which results in the reduced hard failure threshold from \( D_1 \) to \( D_2 \), and the increased degradation rate from \( \beta_1 \) to \( \beta_2 \).

2.2.1.1 Transition Process

According to the proposed generalized mixed shock model, the degradation rate and the hard failure threshold shift simultaneously, if 1) a shock is above a critical value \( D \) in the extreme shock model, 2) a time lag between two sequential shocks is less than \( \delta \) in the δ-shock model, or 3) a run of \( n \) consecutive shocks are greater than a critical level \( D_1 \) \( (D_1 < D_2) \) in the run shock model. These three classic shock models are competing against each other, and whichever occurs first causes the shift in the hard failure threshold and the degradation rate. Every time that a transition occurs, the condition of the system worsens. In practice, the system can only undergo a limited number of transitions, after which any transition can cause the system to fail. The maximum number of possible transitions for a system, \( L \), is predetermined based on the device characteristics. Some important assumptions regarding the transitions include:

1. If more than one shock pattern takes place at the same time, they do not amplify the effects on the hard failure threshold and the degradation rate.
2. The transition is not a onetime event, and can happen multiple times whenever the condition for the generalized mixed shock model is satisfied.
3. After each transition, prior shocks are not taken into account for the next generalized mixed shock model.
4. The trigger shock count and the transition time are being reset after each transition.

The degradation rate increases from \( \beta_j \) to \( \beta_{j+1} \) when the system experiences the \( j^{th} \) transition for \( j = 1, \ldots, l \), where \( l \) is the total number of transitions actually occurred before failure, and \( l \leq L \). The \( \beta \) representing the degradation rate vector is \( \beta = [\beta_1 \ldots \beta_j \ldots \beta_l] \) where the \( j^{th} \) element \( (\beta_j) \) is the degradation rate after the \((j-1)^{th}\) transition. We define \( \beta_{j+1} = \beta_j + \eta \) where \( \eta \) is an independent and identically distributed (i.i.d.) positive random variable and independent of \( \beta_j \).

Similarly, the hard failure threshold declines from \( D_j \) to \( D_{j+1} \) after the \( j^{th} \) transition for \( j = 1, \ldots, l \). We use \( D \) to represent the hard failure threshold vector as \( D = [D_0 \ldots D_j \ldots D_n] \)

where the \( j^{th} \) element \( (D_j) \) is the hard failure threshold after the \((j-1)^{th}\) transition. We define \( D_{j+1} = D_j - d \) where \( d \) is an i.i.d. random variable that is independent of \( D_j \).

The shock that triggers a transition is called the trigger shock. The time lag between two successive transitions is defined as the transition time. The number of shocks leading to the \( j^{th} \) transition is denoted as \( S_j \) after the last transition, which is an unknown random variable. The time lag between the \((j-1)^{th}\) and the \( j^{th} \) trigger shocks is called the \( j^{th} \) transition time, \( T_j \), which is also a random variable. We use \( T \) to denote the transition time vector including \( T_j \) for \( j = 1, \ldots, l \):

\[
T = [T_1 \ldots T_{i-1} \ldots T_i \ldots T_j \ldots T_n]
\]

where \( T_{i+1} = \sum_{j=1}^{i} T_j \) represents the remaining time after the last transition. After the random transition times are realized, the vector \( T \) transfers to be a vector composed of known values of transition times:

\[
\tau = [\tau_1 \ldots \tau_{i-1} \tau_i \ldots \tau_{i+1}]
\]

2.2.1.2 Shock Process and Hard Failure

Random shocks arrive according to a homogeneous Poisson process with rate \( \lambda \). The total
number of shocks by time $t$, $N(t)$, follows a Poisson distribution:

$$P(N(t) = i) = \frac{e^{-\lambda t} (\lambda t)^i}{i!}, \quad i = 0, 1, 2, \ldots \quad (17)$$

The magnitude of a shock is denoted by $W_i$ for $z = 1, 2, \ldots$, a sequence of i.i.d. non-negative random variables with a common cumulative distribution $F_W(w) = P(W_i < w)$. In this section, the traditional extreme shock model is employed for traumatic failure. It means that the unit fails as soon as the size of any shock is greater than the corresponding hard failure threshold at the time of shock arrival.

The magnitude of a shock is denoted by $W_i$ for $z = 1, 2, \ldots$, a sequence of i.i.d. non-negative random variables with a common cumulative distribution $F_W(w) = P(W_i < w)$. In this section, the traditional extreme shock model is employed for traumatic failure. It means that the unit fails as soon as the size of any shock is greater than the corresponding hard failure threshold at the time of shock arrival. The probability that a unit survives after exposing to a single shock is

$$P(W_i < D_1) = F_W(D_1), \quad \text{for } z = 1, 2, \ldots; j=1, \ldots, or l+1. \quad (18)$$

In order for a unit not to experience hard failure by time $t$, it needs to survive all the shocks by that time. By considering the following two scenarios, the probability of no hard failure by time $t$, $R_0(t)$, can be derived further.

i) No transition occurs by time $t$, or the shock count for the first transition is larger than $N(t)$, $S_1 > N(t) = i$. The magnitude of each shock must be less than the initial hard failure threshold, $D_i$. The probability of surviving against hard failure given that no transition occurs by time $t$ is

$$R_n(t|S_i > N(t) = i) = \left(\prod_{j=1}^{i-1} [F_W(D_j)]^{z_j}\right) = F_W(D_i). \quad (19)$$

ii) At least one transition occurs by time $t$, or the shock count for the first transition is less than or equal to $N(t)$, $S_1 \leq N(t) = i$. The magnitude of each shock must be less than the corresponding hard failure threshold at the time of shock arrival. We use $I_{ji} = i - \sum_{m=1}^{j} S_m$, for $j = 1, \ldots, l$, to denote the number of remaining shocks after the $j^{th}$ transition, which is a random variable. After the value for shock counts are realized, it becomes $I_{ji} = i - \sum_{m=1}^{j} S_m$. The probability of no hard failure given $S_1 \leq N(t) = i$ is

$$R_n(t|S_1 = s_1 \leq N(t) = i) = \sum_{z_1=0}^{S_1} \sum_{z_2=0}^{S_2} \cdots \sum_{z_l=0}^{S_l} \left[ \prod_{j=1}^{l} [F_W(D_{j})]^{z_j}\right] \quad (20)$$

We present the further derivation of $P(S_1=s_1)$, the probability that the $j^{th}$ transition occurs at the $s_j$ shock after the last transition, and $P(S_1 \geq I_{ji})$, the probability that no transition occurs in the remaining shocks after the very last transition.

### 2.2.1.3 Degradation Process and Soft Failure

As aforementioned, we model the degradation process as a linear path, expressed as $X(t) = \varphi + \beta t$, where $\varphi$ is the initial degradation level that is a random variable (due to variability in manufacturing and delivering processes) and $\beta$ is degradation rate that is a random variable and varies from part to part.

The shock process impacts the degradation in two ways: causing a sudden increment to the probability that the unit fails as soon as the size of any shock is greater than the corresponding hard failure threshold at the time of shock arrival.

The shock process impacts the degradation in two ways: causing a sudden increment to the probability that the unit fails as soon as the size of any shock is greater than the corresponding hard failure threshold at the time of shock arrival. We use the damage from each shock by $D_j$ for $z = 1, 2, \ldots$, where $D_j$ is an i.i.d. non-negative random variable. The overall degradation including both continuous degradation and instantaneous damage induced by nonfatal shocks can be expressed as

$$x_i(t) = x_i(t) + x_i(t) = x_i(t) + \sum_{j=1}^{N(t)} Y_j. \quad (21)$$

where $Z(t)$ represents the cumulative damage by nonfatal random shocks.

By considering the following two situations, the probability of no soft failure by time $t$, $R_S(t)$, can be derived further.

i) No transition occurs by time $t$, or the shock count for the first transition is larger than $N(t)$, $S_1 > N(t) = i$. Given no transition by time $t$, the initial degradation rate is fixed and the probability of no soft failure by time $t$ is

$$R_n(t|S_1 > N(t) = i) = P\left(x_i(t) < H|S_1 > N(t) = i\right) = P\left(\varphi + \beta t < H\right) \quad (22)$$

ii) At least one transition occurs by time $t$, or the shock count for the first transition is less than or equal to $N(t)$, $S_1 \leq N(t) = i$. When a system experiences a transition, the degradation rate shifts; therefore, the degradation progresses at an increased degradation rate. The probability of no soft failure given at least one transition and the total number of shocks occurred by time $t$ is (see Appendix B)

$$R_n(t|S_1 = s_1 \leq N(t) = i) = \sum_{z_1} \sum_{z_2} \cdots \sum_{z_l} \left[ \prod_{j=1}^{l} [F_W(D_{j})]^{z_j}\right] \times P(S_{1+i} > I_{ji}) \quad (23)$$

In a homogeneous Poisson process, the time for the $j^{th}$ transition $T_j$, given a value of $S_1=s_j$, follows a gamma distribution with the scale parameter of $s_j$, and the shape parameter of $\beta$:

$$f_{T_j}(t) = \frac{\beta^{s_j}}{(s_j-1)!} t^{s_j-1} e^{-\beta t}, \quad \text{for } j = 1, \ldots, l. \quad (24)$$

The probability that the $j^{th}$ transition occurs at the $s_j$ shock after the last transition, $P(S_1=s_j)$, changes based on which shock pattern occurs. Let $t'_j = t - \sum_{i=1}^{j-1} t_i$ denote the remaining time after the $j^{th}$ transition for $j = 1, \ldots, l$. Then the probability in (23) can be calculated as
\begin{align*}
R_t (S_i = s_i, N(n)) &= \sum_{i=0}^{\infty} \sum_{s_i=0}^{\infty} P(S_i = s_i) \left[ \prod_{i=0}^{\infty} P(S_i = s_i) \right] \\
&= \left[ \int \left( \prod_{i=0}^{\infty} P(S_i = s_i) \right) f_I(t) dt \right] \cdots f_I(t) dt \\
&= \left( \prod_{i=0}^{\infty} P(S_i = s_i) \right) \int f_I(t) dt \cdots f_I(t) dt \\
\end{align*}

(25)

2.2.1.4 System Reliability Modeling

We investigate the reliability model for a subject system to the generalized mixed shock model. Given the number of shocks by time \( t \), the soft failure and hard failure processes are conditionally independent, due to the assumption that the shock damage size \( Y_t \) is independent of the shock load \( W_t \). Considering situations where there is no transition, only one transition, and more than one transition by time \( t \), the reliability at time \( t \) for a system that experiences two dependent competing failure processes:

\[
R(t) = P(\theta + \beta < t) e^{-\theta} + \sum_{s_i=0}^{\infty} \sum_{s_j=0}^{\infty} \left[ \int \left( \prod_{i=0}^{\infty} P(S_i = s_i) \right) f_I(t) dt \right] \cdots f_I(t) dt \\
= \left( \prod_{i=0}^{\infty} P(S_i = s_i) \right) \int f_I(t) dt \cdots f_I(t) dt \\
\]  

(26)

2.2.2 Probability of Trigger Shock Count in the Generalized Mixed Shock Model

The probabilities related to \( S_j \) (the trigger shock count for the \( j \)th transition) in (26) are derived in the following sections.

A transition occurs when the condition is satisfied for at least one of three shock patterns in the generalized mixed shock model: extreme shock model, \( \delta \)-shock model, and run shock model. It is likely that the conditions for two or more shock patterns are satisfied at the same shock. In a transition, the hard failure threshold is assumed to shift to a lower level. If the resulting hard failure threshold \( D_j \) is less than \( D_s \) in the extreme shock model or \( D_l \) in the run shock model, the corresponding shock pattern is not applicable in the generalized mixed shock model anymore. For example, when \( D_l \) is less than \( D_s \), a shock larger than \( D_l \) leads to a hard failure even before it is considered in the extreme shock model for triggering a transition.

We derive the probability of the \( j \)th transition occurring at the \( s_j \)th shock (after the last transition), \( P(S_j = s_j) \), in the following:

1. If the hard failure threshold is greater than the critical level of extreme shock model \( (D_s > D_j > D_l) \), the soft shock model is not applicable. Then we have

\[
P(S_j = s_j) = P(S_j = s_j) + P(S_j = s_j) - P(S_j = s_j) - P(S_j = s_j) \left( \prod_{i=0}^{\infty} P(S_i = s_i) \right) \\
\]  

(28)

3. If the hard failure threshold is less than the critical level of run shock model \( (D_l > D_j) \), only \( \delta \)-shock model is applicable:

\[
P(S_j = s_j) = P(S_j = s_j). \\
\]  

(29)

2.2.3 Probability of Trigger Shock Count

The probabilities for different shock patterns occurring at the \( s_j \)th shock are derived and presented in Table 3. The first column lists different shock patterns that can cause the \( j \)th transition, and the second column presents the probabilities of the corresponding shock patterns.

Table 3: Probability of occurrence of different shock patterns

When the run shock model is involved, we have three subcases based on the relationship between \( s_j \) and \( n \), as presented in Table 1. In general, we defined \( V_{n,j} \) to be the probability that no run of \( n \) consecutive successes (a success is defined as a shock larger than \( D_s \)) occur in a set of \( s_j \) shocks, and \( V_{n,j} \) to be the probability that the first run of \( n \) consecutive successes occurs at the \( s_j \)th shock:

\[
V_{n,j} = \begin{cases} 
1, & \text{if } 0 \leq s_j < n \\
(1 - p^s)^q, & \text{if } s_j = n \\
qV_{n,j-1} + \ldots + qp^{n-1}V_{n,n-1}, & \text{if } s_j > n
\end{cases}
\]  

(30)

where \( p \) is the probability of success, and \( q = 1 - p \). In addition, we have

When \( s_j > n \), to have the first run of \( n \) consecutive successes occurring at the \( s_j \)th trial, the last \( n \) trials must be successes, the \((s_j-n)\)th trial must be a failure, and no consecutive \( n \) successes occur in the first \( s_j-n-1 \) trials.

2014 Annual Report

Grant #: 0970140 and 0969423
2.2.4 Probability of No Transition in Remaining Shocks

When the number of transitions by time \( t \) is \( l \), it implies that there is no transition in the remaining \( i_{l+1} \) shocks after the \( l^{th} \) transition, or the shock count for the \( l+1 \) transition is larger than \( i_{l+1} \). The probability of no transition in the last \( i_{l+1} \) shocks depends on the relationship between the final hard failure threshold \( D_{l+1} \) and the critical levels for extreme shock pattern (\( D_e \)) and run shock pattern (\( D_r \)).

1. If the hard failure threshold after the \( l^{th} \) transition is greater than the critical level of extreme shock model (\( D_{l+1} > D_e > D_r \)), all three shock patterns are applicable. Then the probability of no transition in the last \( i_{l+1} \) shocks is

\[
P(S_{i_{l+1}} > i_{l+1}) = P(S_{i_{l+1}} > i_{l+1} \cap S_{i_{l+1}} > i_{l+1} \cap S_{i_{l+1}} > i_{l+1}) = P(S_{i_{l+1}} > i_{l+1}) P(S_{i_{l+1}} > i_{l+1}) P(S_{i_{l+1}} > i_{l+1}) \tag{31}
\]

2. If the hard failure threshold after the \( l^{th} \) transition is between the critical levels of extreme shock model and run shock model (\( D_e > D_{l+1} > D_r \)), the extreme shock model is not applicable. Then we have

\[
P(S_{i_{l+1}} > i_{l+1}) = P(S_{i_{l+1}} > i_{l+1}) P(S_{i_{l+1}} > i_{l+1}) \tag{32}
\]

3. If the hard failure threshold after the \( l^{th} \) transition is less than the critical level of run shock model (\( D_e > D_{l+1} \)), only \( \delta \)-shock model is applicable:

\[
P(S_{i_{l+1}} > i_{l+1}) = P(S_{i_{l+1}} > i_{l+1}) \tag{33}
\]

The probability of no transition after the \( l^{th} \) transition in the remaining \( i_{l+1} \) shocks, or \( P(S_{i_{l+1}} > i_{l+1}) \), for different scenarios are derived and presented in Table 4.

Table 4: Probability of no transition after the \( l^{th} \) transition

2.2.5 Numerical Example

We implement our new reliability models to the application of micro-engines, because a micro-engine fails due to two competing failure processes. In addition, a micro-engine becomes more susceptible to shocks and the wear volume accumulates faster, after exposure to a certain pattern of shocks or a significantly large shock. The corresponding value for each parameter is presented in Table 5.

For the proposed generalized mixed shock model with increasing degradation rate and decreasing hard failure threshold, the reliability function \( R(t) \) in (26) is plotted in Fig. 8, and sensitivity analysis is performed to measure the effects of the parameters \( D_e, \delta \) and \( D_r \) on the reliability function.

Table 5: Parameter Values

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H )</td>
<td>0.00125( \mu m )</td>
</tr>
<tr>
<td>( D_e )</td>
<td>1.5Gpa</td>
</tr>
<tr>
<td>( \mu_d )</td>
<td>0.2Gpa</td>
</tr>
<tr>
<td>( \sigma_d )</td>
<td>0.01Gpa</td>
</tr>
<tr>
<td>( D_r )</td>
<td>1.4Gpa</td>
</tr>
<tr>
<td>( D_0 )</td>
<td>1.2Gpa</td>
</tr>
<tr>
<td>( \psi )</td>
<td>0</td>
</tr>
<tr>
<td>( \mu_w )</td>
<td>8.4823( \times 10^{-9} \mu m )</td>
</tr>
<tr>
<td>( \sigma_w )</td>
<td>6.0016( \times 10^{-9} \mu m )</td>
</tr>
<tr>
<td>( \mu_Y )</td>
<td>2.4823( \times 10^{-6} \mu m )</td>
</tr>
<tr>
<td>( \sigma_Y )</td>
<td>1.0016( \times 10^{-6} \mu m )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>5( \times 10^{-7} ) revolutions</td>
</tr>
<tr>
<td>( \mu_L )</td>
<td>1.2Gpa</td>
</tr>
<tr>
<td>( \sigma_L )</td>
<td>0.2Gpa</td>
</tr>
<tr>
<td>( \mu_V )</td>
<td>1.0( \times 10^{-3} \mu m )</td>
</tr>
<tr>
<td>( \sigma_V )</td>
<td>2( \times 10^{-3} \mu m )</td>
</tr>
<tr>
<td>( \delta )</td>
<td>2.0( \times 10^{-10} ) revolutions</td>
</tr>
</tbody>
</table>

Fig. 8: Plot of \( R(t) \)

2.3 A Condition-Based Maintenance Strategy for Repairable Deteriorating Systems subject to Generalized Mixed Shock Model:

Based on reliability analysis for a system subject to dependent competing risks of internal degradation and external shocks, we propose a condition-based maintenance policy considering imperfect repair for complex systems. The internal degradation, such as crack growth, erosion or corrosion, can be modeled using a stochastic deterioration process. External shocks occur at random times can be divided into two classes:
(1) **fatal shocks** that can cause the system to fail immediately, if a shock belongs to any of the three classic shock models (i.e., extreme shock model, run shock model, and \( \delta \)-shock model), or the generalized mixed shock model; (2) **non-fatal shocks** that can damage the system by randomly increasing the degradation level. Under the proposed condition-based maintenance policy, the system is inspected at fixed time intervals and a decision for an appropriate maintenance action (i.e., no action, imperfect repair, preventive or corrective replacement) is made based on the actual health condition of the system detected through inspection. The imperfect repair restores the system by lowering the degradation level to a certain level. The objective is to determine the optimal inspection interval that minimizes the expected long average maintenance cost rate. A micro-electro-mechanical system example is used to evaluate the efficiency of developed reliability and condition-based maintenance model.

### 2.3.1 System Failure Modeling

Consider a repairable system whose failure is due to the competing risks of degradation and shocks. Figure 9 illustrates an example of the two dependent failure processes. The first failure is the system experiences a **shock-based failure** when the magnitude of the shock arrived at time \( t_2 \) is greater than the critical level for the extreme shock model \( W_2 > D_1 \). The system experiences the second failure at time \( t_3 \) when \( B_3 \) is less than the threshold \( \delta \) (\( \delta \)-shock model). The third failure is a **degradation-based failure** because the accumulated degradation level exceeds the threshold \( H \). The last failure occurs at time \( t_4 \) as soon as there are two successive shocks with their magnitudes greater than the critical level for the run shock model \( W_3 > D_4 \). Because the system is replaced after each failure, the shock count and time origin have been reset after each failure as shown in Figure 9.

![Figure 9](image_url)

Figure 9: Two dependent competing failure processes: (a) degradation-based failure, (b) shock-based failure

### 2.3.1.1 Degradation-based Failure Modeling

We assume that degradation: \( X(t) = \varphi + \beta t + \varepsilon \), where \( \beta \) is the degradation rate that is a random variable and varies from part to part, and \( \varepsilon \) is the random error term following a normal distribution \( \varepsilon \sim N(0, \sigma^2) \).

When a non-fatal shock arrives, we assume that it damages the system by increasing the degradation level. We denote the damage caused by the \( k^{\text{th}} \) shock by \( Y_k \) for \( k = 1, 2, \ldots \), where \( Y_k \) is an independent and identically distributed (i.i.d.) non-negative random variable.

\[
X_i(t) = X(t) + Z(t) = X(t) + \sum_{i=1}^{N(t)} Y_i,
\]

where \( Z(t) \) represents the cumulative damage by non-fatal random shocks, and \( N(t) \) is the number of shocks by time \( t \). Therefore, the probability of no degradation-based failure by time \( t \) is

\[
P(X_i(t) < H) = \sum_{i=1}^{N(t)} P(X(t) + \sum_{i=1}^{N(t)} Y_i < H | N(t) = i) P(N(t) = i). \tag{35}
\]

We assume that shocks arrive according to a homogeneous Poisson process with a constant rate of \( \lambda \). If we consider \( F_X(x; \varphi, \beta, t) \) to be the cumulative distribution function (cdf) of continuous degradation, \( X(t) \), and \( f_Z(z) \) to be the probability density functions (pdf) of the accumulated damage caused by shocks, \( Z(t) \), then the cdf of \( X_i(t) \), \( F_{X_i}(x; \varphi, \beta, t) \), can be derived using a convolution integral:

\[
f_{X_i}(x; \varphi, \beta, t) = F_{X_i}(x < x) = F_X(x; \varphi, \beta, t) e^{-\lambda t} \sum_{i=0}^{\infty} \left( \int f_Z(x - w; \varphi, \beta, t) f_Z(w; \varphi, \beta, t) dw \right) \frac{e^{-\lambda t}}{i!}. \tag{36}
\]

The model in (36) is general and can accommodate different distributional assumptions. In a specific case, we assume that \( \varphi \) is a constant, \( \beta \) and \( Y_k \) are normally distributed, i.e., \( \beta \sim N(\mu_\beta, \sigma_\beta^2) \) and \( Y_k \sim N(\mu_Y, \sigma_Y^2) \). For this specific case, the resulting total degradation, \( X_i(t) \), follows a normal distribution. Then the probability in (36) can be expressed as

\[
F_{X_i}(H; \varphi, \beta, t) = \sum_{i=0}^{\infty} \frac{H - (\varphi + \mu_\beta + \mu_Y)}{\sqrt{\sigma_\beta^2 + \sigma_Y^2}} \left( \frac{1}{\sqrt{\pi}} \frac{e^{-\lambda t}}{i!} \right). \tag{37}
\]

### 2.3.1.2 Degradation-based Failure Modeling

Only one shock model (e.g., extreme shock model, \( \delta \)-shock model) is typically considered to result in an immediate failure. We introduce a generalized mixed shock model, a combination of three shock models including extreme shock model, \( \delta \)-shock model, and run shock model, as the direct cause for the **shock-based failure**. The system suffers shock-based failure as soon as the condition for one of the three shock models is realized: 1) the magnitude of one shock is above the critical value \( D_r \) (extreme shock model), 2) the time lag between two sequential shocks is less than the threshold \( \delta \) (\( \delta \)-shock model), or 3) a set of \( n \) consecutive shocks with magnitudes that are greater than the critical level \( D_s \) \((D_s < D_r)\) occurs (run shock model). These three classic shock models are competing against each other, and whichever occurs first causes the system to experience shock-based failure.
**Scenario (1):** If the magnitude of the \(i\)th shock is greater than the critical threshold \(D_r\), the system fails according to the extreme shock model, and the \(j\)th shock is the fatal shock. We use \(S_i\) to denote the **fatal shock count** due to the extreme shock model. The probability that the system does not experience extreme shock model by the \(i\)th shock is derived to be

\[
P(S_i > N(t) | N(t) = i) = P\left(\prod_{i=1}^{\infty} |W_i < D_i|\right) = F_{\lambda}(D_i)^{-1} 
\]

(38)

**Scenario (2):** If the time interval between the \((i-1)\)th shock and the \(j\)th shock is less than the threshold \(\delta\), the system breaks down and the \(i\)th shock is considered as the fatal shock. We use \(S_j\) to denote the **fatal shock count** due to the \(\delta\)-shock model, which is a random variable. The probability that the system does not fail according to the \(\delta\)-shock model by the \(i\)th shock is derived to be

\[
P(S_j > N(t) | N(t) = i) = P\left(\prod_{i=1}^{\infty} |B_i > \delta|\right) = e^{-it\delta} 
\]

(39)

where \(B_i\) is the arrival time of the first shock or the time interval between the \((k-1)\)th and \(k\)th shocks.

**Scenario (3):** If the first run of \(n\) consecutive shocks that are greater than the critical level \(D_r\) occurs by the \(i\)th shock, it causes the system to break down according to the run shock model. The \(i\)th shock is considered to be the fatal shock. We use \(S_n\) to denote the **fatal shock count** due to the run shock model, which is a random variable.

In general, if we define \(V_m\) to be the probability that no run of \(m\) consecutive successes (in our case, a success is defined as a shock greater than \(D_r\)) occurs in a set of \(m\) shocks, it has been shown to satisfy the recursive equation (14):

\[
V_m = qV_{m-1} + \ldots + p^{m-1}qV_0, \quad \text{for } m > n, \quad (40)
\]

Based on Eq. (40), the probability that no run of \(n\) consecutive shocks greater than \(D_r\), happens by time \(t\), or the probability that the fatal shock count due to the run shock model, \(S_n\), is larger than \(N(t)\), is derived to be

\[
P(S_n > N(t) | N(t) = i) = P(W_{i-1} < D_i)P(S_n > N(t) | N(t) = i) + \ldots
\]

\[
+ P(W_{n-m+1} < D_i)P(S_n > N(t) | N(t) = i) + \ldots
\]

\[
+ P(W_{n-1} < D_i)P(S_n > N(t) | N(t) = i).
\]

(41)

For a system not to experience a **shock-based failure**, the condition for the above three scenarios should not be met. Therefore, the probability that a system does not experience the **shock-based failure** by time \(t\), i.e., the fatal shock count is greater than \(N(t)\), is derived as follows:

\[
P(S > N(t) | N(t) = i) = P(S_i > N(t) \cap S_j > N(t) \cap S_n > N(t) | N(t) = i)
\]

\[
= P(S_i > N(t) | S_j > N(t) | S_n > N(t) | N(t) = i)P(S_j > N(t) | N(t) = i)P(S_n > N(t) | N(t) = i).
\]

(42)

By considering the following situations, the probability of no **shock-based failure** by time \(t\) is derived further:

\(i\) When the number of shocks by time \(t\) is less than \(n\), or \(N(t) \leq n - 1\), we have

\[
P(S > N(t) | N(t) = i) = 1 \times F_{\lambda}(D_i)^{-1} \cdot e^{-it\lambda} \cdot (43)
\]

\(ii\) When the number of shocks by time \(t\) is equal to \(n\), or \(N(t) = n\), we have

\[
P(S > N(t) | N(t) = n) = \left\{ \frac{F_{\lambda}(D_i) - F_{\lambda}(D_i)^{-1}}{F_{\lambda}(D_i)} \right\} e^{-it\lambda} \cdot (44)
\]

\(iii\) When the number of shocks by time \(t\) is larger than \(n\), or \(N(t) > n\), we have

\[
P(S > N(t) | N(t) = i > n) = F_{\lambda}(D_i) + F_{\lambda}(D_i) - F_{\lambda}(D_i) + \ldots
\]

\[
+ F_{\lambda}(D_i) \frac{F_{\lambda}(D_i) - F_{\lambda}(D_i)^{-1}}{F_{\lambda}(D_i)} \cdot P(S_{i-1} > n | S_{i-1} > n) e^{-it\lambda} \cdot (45)
\]

Now by summing them all, we finally have the probability that a system does not experience the **shock-based failure** by time \(t\), i.e., the fatal shock count is greater than \(N(t)\):

\[
P(S > N(t)) = \sum_{i=0}^{\infty} P(S > N(t) | N(t) = i)P(N(t) = i)
\]

\[
= \sum_{i=0}^{\infty} F_{\lambda}(D_i)^{-1} \cdot e^{-it\lambda} \cdot \left\{ \frac{F_{\lambda}(D_i) - F_{\lambda}(D_i)^{-1}}{F_{\lambda}(D_i)} \right\} e^{-it\lambda} \cdot (46)
\]

2.3.1.3 System Reliability Function

For a system to survive by time \(t\) in this DTS model, we need to ensure that **degradation-based failure** and **shock-based failure** do not occur by time \(t\). The reliability function for such a system using Eqs. (2) and (46) is derived as

\[
R(t) = \sum_{i=0}^{\infty} P(S > i)P(X(t) + \sum_{i=0}^{N(t)} B_i < H) | N(t) = i \cdot (47)
\]

2.3.2 Condition-based Maintenance Strategy

In this section, we propose a condition-based maintenance model with periodic inspection. The proposed CBM takes into account corrective replacement, preventive replacement, and imperfect repair actions. The maintenance actions take place according to the detected condition of the system through inspections. The proposed condition-based maintenance model is illustrated in Figure 10, which presents the critical level for preventive replacement, \(H_p\), the critical level for imperfect repair, \(H_{ip}\), and the restored level from imperfect repair, \(H_r\).
Figure 10: Condition-based maintenance policy

Assumptions

If the system is detected to be failed due to shock-based or degradation-based failure, a corrective replacement is performed. If the degradation is beyond the cut-off limit, $H_D$, a preventive replacement is implemented even though the device is still functioning. If the degradation exceeds the warning limit value $H_W$, but is not beyond $H_D$, an imperfect repair takes place that restores the degradation level to some younger system status, $H_s$. $H_s < H_D < H_D < H_D$. If the system is operating and the degradation level is less than the warning limit, $H_s$, no action is needed.

From the basic renewal theory, the average long run maintenance cost per unit of time can be calculated as follows:

$$C_{r}(r) = \lim_{r \to \infty} \frac{C(r)}{r} = \frac{E[L]}{E[N]}$$

where $N_t$ is the number of inspections taking place in a renewal cycle.

The expected renewal cycle length, $E[L]$, is formulated as follows

$$E[L] = \sum_{i=1}^{\infty} E[L|N_i = i]P(N_i = i) = \sum_{i=1}^{\infty} i \cdot t \cdot P(N_i = i),$$

where $N_i$ is the number of inspections taking place in a renewal cycle. The expected renewal cycle length is given by $E[L]$.

2.3.2.2 Probability of Performing Preventive Replacement

The system is preventively replaced when the overall degradation is greater than $H_D$, but the system is still functioning. Two scenarios may occur before the preventive replacement is performed.

Scenario 1: No repair is recorded by the $i^{th}$ inspection.

$P(N_{IR} = j|N_{IR} > i-1) = \sum_{i=1}^{\infty} j \cdot t \cdot P(N_{IR} = j)P(N_{IR} = j),$ (51)

where $T_{m}$, the first time that the degradation reaches $H_m$ given the initial degradation level equals $0$

Figure 11: Preventive replacement at the $i^{th}$ inspection

Figure 11 shows an example of Scenario 1 where the degradation level is less than the warning limit $H_W$ at time $(i-1)t$, and it is within the interval of $[H_D, H_D)$ at time $it$ leading to a preventive replacement. Let $P(N_{IR} = i|N_{IR} > i-1)$ denote the probability of performing a preventive replacement at time $it$, given the system has not been repaired yet, and it is derived as follows:

$$P(N_{IR} = i|N_{IR} > i-1) = \sum_{i=1}^{\infty} j \cdot t \cdot P(N_{IR} = j)P(N_{IR} = j),$$

where $T_{m}$, the first time that the degradation reaches $H_m$ given the initial degradation level equals $0$

Scenario 2: One repair action by the $i^{th}$ inspection.

Figure 12: Preventive replacement at the $i^{th}$ inspection

Figure 12 depicts an example of Scenario 2 where the degradation is within $[H_W, H_D)$ at time $jt$, and therefore, an imperfect repair action is taken to restore the degradation level to $H_D$ at the $j^{th}$ inspection. Let $P(N_{IR} = i|N_{IR} = j)$ denote the probability of performing the preventive replacement at the $i^{th}$ inspection given the system has been repaired at the $j^{th}$ inspection, $j < i$, and we have

$$P(N_{IR} = i|N_{IR} = j) = \sum_{i=1}^{\infty} j \cdot t \cdot P(N_{IR} = j)P(N_{IR} = j).$$

2014 Annual Report

Grant #: 0970140 and 0969423
2.3.2.3 Probability of Performing Corrective Replacement

The system is correctly replaced when it fails due to either internal degradation process or external shock process. Two scenarios may happen before the corrective replacement takes place.

**Scenario 1:** No repair is recorded by the $i$th inspection.

\[
\begin{align*}
P(N_{CR} = i | N_{R} = i - 1) &= F_t \left( (i-1) \tau_{i-1} | \varphi = 0 \right) - F_t \left( i \tau_{i-1} | \varphi = 0 \right), \tag{55}
\end{align*}
\]

![Figure 13: performing corrective replacement at the $i$th inspection interval](image)

Figure 13 shows an example of Scenario 1 where the degradation level is less than the warning limit $H_w$ at time $(i-1)\tau$, and the system is failed at time $i\tau$ due to degradation-based failure, leading to corrective replacement. Let $P(N_{CR} = i | N_{R} > i - 1)$ denote the probability of performing the corrective replacement at time $i\tau$, given the system has not been repaired yet:

\[
P(N_{CR} = i | N_{R} > i - 1) = F_t \left( i \tau_{i-1} | \varphi = 0 \right) - F_t \left( (i-1) \tau_{i-1} | \varphi = 0 \right). \tag{56}
\]

**Scenario 2:** One repair action is performed by the $i$th inspection.

\[
\begin{align*}
P(N_{CR} = i | N_{R} = j) &= F_t \left( (i-j) \tau_{i-1} | \varphi = H \right) - F_t \left( (i-j - 1) \tau_{i-1} | \varphi = H \right), \tag{57}
\end{align*}
\]

Then the pdf of $\Theta$ for the specific case when $\beta$ and $Z(t)$ follow normal distributions is derived as follows:

\[
f_{\Theta}(\theta) = \frac{dF_t(\Theta; \varphi)}{d\theta} = \frac{dF_t(H; \varphi)}{d\theta} = \sum_{k=0}^{\infty} \left[ H_k \left( \frac{\sigma^2 \theta^2 + \sigma^2 + \sigma \alpha^2}{\sigma^2 \theta^2 + \sigma^2 + \sigma \alpha^2} \right)^{k/2} \right] \frac{e^{-\lambda^2/2}}{\sqrt{2\pi \sigma^2 \theta^2 + \sigma^2 + \sigma \alpha^2}}. \tag{59}
\]

2.3.2.5 Expected Cost Rate

\[
E(C_i) = E(\text{Inspection Cost}) + E(\text{Imperfect Repair Cost}) + E(\text{Preventive Replacement Cost}) + E(\text{Corrective Replacement Cost}). \tag{60}
\]

**Illustrative Examples**

Numerical examples are presented to illustrate the reliability and maintenance models that were developed. The corresponding values for the parameters in reliability analysis are given in Table 6, where some parameters are adopted from the literature and others are assumptions based on typical and plausible values.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H$</td>
<td>0.00125 $\mu$m$^3$</td>
</tr>
</tbody>
</table>
For the described system, the reliability function $R(t)$ in (47) is plotted in Fig. 15, and sensitivity analyses was performed to measure the effect of the changing parameters $D_n$, $D_1$, and $\delta$ on the reliability function.

![Reliability vs. Time](image)

**Fig. 15:** Plot of $R(t)$

To obtain the optimal time interval for the periodic inspection, we need to minimize the average long-run maintenance cost rate, $C_A(\tau)$. We use the simplex search method of Lagarias et al [15] that is a direct search method without using numerical or analytic gradients. For an example of $C_I=$$10$, $C_{lH}=$$50$, $C_{pH}=$$100$ and $C_{Cr}=$$500$, the minimum average long-run maintenance cost rate is $57,158.50$ at $t^*=7.59\times10^4$ revolutions, the optimum inspection interval. Figure 16 presents $C_A(\tau)$ as a function of $\tau$. We also performed sensitivity analyses to capture the impact of the model parameters ($H$, $H_p$, $H_n$, and $H_H$) on the optimal solutions.

![Average long-run maintenance cost rate versus inspection interval](image)

**Fig. 16:** Average long-run maintenance cost rate versus inspection interval

3. Reliability Analysis Considering Multiple Components: This section presents our work on Task 3, analysis for multiple component devices. We developed reliability models for multi-component systems with Dependent Degrading Components based on Thermodynamic Physics-of-Failure and optimized condition-based maintenance policy [16]. These proposed reliability models can be applied directly or customized to other similar complex systems.

### 3.1 Reliability Analysis and Condition-based Maintenance of Systems with Dependent Degrading Components based on Thermodynamic Physics-of-Failure

In this section, we present a new reliability model and a unique condition-based maintenance model for complex systems with dependent components subject to respective degradation processes, and the dependence among components is established through environmental factors. Common environmental factors, such as temperature, can create the dependence in failure times of different degrading components in a complex system. The system under study consists of one dominant component and $n$ statistically dependent components that are all subject to degradation. We consider two aspects that link the degradation processes and environmental factors: the degradation of dominant component is not affected by the state of other components, but may influence environmental factors, such as temperature; and the $n$ dependent components degrade over time and their degradation rates are impacted by the environmental factors. Based on the thermodynamic study of the relationship between degradation and environmental temperature, we develop a reliability model to mathematically account for the dependence in multiple components for such a system. Considering the unique dependent relationship among components, a novel condition-based maintenance model is developed to minimize the long run expected cost rate. A numerical example is studied to demonstrate our models, and sensitivity analysis is conducted to test the impact of parameters on the models.

#### 3.1.1 Thermodynamic Study for Physics-of-Failure

For a system consisting of one dominant component and $n$ statistically dependent components that are all subject to degradation, we consider two aspects that link the degradation processes and environmental factors [17]:

- The dominant component degrades over time, and its degradation rate or lifetime distribution is not affected by the state of other components. However, the degradation process of the dominant component may influence environmental factors, such as temperature. For example, the wear degradation of a microengine increases ambient temperature.
- The $n$ dependent components degrade over time and their degradation rates are impacted by the

<table>
<thead>
<tr>
<th>$H_p$</th>
<th>0.00100 μm$^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_n$</td>
<td>0.00075 μm$^3$</td>
</tr>
<tr>
<td>$H_H$</td>
<td>0.00050 μm$^3$</td>
</tr>
<tr>
<td>$D_n$</td>
<td>1.55 Gpa</td>
</tr>
<tr>
<td>$D_1$</td>
<td>1.20 Gpa</td>
</tr>
<tr>
<td>$\varphi$</td>
<td>0</td>
</tr>
<tr>
<td>$\mu_0$</td>
<td>$8.4823\times10^{3}μm^3$</td>
</tr>
<tr>
<td>$\sigma_0$</td>
<td>$6.0016\times10^{10}μm^3$</td>
</tr>
<tr>
<td>$\mu_w$</td>
<td>1.2 Gpa</td>
</tr>
<tr>
<td>$\sigma_w$</td>
<td>0.2 Gpa</td>
</tr>
<tr>
<td>$\mu_I$</td>
<td>$1.0\times10^{4}μm^3$</td>
</tr>
<tr>
<td>$\sigma_I$</td>
<td>$2\times10^{7}μm^3$</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>$1\times10^{10}μm^3$</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$2\times10^{3}$revolutions</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>$5\times10^{7}$revolutions</td>
</tr>
</tbody>
</table>
environmental factors. For instance, the elevated temperature accelerates the degradation of resistors.

To demonstrate the thermodynamic analysis and reliability modeling, we use an example application. The dominant component in an example system is a microengine that experiences wear degradation over time, and the wear-out process increases the ambient temperature. In the system, there are $n$ temperature-sensitive thin-film resistors whose resistances increase over time, and the degradation rates increase as the temperature elevates due to the wear-out process of the microengine. In order to analyze reliability performance of this system, we need to understand physics-of-failure mechanisms for these degradation processes, specifically through the study of thermodynamics.

The relationship between wear degradation and temperature has been of great interest to many researchers in thermodynamics [18].

### 3.1.1.1 Wear Degradation and Thermal Response

For the dominant component (e.g., a microengine), we assume its wear degradation $X(t)$ follows a linear degradation path, $X(t) = \varphi + \beta t + \epsilon_0$, where the initial value $\varphi$ is a constant. The degradation rate $\beta$ follows a normal distribution, $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, characterizing the unit-to-unit variability; and $\epsilon_0$ is the random error term following a normal distribution, $\epsilon_0 \sim N(0, \sigma_\epsilon^2)$, capturing the temporal variability. The microengine is considered to be failed when the wear degradation is greater than a failure threshold value $H$.

The degradation of the dominant component leads to the rise of ambient temperature. According to Amiri et al. [19], the temperature rise $\Delta T$ at the interface during steady state operation has a linear relationship with the wear degradation rate:

$$\Delta T = \frac{\Psi}{\xi} \beta,$$

where $\xi = \frac{K}{h}$, $K$ is the wear coefficient, $\eta$ is the heat partitioning factor, $\mu_{ave}$ is the friction coefficient, $h$ is the material hardness, and $\Psi$ is a constant. Because $\beta$ follows a normal distribution, $\beta \sim N(\mu_\beta, \sigma_\beta^2)$, the temperature rise $\Delta T$ at the steady state is also a normal random variable with mean of $\mu_{\Delta T}$ and variance of $\sigma_{\Delta T}^2$.

### 3.1.1.2 Arrhenius Relationship

Similar to the degradation process modeling of the dominant component, we want to incorporate both unit-to-unit variability and temporal variability in the degradation process modeling of the dependent components as well. For the $n$ dependent components, such as thin film resistors, the resistance $r_i(t)$ increases linearly over time, $r_i(t) = r_0 + \rho t + \epsilon_i$, where $r_0$ is the initial resistance of component $i$, $p_i$ is the degradation rate of component $i$, $\epsilon_i$ is the random error with a normal distribution, $\epsilon_i \sim N(0, \sigma_\epsilon^2)$ for component $i$, $i = 1, 2, ..., n$.

It is known that the degradation rate $p_i$ is affected by the temperature via the Arrhenius relationship [20]:

$$p_i = r_0 A \exp \left( -\frac{E_a}{kT} \right),$$

where $E_a$ is the activation energy in eV, $k$ is the Boltzmann constant, $T$ is the temperature in Kelvin, and $A$ is a constant. Therefore, the resistance is expressed as

$$r_i(t) = r_0 + \rho t + \epsilon_i = r_0 \left( 1 + \Delta t \exp \left( -\frac{E_a}{kT} \right) \right) + \epsilon_i.$$

A thin film resistor is considered to be failed when the resistance is beyond the failure threshold value, $L_i$, $i = 1, 2, ..., n$. Figure 17 shows 30 pairs of simulated degradation processes of a dominant component and a dependent component. We can notice that the lifetime of the dependent component has a much larger variance than that of the dominant component, because the degradation rate of the dominant component significantly affects the degradation rate of the dependent component. For a series system with dependent components, we develop its reliability function and a unique condition-based maintenance policy in the following sections.

![Figure 17. Simulation of the stochastic degradation processes for dominant and dependent components](image)

#### 3.1.2 System Reliability Modeling

Consider a microengine and $n$ thin-film resistors connected in series. System reliability at time $t$ is the probability that it survives by time $t$, that is, the degradation level of each component should be less than the corresponding failure threshold level [27]:

$$R(t) = P \left( X(t) < H, r_1(t) < L_1, ..., r_n(t) < L_n \right).$$

Because the degradation processes of these components are dependent through temperature change, we need to compute it by finding the conditional probability given $\Delta T$. Based on the law of total probability, we then integrate this conditional probability multiplied by the probability density function (pdf) of $\Delta T$ to derive the system reliability:
\[ R(t) = P(X(t) < H, \tau_1(t) < L_1, \ldots, \tau_n(t) < L_n) \]
\[ = \int_{-\infty}^{\infty} P(X(t) < H, \tau_1(t) < L_1, \ldots, \tau_n(t) < L_n | \Delta T = \delta) f_{\Delta T}(\delta) d\delta \]
\[ = \int_{-\infty}^{\infty} P(X(t) < H | \Delta T = \delta) \prod_{i=1}^{n} P(\tau_i(t) < L_i | \Delta T = \delta) f_{\Delta T}(\delta) d\delta, \] (69)

The temperature rise \( \Delta T \) is a normal random variable with mean of \( \mu_{\Delta T} / \xi \) and variance of \( \sigma_{\Delta T}^2 / \xi^2 \), and its pdf can be expressed as
\[ f_{\Delta T}(\delta) = \frac{\xi}{\sigma_{\Delta T} \sqrt{2\pi}} \exp \left( -\frac{(\delta - \mu_{\Delta T} / \xi)^2}{2 \sigma_{\Delta T}^2} \right). \] (70)

Finally, the reliability function is expressed as
\[ R(t) = \int_{-\infty}^{\infty} \Phi \left( \frac{1}{\sigma_{\Delta T}} \left( H - \varphi - \frac{\xi}{\varphi} \right) \right) \prod_{i=1}^{n} \Phi \left( \frac{1}{\sigma_i} \left( L_i - r_i - r_i/\Delta T \exp \left( -\frac{E_i}{k(T_i + \delta)} \right) \right) \right) \times \frac{\xi}{\sigma_{\Delta T} \sqrt{2\pi}} \exp \left( -\frac{(\delta - \mu_{\Delta T} / \xi)^2}{2 \sigma_{\Delta T}^2} \right) d\delta. \] (71)

### 3.1.3 Condition-based Maintenance Modeling

Due to the unique relationship between the dominant and dependent components and their characteristics, we propose a new maintenance model. Since the dominant component plays a key role in this system and it is typically expensive, we consider the case when the replacement cost of the dominant component is much higher than the replacement cost of all the dependent components combined (or the subsystem). We assume that the system is non-repairable or not worth repairing rather than replacing. The replacement time for the whole system and the subsystem of all dependent components is negligible. With more attention on the expensive dominant component, the proposed maintenance strategy is designed as follows and illustrated in Figure 18:

- **Periodic inspection of length \( \tau \) is carried out to observe or measure the degradation level \( X(t) \) of the dominant component. If the degradation level is less than a warning limit, \( D \), no action is taken; and if the degradation level is between the warning limit \( D \) and the failure threshold \( H \), preventive replacement takes place.
- If the dominant component fails (the degradation level is beyond the failure threshold \( H \)) between two inspection actions, it is self-announcing and corrective replacement is implemented.
- Every time the dominant component is replaced preventively or correctly, the whole subsystem of dependent components is replaced preventively for the purpose of saving time/labor, shown as ‘PM’ in Figure 18.
- The conditions of dependent components are not checked during the periodic inspection actions.

However, the failure of any dependent component is self-announcing. If one of the dependent components in the subsystem fails, we replace the whole subsystem correctly for the purpose of saving time/labor.

![Condition-based maintenance model](image)

To determine the inspection interval \( \tau \) and the warning limit \( D \), we need to derive and optimize the expected total maintenance cost per unit of time:

\[ \text{Expected cost rate} = \frac{E(\text{Total Cost})}{E(\text{Cycle Length})} \] (72)

As illustrated in Figure 18, a renewal cycle of the dominant component can be terminated due to a preventive replacement (the degradation level is between \( D \) and \( H \)) or a corrective replacement (the degradation level is beyond \( H \)). To find the expected cost per renewal cycle and the expected renewal cycle length, we need to consider these two cases.

#### 3.1.3.1 Renewal Cycle Terminated due to Preventive Replacement

We start with the case that a renewal cycle is terminated when the degradation level exceeds the warning limit, and therefore, preventive replacement is performed for the dominant component. Let \( N_{PM} \) denote the inspection count at which a preventive maintenance/replacement is implemented. The probability of performing preventive replacement is derived as follows.

1) The preventive replacement is performed at the 1st inspection, or \( N_{PM} = 1 \):
\[ P(N_{PM} = 1) = \Phi \left( \frac{H - \varphi - \mu_{\Delta T} \varphi}{\sqrt{\sigma_{\Delta T}^2 + \sigma_{\varphi}^2}} \right) - \Phi \left( \frac{D - \varphi - \mu_{\Delta T} \varphi}{\sqrt{\sigma_{\Delta T}^2 + \sigma_{\varphi}^2}} \right) \] (73)

2) The preventive replacement is performed at the \( i \)th inspection, or \( N_{PM} = i > 1 \):
\[ P(N_{PM} = i > 1) = P(D < X(i\tau) < H, (H - \Delta T(i\tau) < \varphi)) \]
\[ = \int_{-\infty}^{\infty} P(D < X(b \tau) < H, (H - \Delta T(b \tau) < \varphi)) f_{\Delta T}(b) db \]
\[ = \int_{-\infty}^{\infty} \left( \Phi \left( \frac{D - \varphi - b \tau}{\sigma_{\varphi}} \right) - \Phi \left( \frac{H - \varphi - b \tau}{\sigma_{\varphi}} \right) \right) f_{\Delta T}(b) db \]
\[ + \int_{-\infty}^{\infty} \left( \Phi \left( \frac{H - \varphi - r_i}{\sigma_{\varphi}} \right) - \Phi \left( \frac{H - \varphi - r_i - r_i/\Delta T(i\tau)}{\sigma_{\varphi}} \right) \right) f_{\Delta T}(b) db. \] (74)
3.1.3.2 Renewal Cycle Terminated due to Corrective Replacement

When the degradation level of the dominant component exceeds the failure threshold $H$, a renewal cycle is terminated and corrective replacement is performed. To find the probability of performing corrective replacement upon failure, we need to derive the failure time distribution of the dominant component. The degradation process $X(t)$ follows a normal distribution with mean $\mu_X(t) = \phi + \mu_\beta t$, and variance $\sigma_X^2(t) = \sigma_\rho^2 + \sigma_\delta^2 + \sigma_\kappa^2$. If we denote $T_x$ as the time of the degradation path reaching a threshold $x$, then the cumulative distribution function (CDF) of $T_x$ is

$$P(T_x < t) = P(\phi + \mu_\beta t + \epsilon_t > D) = 1 - \Phi \left( \frac{D - \phi - \mu_\beta t}{\sigma_\rho} \right).$$

Its pdf can be calculated by taking the first derivative of the cdf with respect to $t$, which is

$$f_{T_x}(t) = \frac{dP(T_x < t)}{dt} = \frac{\phi (D - \phi - \mu_\beta t)}{\sigma_\rho (D - \sigma_\rho)}.$$

In the case of a failure occurring between inspections, it indicates that at the previous inspection the degradation level of the dominant component does not reach the warning limit $D$ yet. We need to include this condition in our derivation of the failure distribution for the dominant component. The cdf of the failure time $T_H$ conditioning on $T_D$ is

$$P(T_H < t | T_D = t_D) = P(X(t) > H | X(t_D) = D) = 1 - \Phi \left( \frac{H - D - \mu_\beta (t - t_D)}{\sigma_\rho (t - t_D)} \right).$$

Similarly, the pdf of the failure time $T_H$ conditioning on $T_D$ can be derived as

$$f_{T_H|T_D}(t | t_D) = \frac{dP(T_H < t | T_D = t_D)}{dt} = \frac{\phi (H - D - \mu_\beta (t - t_D))}{\sigma_\rho (t - t_D)}.$$

3.1.3.3 Optimization Model

The dominant component is either preventively replaced at inspection or correctively replaced upon failure between inspections. Based on Eqs. (73)-(77), the expected renewal cycle length can be found as:

$$E(Cycle\ Length) = \sum_{i=1}^{n} E[N_{t_D(i)}] + \sum_{i=1}^{n} \int_{t_{D(i)}}^{t_{D(i)+t_{max}}} f_{T_H|T_D}(t | t_D) dt.$$

The overall maintenance cost includes preventive and corrective replacement costs for the dominant component, $C_{PI}$ and $C_{CI}$, preventive and corrective replacement costs for the subsystem of all dependent components, $C_{PD}$ and $C_{CD}$, and the inspection cost $C_I$. The system downtime cost is not considered, as we assume that the time for maintenance actions, such as inspection and replacement, is negligible.

When a renewal cycle is terminated at the $i$th inspection due to preventive replacement, the incurred cost includes the preventive replacement cost of the dominant component, $C_{PI}$, the preventive replacement cost of the subsystem, $C_{PD}$, the cost for $i$ inspection actions, and the cost to correctively replace the subsystem before it. The subsystem can be correctly replaced multiple times whenever one of the dependent components fails before it. The number of corrective replacements (or the number of failures) of the subsystem prior to $it$ can be calculated by the renewal function, $M(t)$, which is derived in the next section. When a renewal cycle is terminated between $(i-1)r$ and $ir$ due to failure, the incurred cost includes the corrective replacement cost of the dominant component, $C_{CI}$, the preventive replacement cost of the subsystem, $C_{PD}$, the cost for $i-1$ inspection actions before failure, and the cost to correctly replace the subsystem before failure. Then the expected total maintenance cost is derived to be:

$$E(Total\ Cost) = \sum_{i=1}^{n} (C_{PI} + C_{PD} + iC_I + M(t)C_{CI}) \sum_{j=1}^{n} \int_{t_{D(i)}}^{t_{D(i)+t_{max}}} f_{T_H|T_D}(t | t_D) dt.$$

Based on Eqs. (79) and (80), we propose the following constrained non-linear optimization problem for the maintenance optimization:

$$\min_{\tau(D)} \frac{E(Total\ Cost)}{E(Cycle\ Length)}$$

Subject to: $0 < D < H$

where $t_{max}$ is the allowed upper bound of the inspection interval. The Sequential Quadratic Programming algorithm (Matlab optimization toolbox) is used to solve this constrained non-linear optimization problem.

3.1.3.4 Renewal Function of the Subsystem

To calculate the expected total cost per cycle, we need to have the number of corrective replacements (or the number of failures) of the subsystem in a renewal cycle, namely, the renewal function, which requires the reliability function of the subsystem of dependent components, $R_{Sub}(x)$:

$$R_{Sub}(x) = P(T_H < D | T_D = D) = \sum_{i=1}^{n} \int_{t_{D(i)}}^{t_{D(i)+t_{max}}} \frac{1}{\sigma_\rho} \left( \int_{t_{D(i)}}^{t_{D(i)+t_{max}}} f_{T_H|T_D}(t | t_D) dt \right) dt.$$

The renewal function is calculated to be

$$M(t) = F_{Sub}(t) + \int_{0}^{t} M(t - u) f_{Sub}(u) du,$$

where $F_{Sub}(t)$ and $f_{Sub}(t)$ are the cdf and pdf of the subsystem, respectively. It is difficult to derive the closed form of the renewal
function given the complicated subsystem reliability function. Estimation of the renewal function is typically applied:

$$M(t) = F_{sub}(t) + \int_0^t \frac{F_{sub}^2(u)}{\int_0^u R_{sub}(v)dv} \, du$$

Even using the estimate of the renewal function, the complex non-linear optimization model is still difficult to solve mathematically. One approach demonstrated in the numerical example is to simplify the subsystem reliability function by fitting it to a simple regression model that could lead to a closed-form renewal function. For example, when the failure time follows an exponential distribution with arrival rate $\lambda$, its renewal function is simply $M(t) = \lambda t$.

### 3.1.4 Numerical Example

In this numerical example, a system consisting of one microengine (the dominant component) and two identical resistors (dependent components) is studied. The three components are dependent because the degradation of the microengine causes the temperature rise in the surrounding environment, while the temperature rise accelerates the degradation of both resistors. For this type of system, we are interested in determining reliability over time and the optimal maintenance strategies using the reliability and maintenance models that we developed. The parameters and their values used in our models are listed in Table 7. Figure 19 plots the reliability of the system over time.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k$</td>
<td>8.671x10^7 eV/K</td>
</tr>
<tr>
<td>$E_0$</td>
<td>1.29 eV (for TaN)</td>
</tr>
<tr>
<td>$A$</td>
<td>1.162x10^6 (for TaN)</td>
</tr>
<tr>
<td>$T_0$</td>
<td>293 K</td>
</tr>
<tr>
<td>$h$</td>
<td>11.5 Gpa</td>
</tr>
<tr>
<td>$K$</td>
<td>1x10^4</td>
</tr>
<tr>
<td>$\mu_{rel}$</td>
<td>0.7</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0.5</td>
</tr>
<tr>
<td>$\psi$</td>
<td>2.484x10^{-14}pa</td>
</tr>
<tr>
<td>$\mathcal{P}$</td>
<td>4.55x10^{13} K/W</td>
</tr>
<tr>
<td>$H$</td>
<td>0.005 $\mu m)$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$N(\mu_0, \sigma_0^2)$, $\mu_0 = 8.4823x10^{-9} \mu m$/time unit, $\sigma_0 = 6.0016x10^{-10} \mu m$/time unit</td>
</tr>
<tr>
<td>$\psi$</td>
<td>0</td>
</tr>
<tr>
<td>$\epsilon_0$</td>
<td>$N(0,\sigma_{\epsilon_0}^2)$, $\sigma_{\epsilon_0} = 5.0000x10^{-8} \mu m)$</td>
</tr>
<tr>
<td>$\epsilon_1$</td>
<td>$N(0,\sigma_{\epsilon_1}^2)$, $\sigma_{\epsilon_1} = 5.0000x10^{-1} \Omega$</td>
</tr>
<tr>
<td>$\epsilon_2$</td>
<td>$N(0,\sigma_{\epsilon_2}^2)$, $\sigma_{\epsilon_2} = 5.0000x10^{-1} \Omega$</td>
</tr>
<tr>
<td>$\sigma_{R_{01}}, \sigma_{R_{02}}$</td>
<td>250.48 $\Omega$</td>
</tr>
<tr>
<td>$L_{s1}, L_{s2}$</td>
<td>300.58 $\Omega$</td>
</tr>
</tbody>
</table>

Figure 19. System reliability over time

For the condition-based maintenance model, we assume that the preventive and corrective replacement costs for the subsystem of two dependent components are 40 and 50, respectively. Because the replacement cost of the dominant component is far more expensive than that of the dependent components, the preventive and corrective replacement costs for the dominant component are 400 and 500, respectively. The inspection cost is 10 per inspection for the dominant component.

It becomes difficult and inefficient to directly solve the optimization problem, because of the complex form of the subsystem reliability function and the resulting renewal function. An alternative way is to simplify the subsystem reliability function by using a regression model to approximate it. After fitting the subsystem reliability to an exponential regression model, the renewal function has a simple form, which is $M(t) = 1.332x10^{-4}t$. Then we use the Sequential Quadratic Programming algorithm (in Matlab R2013a) to solve this constrained nonlinear optimization problem and obtain the minimum expected cost rate, 8.98x10^{-4}, when $r^* = 5.57x10^4$ and $D^* = 0.0025$. The expected cost rates at different $r$ and $D$ levels are plotted in Figure 20.

Figure 20. 3D plot of the expected cost rate vs $r$ and $D$
cost-reliability and maintenance optimization considering uncertain future usage scenarios [21].

4.1 Two-Stage Stochastic Algorithm for System Cost-Reliability and Maintenance Optimization Considering Uncertain Future Usage Scenarios:

We develop a two-stage stochastic cost-reliability optimization model for multi-component systems subjected to uncertain stress exposure which leads to uncertain component and system degradation and wear. Information of conditions and inputs to determine an optimum preventive maintenance policy. In this formulation, the system is exposed to distinct usage scenarios that are collectively represented as the future usage profile. This profile is described by a set of specific scenarios and their probability of occurrence or likelihood. Once these future usage profiles are known or can be estimated, the cost rate for prospective system configurations and associated maintenance can be modeled. The cost-reliability system design with maintenance modeling problem is defined as a two-stage stochastic cost-reliability programming problem with recourse. The decision variables for the first-stage are the selection of components and the number of components to be used in the system, where the cost rate objective function is constrained by component availability. The second stage variables are defined by the corrective and preventive maintenance plan in order to optimize the maintenance time interval for planned replacement of components in the system. Heuristics, meta-heuristics or exact algorithms have been used for solving redundancy allocation problems [22]. Maintenance optimization is formulated to minimize costs of maintenance and/or to increase system availability [23]. System reliability optimization has been studied considering uncertain future operating conditions and usage stresses [24]. Expected reliability for systems under uncertain usage environment with multiple dependent competing failure processes was studied [25].

4.1.1 Future usages Profile of Operating Conditions and Stresses: Uncertainty in future usage can involve several factors of uncertain stress/load to components within the future usage profile, as shown in Figure 21. In some applications, system reliability estimation is often problematic due to unplanned variation, or changing operating stresses. More often, not only operating conditions and stresses but also environmental stresses such as temperature or vibration can be uncertain and dependent on usage conditions of components and systems in the future.

Component stress/load parameters are represented in the future usage scenarios as vectors, which describe how changing stresses in different future usage scenarios affect reliability functions. Uncertain stress factors are demonstrated in the form of a random vector $U$ in the future usage profile. The model parameter $U_k$ is a random variable of the $k^{th}$ stress factor that the component or system experiences, e.g., temperature, humidity, voltage or others. A random future usage vector is represented with $c$ different operating usage and stress factors where each uncertain usage condition has different effects on each component, $U = (U_1, U_2, ..., U_c)$.

The future usage profile defines possible future usage scenarios defined by future usage stress vectors. In practice, the possible future usage scenarios can be enumerated from prediction of how the system will be used and possible occurrences of each future usage. The future usage stress vector in each scenario $l$ is determined from random future usage vector $U$. There are $c$ different factors of future usage stresses which determine how stresses of different factors change in that future usage scenario. For example, the determined future usage vector for possible future scenario 1 is represented by vector $u_1$, vector $u_2$ for possible future scenario 2 and so on.

![Figure 22. Discrete scenarios in future usage profile](image)

In Figure 22, the possible future usage vectors in the future usage profile are defined as $u_l$ and each future usage scenario is associated with a probability $p_l$. Let $u_l = (u_{l1}, u_{l2}, ..., u_{lc})$ represent a determined vector of operating usage conditions and/or stresses. For example, for electronic components, temperature is a critical contributor to component failure ($u_{l1}$), and for a given future usage scenario, the risk of component failure increases along with increasing temperatures. For mechanical components, mechanical loading ($u_{l2}$) and stress ($u_{l3}$) are important factors. The current operating condition and stress vector is $u_0$ and it is known with certainty and given as $u_0 = 0$. Each operating stress from future usage profile is defined by the usage vector $U$. All stresses have been scaled from 0 to 1. For example, $u_{l2}$ is a second usage variable at future $l$.  

---

2014 Annual Report
Grant #: 0970140 and 0969423
4.1.2 Future usages Profile of Operating Conditions and Stresses: For this section, a parallel system configuration is introduced for the stochastic reliability model, as shown in Figure 3. Since the number of identical components, \( x \), is connected in parallel within a system, the system reliability is given in Eq. (84).

\[
R(x;t) = 1 - (1 - r(x;t))^x
\]

(84)

For our proposed system reliability function, the future usage profile of operating conditions and stresses is considered for developing a system reliability model. The component reliability \( r(x) \) is no longer constant, but a random variable. The number of identical components, \( x \), is connected and the system reliability is given as follows

\[
R(x;U;t) = 1 - (1 - r(U;t))^x
\]

(85)

\( U \) is a random future usage vector, so \( r(U;t) \) and \( R(x;U;t) \) are also random variables. The system reliability equation is derived from the component reliability based on the future usage profile.

4.1.2.1 Maintenance Policy with replacement time interval variable

An age replacement maintenance policy of stochastic system reliability is modeled from a renewal reward process that contains multiple random variables. The system is replaced preventively with a new one at age \( y \) while the system that fails before age \( y \) is replaced correctly. The replacement for the entire system with a new system is applied after the time interval \( y \). When the system fails before time \( y \), the replacement is done immediately. Long run average replacement cost per unit time can be evaluated as

\[
\lim_{n \to \infty} \frac{\text{Expected maintenance cost between two replacements}}{\text{Expected time between two replacements}} = \frac{E(TC)}{E(U)}
\]

(86)

where \( TC \) is the total cost of a renewal cycle, and \( U \) is the length of the renewal cycle. The cycle time is equal to \( y \) for preventive replacement. Then the expected total maintenance cost per system is given as

\[
E(TC) = (C_r + C_a)(1 - R(y)) + C_r R(y) = C_r (1 - R(y)) + C_a
\]

(87)

In each cycle, we have one replacement, and the cost \( C_r \) should be considered in all cases. However, if the system fails before the fixed replacement interval, \( C_a \), the cost on an unscheduled failure, should be also included, and the probability of failure before replacement interval \( y \) is \( 1 - R(y) \). Expected time between two replacements is

\[
E[U] = \int_0^y r(t;U) \, dt = \frac{y}{r(U)}
\]

(88)

Based on Equations (87) to (88), the average long-run maintenance cost rate is given as

\[
CR(y) = \frac{C_r (1 - R(y)) + C_a}{\int_0^y r(t;U) \, dt} = \frac{C_r (1 - R(y)) + C_a}{\int_0^y r(U) \, dt}
\]

(89)

4.1.3 Mathematical Formulation

To illustrate the proposed modeling, the two-stage stochastic problem is presented to model system design reliability under uncertainty. The long-term forecast is used for making decisions about the system configuration to be implemented in different future usage scenarios. However, the future conditions cannot be explicitly predicted. The decision makers must decide without fully understanding of future information. The decision variables in the first-stage are represented as the number of components composed in the system. The second-stage decision variables represent when the preventive or corrective actions are taken. Decision variables of maintenance replacement time interval are determined to optimize the system maintenance cost rate.

The number of components and maintenance time interval are decision variables in the objective function. Future usages depending on stress factors are also the parameters in the model and considered as random variable. This two-stage stochastic programing with recourse represents the cost-rate system design with maintenance modeling in Eq. (90).

\[
\min_c c_j(x) + Q(x,y)
\]

s.t. \( 1 \leq x \leq x_{\text{max}} \)

\( x \in \{0,1,2,3,...\} \)

Where

\[
Q(x,y) = \sum_{i=1}^{r} p_i Q(u_i; y; x)
\]

\[
c_j(x) = c(x/P, \text{int} \%, N)
\]

Reliability of system: \( R(x,y,U) = 1 - (1 - r(U))^x \)

Component reliability: \( r(y,U) = \exp(-\frac{y}{U}) \)

The net present value formula \( (A/P, \text{int} \%, N) = \frac{\text{int}(1+\text{int})^N}{(1+\text{int})^N - 1} \) is the relationship between \( P \) and \( A \) for finding a uniform series of end-of-period cash flows of amount \( A \) for \( N \) periods. \( c(x) \) is the equivalent end of period cash flows in a uniform series for a specified number of periods.

---

**Notation**

- \( U \) = random stress vector, \( U = (U_1, U_2, ..., U_l) \)
- \( u_i \) = Random stress variable \( U \in \{u_1, u_2, ..., u_l\} \)
- \( C \) = cost constraint
- \( a \) = Age
- \( a_k \) = Sensitivity coefficient for component \( i \) and stress factor \( k \)
- \( x \) = Design variable of number of components
- \( y \) = Maintenance inspection interval variable in different scenarios (\( j \))
- \( C(t) \) = cumulative maintenance cost by time \( t \)
- \( CR(y) \) = average long-run maintenance cost rate
- \( E[U] \) = expected value of the renewal cycle length
- \( E[T] \) = expected value of the total maintenance cost of the renewal cycle
- \( C_r \) = replacement cost per unit
- \( C_a \) = cost of replacement caused by failure

---

Grant #: 0970140 and 0969423
starting at the end of the first period and continuing through the last period.

The optimization model given in (90) is known as implicit representation of the stochastic program. By considering the value function or recourse function in (91), the condense representation can be shown in Eq. (92)

\[
\min c(A/P, int\%, N) + \sum_{j=1}^{v} p_j Q(u_j, y; x) \\
\text{s.t. } 1 \leq x \leq x_{\text{max}} \\
x \in \{0,1,2,3,...\}
\]

Where \(Q(y,U;x) = C_R[(1-r(y,U))x]+C_F \int_{0}^{\infty} R(t;U,x) dt\)

In the model, the decision variable, \(x\), is required to have integer solution while another decision variable, \(y\), is continuous. The technique presented to solve the two-stage stochastic model is a heuristic approach, which is used to obtain integer solutions. In order to simplify computational complexity from mixed-integer programming, one decision variable is searched along with another to optimize the best solutions.

### 4.1.4 Numerical Example

Future usage profile is expressed in Fig. 23 with three future usage stress variables and there are four probabilities \(p_j = (0.4 \ 0.4 \ 0.1 \ 0.1)\) for four scenarios. The future usage stress variables are related to changes of stress, operating conditions, environmental vibration and/or temperature and have been scaled from zero to one. Each possible scenario is used for the decisions corresponding to each possible future usage.

![Future usage probability](image)

Consider a parallel system where number of component represented by decision variable \(x\). Component cost is \(c\). The net present value and equivalent cash flow annual cost method are used to evaluate the annual system cost and budget with interest rate \((int = 0.2)\) and period of time \((N = 20)\).

The decision makers make a decision at the current time from the best known information in the future usage profile. Given the four future usage scenarios, the second-stage or correction action can be computed from \(p_j Q(u_j, y; x)\). The maintenance time interval is represented by decision variable \(y_j\). The pattern search algorithm in the MATLAB toolbox and combinatorial neighborhood search are used to find the recommended solutions of the number of component and the optimal maintenance policy.

This example demonstrates the model feasibility in a parallel system and the solutions of two-stage stochastic model are shown in Table 8. The recommended number of components that optimize annual system cost and maintenance is 4 components in parallel. The maintenance replacement intervals from the second-stage decision are taken for each scenario and also shown in Table 8.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Maintenance time interval</th>
<th>Number of components</th>
<th>Max values</th>
</tr>
</thead>
<tbody>
<tr>
<td>First stage</td>
<td>Parallel</td>
<td>Max</td>
<td>values</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>Maintenance time interval</td>
<td>306</td>
<td>–</td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Maintenance time interval</td>
<td>816</td>
<td>–</td>
</tr>
<tr>
<td>Scenario 3</td>
<td>Maintenance time interval</td>
<td>958</td>
<td>–</td>
</tr>
<tr>
<td>Scenario 4</td>
<td>Maintenance time interval</td>
<td>1319</td>
<td>–</td>
</tr>
<tr>
<td>Overall</td>
<td>Annual system cost ($/unit time)</td>
<td>2.14</td>
<td>–</td>
</tr>
</tbody>
</table>

The future usage profile displays the probabilities associated with operating condition and stress \(u\) vector. Each future usage vector represents one scenario that the components in the system may experience. The current operating and usage stress vector is given as \(u_0 = 0\).

An example problem is solved to demonstrate the numerical results by using the MATLAB optimization toolbox for the optimal continuous solution. Neighborhood search algorithm is performed to solve the mathematical formulation for the recommended integer solution.

When the optimal maintenance time interval is determined, the solutions of number of components in Figure 24 can be determined with the minimum value, and it is used to construct the parallel system. The maintenance policy is to replace the system at a fixed time interval when choosing \(C_F = $1500\) and \(C_R = $500\). We can find the minimum cost-reliability maintenance rate of $2.1459 cycle, which is obtained at each possible future usage scenario \(y^*=[306, 816, 958, 1319]\).
The objective value reduces significantly at the beginning and then the optimization algorithm keeps finding the optimal objective value. The minimum value can be obtained when any set condition is reached as shown in Figure 25.

5 Literature Review on Micro- and Nano-electromechanical System Reliability: In this section, we have systematically searched and reviewed the published papers related to reliability of micro and nano electromechanical systems [26]. Among more than 300 papers that were primarily considered, about 100 of them were selected to be integrated. The main objective is to cover the wide spectrum of research in this area, and integrate them to provide an insight for researchers who are interested in reliability research of MEMS and NEMS devices.

5.1 Reliability Research on Micro and Nano Electro-Mechanical Systems- A Review: Research on micro/nano-electro-mechanical (MEMS/NEMS) reliability is of crucial importance, due to the fact that we are facing an era in which MEMS and emerging NEMS are expected to have a major impact on our lives. Over the last decade, significant efforts have gone into the reliability study of MEMS/NEMS. Failure roots in micro and nano scales can be addressed by mechanical, electrical, chemical, thermal factors, or combinations of them, which can occur during different manufacturing and post-manufacturing phases. This section reviews and integrates the reliability issues of MEMS and NEMS in different phases of their life cycles including design, manufacturing, logistics, and operation. This section surveys the common failure modes and mechanisms (i.e., wear, degradation, adhesion, stiction, deformation, packaging, contamination, and electrical failure modes), the reliability aspects of design and manufacturing, as well as reliability evaluation and testing techniques for MEMS/NEMS. An extensive collection of papers were selected and integrated to cover various studies in this area, in order to provide an intuition and insight for researchers who are interested in reliability research of MEMS/NEMS. Systematic literature search shows the lack of research in system-level and probabilistic reliability analysis for MEMS/NEMS.

5.1.1 Related Review Work: As one of the earliest published review papers, Tanner (2000) addressed some of the early investigations in MEMS reliability. It was concluded that stiction, wear and resulting adhesion are the main concerns of MEMS reliability. Moreover, fatigue was not a reliability concern yet for polysilicon, while this issue can be a concern for other materials used in these devices. van Spengen (2003) presented an interesting overview of MEMS failure mechanisms and reliability issues. The major failure modes and mechanisms were discussed as follows: fracture, creep, stiction, electromigration, wear, degradation of dielectrics, delamination, contamination, pitting of contacting surfaces, and electrostatic discharge. In a recently published work by Huang et al. (2012), they have updated van Spengen (2003)’s review by focusing on failure mechanisms in MEMS, such as fracture, fatigue, stiction, wear, creep, electrical failures, and contamination. Furthermore, they analyzed the effects of failure modes on performance of devices, the related inspection techniques, as well as the measures to be taken in order to reduce the failure rates. Zhao et al. (2003) reviewed the mechanics of adhesion in MEMS and discussed the governing dimensionless numbers such as Tabor number, adhesion parameter and peel. The importance of surface roughness for adhesion contact, and three types ofasperity height distributions including Gaussian, fractal, and exponential distributions, as well as microscale plastic adhesion contact theory, were analyzed and discussed. Srikar & Spearing (2003) presented a critical review on several measurement techniques for MEMS, such as microbeam bend test, microtension test, wafer curvature measurements, dynamic tests, axisymmetric plate bend test, fabrication of passive strain sensors, and Raman spectroscopy. The common causes for structure-related errors in measurement and characteristics of typical test structures were discussed. Moreover, a rational approach to select an appropriate testing technique for microsystems design was proposed.
Bhushan (2003) presented a critical review on mechanisms of adhesion and stiction, along with different measurement techniques and methods to reduce the stiction in commercial MEMS/NEMS and magnetic storage devices. Several techniques for adhesion measurement, i.e., surface force apparatus, atomic force microscopy (AFM), microtriboapparatus, and cantilever beam array, were addressed. Yapu (2003) provided a review on stiction and anti-stiction research for MEMS/NEMS devices. Several experimental observations of stiction in micromachined accelerometers and RF-MEMS switches were presented, and some criteria for stiction of micro and nanostructures in MEMS and NEMS resulting from surface forces were discussed. The impacts of surface roughness and environmental conditions on stiction phenomenon, and methods to reduce stiction in MEMS and NEMS were reviewed as well. Jeng et al. (2007) introduced the concepts of nanotechnology reliability, including MEMS/NEMS devices. Various failure mode and mechanism analysis, manufacturing reliability, aging and degradation models, failure and lifetime models, reliability testing, evaluation and measurement techniques, as well as structure and parameter design techniques for reliability of nano devices were reviewed.

Bhushan (2007) presented an analytical review on nano-tribology and nano-mechanics of MEMS/NEMS and BioMEMS/BioNEMS materials and devices. Micro/nano-scale adhesion, friction, and wear studies of materials, lubrication studies, and component level stiction phenomena for aforementioned devices were reviewed. It was concluded that adhesion, stiction/friction and wear reduce the lifetimes and thus the reliability of MEMS/NEMS and BioMEMS/BioNEMS devices. The tribological needs of these devices were addressed as well.

Tanner (2009) reviewed the major successes in different MEMS devices from a reliability point of view. The reliability issues of a variety of MEMS devices (such as pressure sensors, ink jet printhead, micro-mirror arrays, and RF switches and resonators) were discussed. It was concluded that the MEMS products are fulfilling expectations in terms of performance, cost, and reliability. In another review work, Fonseca et al. (2011) discussed and reviewed the reliability issue in the context of MEMS functionality. They also briefly reviewed the failure mechanisms for MEMS devices.

Table 9. The list of articles reviewed on each subject

**5.1.2 Failure Mode and Mechanisms:**

Failures in MEMS and NEMS can be thermal, mechanical, chemical, electrical, and biological (for BioMEMS and BioNEMS) related, or combinations of them. In this section, we review the most common failure modes of MEMS and NEMS, and introduce research on modeling these failure modes, including wear and degradation, adhesion and stiction, deformation, electrical failure modes, and finally, packaging and contamination.

**5.1.2.1 Wear and Degradation**

Tanner et al. (2000) at Sandia National Laboratories conducted a comprehensive experimental study on reliability of MEMS devices by developing a testing infrastructure including hardware (SHiMMeR, and SHiMMeR Lite), software (Super μDrive), and failure analysis techniques. They designed and characterized reliability test structures to investigate failure modes in various environments, i.e., temperature, humidity, vibration, shock, and storage. The causes of failures for both non-operating and operating MEMS devices were studied. The results of their study indicate that the wear of the polysilicon rubbing surfaces is the dominant failure mechanism for operating MEMS devices. Later on, Tanner et al. (2001) tested the reliability of a MEMS Torsional Ratcheting Actuator (TRA) under stress condition, and developed a wear model similar to the one for microengines. Different dominant failure modes for TRAs under different frequencies were observed: wear of rubbing surfaces resulting in stiction between the alignment guide and the ring gear, when the device is running at a high frequency; “wiggling” phenomena where a TRA cannot perform a complete revolution, when the device is running at a low frequency. This study demonstrates that the TRA is a reliable substitute for the microengine. Namazu & Isono (2003) studied the impacts of specimen size, frequency, and temperature on fatigue lives of nano-scale single crystal silicon (SC-Si) and
silicon dioxide (SiO) wires, for determining reliable design parameters of NEMS.

5.1.2.2 Adhesion and Stiction

The large area-to-volume ratio of MEMS and NEMS leads to conditions that surface forces (such as van der Waals, capillarity, and hydrogen bonding) become dominant over inertial forces, which results in an abrupt stop or failure during operation (Tayebi, 2005; Williams & Le, 2006). Deb & Blanton (2000) analyzed the impacts of defects resulting from vertical stiction, foreign particles, and etch variations on the resonant frequency of surface-micromachined MEMS. They also analyzed the relationships between different behaviors due to the failure sources. It was demonstrated that a single stuck finger resulting from stiction has a catastrophic influence on resonant frequency of the device.

5.1.2.3 Deformation

Bergers et al. (2011) developed a mechanical experimental methodology to analyze the size effect on the time-dependent deformation of metallic MEMS as an important failure mode of the device. The deformation was measured based on surface topographies using the confocal optical profilometry. Simple image correlation and levelizing algorithms based on elastic beam theory were applied to correct profilometer stage drift, and improve the measurement precision, resulting in 3 nm precision, corresponding to 7% of the surface roughness. Lockwood et al. (2012) studied and proposed a new experimental testing method for deformation behaviors of polysilicon components of MEMS. Nanoscale polysilicon beams were experimentally deformed and analyzed using finite element analysis for fixed-fixed and portal-frame-type beam geometries. This study demonstrates that the mechanical performance of nanoscale polysilicon beams significantly depends on both geometry and dimension of the beam support.

5.1.2.4 Electrical Failure Modes

Apart from mechanical design, the electrical design of MEMS/NEMS devices has a great impact on their performance and reliability. Shea et al. (2004) provided guidance for selecting the gap and dielectric properties that balance performance and reliability of MEMS devices. They discussed the relationship between leakage currents and the accumulation of quasistatic charge in dielectrics, along with techniques to reduce the charging and the associated drift in MEMS devices. Their research demonstrates that the electrode geometry and the conductivity of the dielectric are two important parameters for electrical design and reliability of MEMS.

5.1.2.5 Packaging and Contamination

Economic and reliable packaging plays a critical role in successful commercialization of MEMS and NEMS. Because of rapid development of micro and nano technologies, the conventional packaging needs to be improved to meet the requirements of complex system integration methods in terms of higher reliability and performance, and lower cost. Lee et al. (2003) discussed several challenging issues including reliability in MEMS/NEMS packaging, and presented two case studies for MEMS accelerometers and BioMEMS. In addition, they presented another case study on a silicon nanowire biosensor and two examples from molecular biology to demonstrate that the self-assembly of nanodevices is an important issue in NEMS packaging. Pieters (2005) presented a number of wafer level packaging techniques and important reliability analysis tests in order to enable a more cost-effective micro/nanosystems packaging process. Several important wafer level encapsulation techniques such as wafer-to-wafer, cap-to-wafer, and thin film encapsulation were presented, and reliability analysis techniques such as delamination analysis, motion analysis, and life-time testing were described. This study demonstrates the advantage of using wafer level packaging processes for micro/nanodevices including MEMS/NEMS by improving yield and enhancing reliability.

5.1.3 Reliability Issues in Design & Manufacturing

Reliability issues in design and manufacturing of MEMS/NEMS is another challenging problem for further development of these devices (Kuo, 2006). In this section, we review several related studies in this area.

5.1.3.1 Reliability Issues in Design

The pin joint of a microengine is one of the most critical components of this device. Tanner et al. (2002) developed accelerated stress experiments for pin joints, and analyzed the effects of surface coating in order to evaluate its reliability for design purposes. The results suggest that the gap between surfaces is a critical design parameter for reliability of microengines.

Loh et al. (2011) investigated the failure modes of carbon nanotube-based NEMS through a combined experimental-computation. Using in-situ electromechanical characterization, they identified the robust, failure-free region. Dynamic multiphysics models were employed to find the root causes of failure. It revealed that, besides irreversible stiction, another dominant failure mode occurs due to Joule-heating-induced ablation. The findings demonstrate that the usable design space significantly increases by using novel electrode materials such as diamond-like carbon. Using the diamond-like carbon electrodes, not only reduces the stiction, but also completely eliminates the ablation failure mode.

5.1.3.2 Reliability Issues in Manufacturing

Zou et al. (2005) proposed a novel methodology to selectively manufacture micro/nano-textured surfaces for applications in MEMS/NEMS in order to improve
adhesion/stiction and friction performances by reducing the contact area. The results demonstrate that compared to non-textured surface areas, surface textured areas considerably reduce the adhesion/stiction forces and coefficients of friction as well. Gao et al. (2006) proposed a finite element method (FEM) and derived an FEM formulation from the potential energy functional to study the size-dependent mechanical behavior of nanostructures that can be used in NEMS. They considered a surface element to analyze the surface elastic effect. Their study demonstrates that the surface elasticity has a significant influence on the nature of interaction forms and the effective moduli of nanoporous materials.

5.1.4 Reliability Evaluation and Testing
The users of MEMS/NEMS devices require high reliability, as well as quantitative demonstration of the reliability level of such devices. In December 2003, the MEMS Industry Group (MIG) including dozens of companies in MEMS industry (e.g., Intel, Texas Instruments, and Samsung) published their annual report entitled “Focus on Reliability.” The main theme of the report is that “demonstration of reliability is required by customers” (MIG, 2003). With ongoing miniaturization towards MEMS through NEMS, there is an essential need for new reliability engineering concepts and methodologies to measure and evaluate the reliability of these devices. Research work in this area is introduced in this section.

5.1.4.1 Reliability Testing
Experimental verification is the main means for understanding the theoretical concepts and simulation models. There is a need for deploying measurement methods that are capable of evaluating strain fields with nano-scale resolution (Keller et al. 2005). Tanner et al. (1997) presented the first reliability stress test on micromachined surface of microengines developed at Sandia National Laboratories. 41 microengines were stressed at 36,000 RPM and their functionality was inspected at 60 RPM. Infant mortality, low failure rate, and no wearout regions were observed in their study. In subsequent research work by Tanner et al. (2000), the wearout of contacting surfaces was observed and addressed as one of the dominant failure modes of the microengines.

5.1.4.2 Reliability Evaluation
By reducing one dimension of a component to nano-range while the other two dimensions remain large, the obtained nano-component structure is known as a quantum well. By reducing two dimensions of a component to nano-range while one remains large, the obtained nano-component structure is known as a quantum wire. By reducing all three dimensions to nano-range, the obtained nano-component structure is called quantum dot. Sustaining high reliability is inevitable to guarantee the advancement and utilization of today’s nano-systems and nano-components due to the fact that they account for a large proportion of costs of multi-scale systems. Ebrahimi (2008) introduced general methodologies to assess a 1-D nano-component’s limiting reliability from its atoms. Later on, he developed general methodologies for assessing the reliability of 2D nano-components, e.g., nano-films, and nano-discs (Ebrahimi, 2010).

For more literatures related to this section, please refer [26].

6. Conclusions and Future Research: This technical report presents the accomplishments of the collaborative research awards. The research findings for Tasks 2, 3 and 4 since the date September 1, 2013 are described, which essentially contained six manuscripts that have been published or submitted. For the remaining project period of one year, we will continue to explore interesting and challenging research problems for advanced and evolving technologies in all tasks, especially Tasks 4 and 5.

7. Acknowledgements: This research is based upon work supported from NSF under Grant # 0970140 and 0969423.

8. References:


