The Role of Analysis and Testing in the Service Life Assessment of Ion Engines

J. E. Polk*, N.R. Moore*, J. R. Brophy† and D.H. Ebbeler‡
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Abstract

Experience from tests and flights and engineering analysis represent the two sources of information on which to base conclusions on the reliability or failure risk of aerospace flight systems. It is rarely feasible to establish high reliability at high confidence by testing aerospace flight systems or components. The limitations of testing in evaluating failure risk are discussed and an alternate statistical approach which relies on both test experience and analysis to quantitatively assess reliability is outlined. The implementation of this methodology in the service life assessment of ion thrusters is discussed and examples of failure modes being addressed in the NASA Solar Electric Propulsion Technology Application Readiness (NSTAR) program are given.

Introduction

Xenon ion propulsion offers a number of benefits for planetary mission applications. Because of its high efficiency and specific impulse capability, ion propulsion can significantly decrease the propellant mass or deliver more payload mass and, for many planetary missions, substantially decrease the trip time. NASA's 30 cm xenon ion thruster technology is being validated for use in planetary missions in the NASA Solar Electric Propulsion Technology Application Readiness (NSTAR) program. This program is designed to develop the industrial capability to produce flight engine, power processor and propellant feed system hardware and demonstrate that the technology is mature enough for flight applications. This portion of the program is focused largely on providing flight program managers with sufficient information on performance, reliability and spacecraft interactions to give them the confidence to use the technology.

Because ion engines are inherently low thrust devices, extremely long burn times are required. For instance, a typical small body rendezvous mission would require an engine service life of 9000–12000 hours. Demonstrating engine reliability for such a long service life by testing alone would be prohibitively expensive, so lifetime assessment for engine wearout failure modes must rely to a large extent on analysis based on an understanding of the physics of failure. The rationale behind this approach to failure risk assessment is described in detail in the next section, then examples of how the methodology is being implemented in the NSTAR program are presented.

Testing and Analysis in Service Life Validation

Testing to Assess Failure Risk

It is rarely feasible to establish high reliability at high confidence by testing aerospace systems or components. The value of test experience in establishing low failure probability with high confidence for flight configuration systems is limited when testing is halted before failures occur. Failures during testing are often avoided because they can result in the loss of costly
hardware and damage to expensive test facilities. The availability of failure experience for flight hardware is further diminished because failure modes discovered during development testing are corrected by design changes which are intended to render their occurrence highly unlikely during subsequent mission operation. Consequently, test experience for aerospace systems typically does not include failure data for flight configuration systems, but instead consists of tests which are suspended before failures occur, i.e., zero-failure tests.

When the data from tests or flights consists of some number of trials with no failures, the data is a weak information source for reliability demonstration or failure risk assessment. Demonstrating high reliability is equivalent to making statements about the left-hand tail of a failure distribution. Nonfailure test data typically provides very conservative bounding information about location and variability of the failure distribution. However, information about the failure distribution from engineering analysis can be extremely informative about the distribution’s location. Using the approach presented in this paper, the inclusion of engineering analysis allows an improved description of the failure distribution, even with the variability implied by uncertainty in engineering analysis due to sparse information.

The exclusive use of zero-failure tests to establish very low failure risk with high confidence typically implies that an extremely extensive set of test data is required. For example, suppose that each mission simulation test can be treated as an identical independent trial with constant probability of failure \( p \), so that the number of failures is Binomially distributed. The upper confidence limit for \( p \) at confidence level \( C \) is defined by

\[
(1 - p)^n = 1 - C \tag{1}
\]

where \( n \) is the total number of zero-failure mission simulation tests. Reliability \( R \) at a service life \( M \) is the probability of survival to \( M \). For the Binomial failure distribution with \( M \) expressed as a number of missions, reliability is

\[
R = (1 - p)^M \tag{2}
\]

Thus we have

\[
n/M = \ln(1 - C)/\ln R \tag{3}
\]

In order to have even 50% confidence that the probability of failure is no larger than 1/100, 69\( M \) mission simulation tests without failure would have to be conducted.

Failures can be categorized as the consequence of a specific event or the result of accumulated damage. Erosion, fatigue cracking, degradation, and flaw propagation are examples of damage accumulation failure modes wherein failure is a result of the accumulation of aging effects produced by repeated exposure to operating conditions or by environmental parameters which vary cyclically. In contrast, event consequent failure modes are those in which failure is independent of the extent of previous exposure to operating conditions; instead, failure is a consequence of an event such as applied stress exceeding ultimate strength.

Failure prediction for event consequent failure modes is usually much less uncertain than for damage accumulation failure modes, in part because variability of event consequent failure distributions is often much smaller than that of damage accumulation failure distributions. It may be feasible to conduct testing programs to establish low failure probability with high confidence for event consequent failure modes, while such test programs are essentially infeasible for damage accumulation failure modes.

Reliability demonstration by testing a hypothetical component with an event consequent failure mode with failure parameter \( X \) having a Weibull distribution, \( W(\eta, \beta) \), is illustrated in the following. Consider that \( N \) items are tested and that the maximum value of the failure parameter experienced during each test is \( x_1, \ldots, x_N \). Reliability \( R \) at service life \( M \) is the probability of surviving the largest value of the failure parameter experienced during \( M \) missions, \( x_{L(M)} = \max(x_{L1}, x_{L2}, \ldots, x_{LM}) \). At 100\( (1 - \alpha) \)% confidence, reliability \( R \) can be derived from a result given in [1].

\[
\ln R(M; x_{L(M)}) = \frac{-x_{L(M)}^2 - a(2r + 2)x_{L(M)}^\beta}{2 \sum_{j=1}^{N} x_j^\beta} \tag{4}
\]

where \( r \) is the number of failed components. The feasibility of demonstrating reliability by a limited amount of nonfailure testing depends on \( \beta \) being fairly large (implying small variation in the failure parameter) and \( x_{L(M)} \) being smaller than the values of \( X \) during testing, \( \{x_j\} \).
Consider two situations in which failure risk for a particular component is being assessed by testing. In one case, 30 components undergo testing at $X = 6000$ and no failures occur. In the second case, 30 components undergo testing at $X = 6000$ with no failures as before, but, in addition, one component fails in a test at $X = 12000$. The variability of event consequent failure modes can, in some cases, be relatively small. For this hypothetical component, $\beta$ is taken as 15, which implies a failure distribution of small variability. The reliability demonstrated at 90% confidence by the test results of the two situations described above, calculated by Eq. (4), is shown in Table (1). These reliabilities are conservative bounds when the maximum value of the failure parameter experienced in some number of missions, $X_{L(M)}$, does not exceed the test values of 6000, 5000 or 4500. This example illustrates the value of failures in demonstrating reliability. Nonfailure tests alone may be adequate to establish acceptably low failure risk for event consequent failure modes, depending on how much larger test values are than the maximum capability required and the variability of the distribution of the failure parameter.

For a damage accumulation failure mode characterized by a Weibull distribution

$$R = \exp[-(M/\eta)^{\beta}],$$

if there are $N$ zero-failure tests of duration $T_1, T_2, \ldots, T_N$, the lower confidence limit for $\eta$ at confidence level $C$ is defined by

$$\exp\left[-\sum_{i=1}^{N} (T_i/\eta)^{\beta}\right] = \exp[-N(T/\eta)^{\beta}] = 1 - C \tag{6}$$

where $\eta$ and $\beta$ are the Weibull location and shape parameters, respectively, and

$$T = \left[\frac{1}{N} \sum_{i=1}^{N} T_i^{\beta}\right]^{1/\beta}. \tag{7}$$

Table 1: Demonstrated reliability.

<table>
<thead>
<tr>
<th>$X_{L(M)}$</th>
<th>No Failures</th>
<th>With One Failure</th>
</tr>
</thead>
<tbody>
<tr>
<td>6000</td>
<td>0.925</td>
<td>0.99988</td>
</tr>
<tr>
<td>5000</td>
<td>0.995</td>
<td>0.999992</td>
</tr>
<tr>
<td>4500</td>
<td>0.999</td>
<td>0.999998</td>
</tr>
</tbody>
</table>

Thus we have

$$T/M = [\ln(1 - C)/N \ln R]^{1/\beta}. \tag{8}$$

If the $N$ zero-failure tests consist of testing $N$ units for the same number of missions, then the duration of each of those unit tests is $T$ missions, the service life is $M$ missions, and the total test experience is $NT$ missions.

Suppose two units are tested for the same number of missions. In Fig. (1), the plot of Eq. (8) under that constraint shows reliability at 95% confidence as a function of the ratio of unit test duration to service life $T/M$ for $\beta = 1$ and $\beta = 2$. For complex systems subject to multiple failure modes, approximating the distribution of system failures by a Weibull distribution with $\beta = 1$ is more plausible than with $\beta = 2$. The system failure distribution problem is discussed in Section 4.8 of [2], and Chapter 1.12 of [3]. Figure (1) shows that to establish a reliability of .999 at 95% confidence, $T/M = 1498$ for $\beta = 1$ and $T/M = 39$ for...
Thus for $\beta = 1$ the unit test duration must be $1498M$ missions and for $\beta = 2$ the unit test duration must be $39M$ missions.

Acceptable service life reliability cannot be established using a feasibly attainable amount of nonfailure operating experience. Furthermore, a risk control procedure based only on nonfailure operating experience cannot establish service lifetime limits that are commensurate with a component's capability, even when an extensive operating history has accumulated. From nonfailure experience alone, it is not possible to distinguish between components with low failure risk and components with high failure risk which have not yet failed. Relying on nonfailure data alone for reliability assessment results in overstatement of failure risk for reliable components and in ineffective risk control for problem components. Probabilistic analysis which incorporates all available information, including engineering analysis as well as test/flight experience, can be used to evaluate failure risk meaningfully for both problem components and reliable components.

It is usually infeasible to design testing programs for the purpose of demonstrating high reliability at high confidence for flight configuration systems using only nonfailure operating experience. The primary purposes of a testing program should be: (1) to reveal failure modes not identified and to provide information to aid in uncovering analysis oversights or errors; and (2) by means of instrumented tests, to adequately characterize those parameters which drive the results of an engineering analysis of failure modes.

**Integrating Analysis and Testing in Failure Risk Assessment**

In the probabilistic approach to assessing service life presented here, experience and analytical modeling are used in a statistical structure in which uncertainties about failure prediction are quantitatively treated. This probabilistic analysis can be performed with the information available at any particular time to obtain a quantitative estimate of failure risk that is warranted by what is known about a failure mode at that time. This probabilistic method is applicable to failure modes which can be described by analytical models of the failure phenomena, even when such models are uncertain or approximate.

![Figure 2: Probabilistic failure risk assessment.](image-url)
Probabilistic failure modeling. This prior distribution can be modified to reflect available success/failure data in a Bayesian statistical analysis. The probabilistic failure risk assessment approach shown in Figures (2) and (3) is discussed in detail by Moore, et al. [4,5,6,7].

Experience includes physical parameter information in addition to success/failure data. Information about physical parameters can be derived from measurements taken during tests or service, from analyses to bound or characterize parameter values, from applicable experience with similar systems, or from laboratory tests. Measurements of physical parameters used in analytical modeling, e.g., grid impingement current or eroded area, can be an important information source in failure risk assessment. Physical parameter information is incorporated into probabilistic failure modeling and is reflected in the prior failure risk distribution.

Success/failure data can be acquired from testing or service experience. The failure risk distribution resulting from the combination of the prior distribution and the success/failure data is the description of failure risk which is warranted by the information available. As additional information regarding governing physical parameters becomes available it can be incorporated into analytical modeling to obtain a revised prior failure risk distribution. Additional information in the form of success/failure data can be processed by the Bayesian statistical analysis of Fig. (2) to update the prior failure risk distribution using the procedure given by Moore, et al. [6,7].

The analysis procedures used in probabilistic failure modeling, shown in Fig. (3), are directly derived from deterministic methods for analyses of a particular failure mode. The features of this methodology essential to the rational evaluation of failure risk are: (1) inclusion of information from engineering analysis and operating experience, (2) analytical modeling of failure phenomena based on mechanics or physics, (3) representation of the uncertainty about analytical models and their governing parameters, including uncertainty due to both intrinsic variation and limited information, and (4) consideration of failure risk over service life.

A driver for which uncertainty is to be considered must be characterized by a probability distribution over the range of values it can assume. That distribution expresses uncertainty regarding specific driver values within the range of possible values. A driver probability distribution must represent both intrinsic variability of the driver and uncertain knowledge or limited information on which to base the driver characterization.

Stochastic drivers are characterized by using the information that exists at the time of analysis. If driver information is sparse, the probabilistic characterization of such a driver must reflect that sparseness. If extensive experimental measurements have been performed for a driver, its nominal value and characterization of its variability can be inferred directly from empirical data. However, if little or no directly applicable empirical data is available, analysis to characterize a driver or experience with similar or related systems must be used. Driver distributions must not overstate the precision implied by the available information.

Some general guidelines for characterizing stochastic drivers have emerged from case studies conducted to date as given in Moore, et al. [5,6]. For drivers which have physical bounds, such as controlled dimensions or parameters with physical upper limits,
the Beta distribution parameterized with location, shape, and scale parameters has been successfully used. If only bounds are known, a Uniform distribution is appropriate. For a driver whose variation can be thought of as due to the combined influence of a large number of small independent effects, the Normal distribution can be used. Past experience in characterizing a particular driver such as a material property may suggest the use of a particular distribution, for example, Weibull, Normal, or Lognormal.

A hyperparametric structure for driver distributions has been found useful in describing available information about a driver. For example, to characterize a driver, information from engineering analysis might be used to establish upper and lower bounds. In order to capture the fact that the mean value may not be known with certainty, the mean value may be represented by a Uniform distribution between the upper and lower bounds. This Uniform distribution is the hyperdistribution associated uncertainty about the true mean value, and its parameters are the associated hyperparameters.

Monte Carlo simulation has been used as the principal computational method in probabilistic failure modeling because it is a general method that can be used with failure models of any complexity. Continually increasing computer power due to improving hardware and software is steadily expanding the practical application of Monte Carlo simulation. Efficient Monte Carlo techniques can be used to reduce the number of simulation trials when computational time is an issue. Certain analysis methods such as plasma particle simulation models, may be too computationally intensive for practical use in Monte Carlo simulation. However, the output of these models can be represented as response surfaces over the range of variation of significant parameters, see Moore, et al. [4]. The uncertainties of response surface representations must be treated as drivers if significant.

Alternative computational methods, for example, FORM/SORM, see Madsen, et al. [8], may fail to give accurate results for problems in which significantly nonlinear models are employed and driver uncertainty is large. Computational methods are discussed further by Moore, et al. [7].

Implementation in the NSTAR Program

General Approach

As discussed above, it would be prohibitively expensive to characterize the engine failure probability by testing alone. In addition, problems encountered in the first long duration test of the NSTAR engine led to design changes which essentially invalidate that test as a demonstration of the engine service life capability [25]. Finally, most of the interesting planetary missions require that the thruster be throttled over a large power range. Without understanding the processes involved in the dominant failure modes, test experience with a particular design at one operating point cannot be generalized to other designs or throttle points. The NSTAR program is therefore relying heavily on probabilistic failure analysis to demonstrate that the engine will operate reliably for the service life required for the planetary missions of interest.

The initial focus of the service life assessment for the NSTAR engine was on known engine failure modes such as accelerator grid structural failure [13]. As a part of NSTAR, a number of long duration tests are being conducted to expose additional damage accumulation failure modes. Where necessary, additional experiments are being performed to identify physical mechanisms responsible for wear observed in the long duration tests. Models of the physical processes governing the important failure mechanisms are being developed. Additional testing and analysis are being pursued to refine these models and characterize the governing parameters. This information is then integrated into the probabilistic analysis framework outlined above to determine the engine failure probability as a function of lifetime. The primary product of this effort is the quantitative assurance that a particular failure mode does not pose unacceptable risk over the lifetime required for a particular mission. In addition, the analysis is being used to identify the primary drivers that limit lifetime, information that can be used to guide engine improvements. This effort also provides the tools required to assess the impact of design changes or changes in operation on engine reliability. Three potential failure modes have been chosen to illustrate this approach. The identification of the failure mode, cur-
rent understanding of the mechanisms involved, and recommended experimental and analytical strategies to implement the probabilistic failure analysis for the three examples are discussed below.

Accelerator Grid Structural Failure

This example is discussed first because the failure mechanism is relatively well understood and quantitative analyses of service life based on a model of the physics of failure have been completed. Accelerator grid mass loss due to ion bombardment is observed at some level in all ion engines, but structural failure caused by sputter erosion was identified as a potential failure mode in inert gas ion thrusters after several long duration wear tests [9,10]. Some of the slow ions created downstream of the accelerator grid by charge exchange reactions between the beam ions and neutral atoms strike the grid and remove material by sputtering. In space, the engine is the only source of neutral gas. An inevitable consequence of propellant utilization inefficiencies is the leakage of neutral atoms from the discharge chamber, although the density of these source atoms drops rapidly downstream from the engine propellant flow rate and the pumping capability of the facility. Most of the charge exchange ions collected by the accelerator grid are generated downstream and are focussed into the center of the webbing surrounding the apertures. This creates a hexagonal wear pattern composed of grooves located between apertures with deeper pits at the intersection points between three holes, as shown in Figs. (4) and (5). After the pits wear through the grid, penetration proceeds along the channels until complete structural failure occurs.

To study the geometry of the grid erosion and the approach to structural collapse a two grid, 30 cm ion accelerator system was tested to failure [12]. The wear rate was artificially accelerated by operating at a vacuum chamber pressure of $3.5 \times 10^{-3}$ Pa and a comparatively high accelerator grid voltage of -500 V. The engine actually failed after 633 hours when the grids were shorted together by a flake of material that had been sputter-deposited on the screen grid from the accelerator grid, but the grid was also very near structural failure. Inspection of the grid after 145 hours and 633 hours showed that the erosion was largely confined to the pits and grooves pattern and that the fraction of the webbing area encompassed by this pattern (determined by the diameter of the pits and the width of the grooves) was approximately constant with radial position on the grid and over the course of the test. These observations led to the definition of a failure criterion based on the removal of all of the material in the pits and grooves pattern in the center of the accelerator grid where the impingement current density is highest, with the assumption that the grooves are square channels with straight, vertical walls. The total mass loss at failure is then given by the expression

$$M_f = A_b t_o \rho (1 - \phi_o) \alpha f_o,$$

where $A_b$ is the active grid area, $t_o$ is the grid thickness, $\rho$ is the density, $\phi_o$ is the physical open area fraction of the grid, and $\alpha$ is the fraction of the webbing area covered by the pits and grooves pattern. The accelerator grid current flatness parameter $f_o$ is defined as the ratio of the average impingement current density or mass loss per unit area to the peak current density or mass loss per unit area and accounts for the nonuniform radial distribution of mass loss. Using measured values of $\alpha$ and $f_o$ for this test, this equation yields a total mass loss at failure of 63.5
an acceptable local mass loss in the erosion pattern in the center of the grid and the mass loss rate was calculated based on the local current density and the average net sputter yield. The average local current density was determined from the total beam current, the ratio of impingement current to beam current, the grid current flatness parameter and the eroded area fraction. This approach accounts for the radial variation in current density and the concentration of impingement current in the pits and grooves pattern, but does not distinguish the current density distribution between pits and grooves. The mass loss per unit area in the pits and grooves pattern in the grid center $M_{pg}$ can then be expressed as

$$M_{pg} = \frac{m_p}{e} \left( \frac{\lambda Y}{(1 - \phi_a) \alpha f_a} \right) \left( \frac{J_a Y t}{A_0} \right).$$

(11)

where $t$ is the operating time. The cross-sectional area of the erosion pattern was also modified by a stochastic shape parameter $\lambda_s$. This allowed variation of the eroded geometry at failure from a channel with a square cross-section to a trench with a parabolic cross-section. This reflects the experimental observation that the cross-section in the bridge region is approximately parabolic when the bridge is finally penetrated in the center while the cross-section in the pits at failure is much closer to square. The failure criterion is defined as the point at which sufficient mass has been removed from the erosion pattern with the chosen cross-sectional area that the bridges surrounding the center hole collapse.

The model drivers such as grid current flatness parameter and the ratio of impingement current to total beam current were assigned distributions representing the uncertainty or intrinsic variability in these parameters. These distributions were calculated from test data available at the time or based on reasonable estimates of upper and lower bounds. The eroded area fraction was treated parametrically, because insufficient information was available to bound the range of possible values. The net sputter yield parameter $\lambda_Y$ was given a conservative value of unity, reflecting a lack of knowledge about its variability. This model was then used in a Monte Carlo simulation to calculate the time to failure based on samples drawn from these distributions. The results of approximately 20000 such calculations were used to calculate the probability of failure as a function of

Figure 5: Photograph showing the erosion structures typically found on the upstream surface.
I three grids used in long duration tests show that the groove width increases slightly with radius [14]. This radial variation in \( \alpha \) must be incorporated into the model. The eroded area fraction is a measure of the degree of ion focusing, so these results suggest that the decreasing beamlet current density with increasing radius results in less focusing of the impingement ions. A 3D particle-in-cell (PIC) code developed by ERC, Inc. which simulates the distribution of impingement ion current density around a single aperture [15] will be used in an attempt to simulate this behavior. In addition, a series of tests with varying beamlet current densities may be performed to help validate the PIC code. The erosion pattern can be visualized in relatively short duration tests by coating the downstream face of the accelerator grid with a thin layer of an easily eroded metal or alternating layers of two different metals [16]. These sources of information can then be used to specify a value of the eroded area fraction which properly scales the total mass loss into a local mass loss in the grid center.

Figure 6: Probabilistic life assessment of the NSTAR engine for operation at 2.3 kWe.

operating time to assess the service life capability of the 30 cm diameter xenon ion engine proposed for the NSTAR program. The results are shown in Fig. (6). At a power level of 2.3 kWe the service life with a 50% failure probability, which corresponds to the lifetime predicted deterministically using nominal values of the input parameters, is 21000 hours for a nominal eroded area fraction of 0.4. The service life achievable with a one-in-a-thousand failure risk, however, is only 11600 hours. This shows that the lifetime with acceptable failure risk can be substantially lower than that estimated deterministically using nominal parameter values. This is the primary product of the service life assessment effort; a quantitative evaluation of whether the engine can provide the required service life at an acceptable failure risk level for a given failure mode. The probabilistic model was also used to perform a sensitivity analysis which revealed that uncertainties about the eroded geometry at failure, the net sputter yield and the impingement ion current in space are the primary risk drivers.

Current efforts to refine the model are focussed in two areas: the relaxation of several model assumptions and improved driver characterization. The model assumes that the eroded area fraction is constant in order to scale the macroscopic mass loss into the local mass loss in the grid center. However, more detailed measurements of the erosion geometry on three grids used in long duration tests show that the groove width increases slightly with radius [14]. This radial variation in \( \alpha \) must be incorporated into the model. The eroded area fraction is a measure of the degree of ion focusing, so these results suggest that the decreasing beamlet current density with increasing radius results in less focusing of the impingement ions. A 3D particle-in-cell (PIC) code developed by ERC, Inc. which simulates the distribution of impingement ion current density around a single aperture [15] will be used in an attempt to simulate this behavior. In addition, a series of tests with varying beamlet current densities may be performed to help validate the PIC code. The erosion pattern can be visualized in relatively short duration tests by coating the downstream face of the accelerator grid with a thin layer of an easily eroded metal or alternating layers of two different metals [16]. These sources of information can then be used to specify a value of the eroded area fraction which properly scales the total mass loss into a local mass loss in the grid center.

In addition, the model assumes that all mass loss is confined to the pits and grooves pattern and ignores the detailed distribution within this pattern. However, more recent measurements indicate that the detailed distribution of mass loss cannot be ignored, particularly the contribution that comes from chamfering on the upstream edges of the pits once they wear through the grid [14]. While this undercutting clearly does remove material from the upstream side of the bridges connecting the ringlets surrounding the holes, it also removes material from the ringlets themselves. The material lost in the ringlets does not appear to contribute to the approach to structural collapse. This effectively results in mass loss which is outside the pits and grooves pattern as it is defined in this model, so the results are probably overly conservative. Relaxing this conservatism will require the development of tools to predict the distribution of mass loss in the hexagonal pattern and in the undercutting, perhaps based on ERC's PIC code.

Reduction of the uncertainty in the driver characterization translates directly into a reduction in the calculated failure risk. The eroded area fraction was treated parametrically rather than characterized by a probability distribution because so little is currently known about how this parameter varies with operating conditions and environment. There are three
sources of information that will be exploited to properly characterize this important driver. Data generated in long duration tests with the NSTAR engine are invaluable. An eroded area fraction of 0.37 was measured on the accelerator grid used in a 2000 hour test of the first engineering model thruster. However, subsequent development tests with this grid set at a higher tank pressure and possibly with a lower propellant utilization efficiency appeared to cause a broadening of the erosion pattern, suggesting that \( \alpha \) is sensitive to operating conditions or test environment [17]. Additional data will soon be available from an ongoing test of this engine with a new grid set which is scheduled to last 1000 hours. The 3D PIC code will also be used to examine the sensitivity to operating and environmental conditions. Finally, additional tests using thin film diagnostics will be used to experimentally characterize the sensitivity to beamlet current density, tank pressure and propellant utilization efficiency as well as suspected drivers such as grid voltage and geometry. Once the code is validated by comparison with test data, it can be used to generate a response surface which describes the variation in eroded area fraction with variation in these other drivers, which are more easily characterized. This can then be incorporated into the probabilistic analysis to assess the impact of uncertainties in these parameters.

The sputter yield parameter \( \lambda_Y \) was also not treated probabilistically, but was given a conservative value of unity. The model should incorporate a distribution which characterizes the range of values that this parameter might span. Test data available now indicate values ranging from 0.6 to 1.0, but the variation is not understood. A simulation code which tracks the trajectories of sputtered atoms in an erosion groove which is evolving with time is being developed to study the importance of redeposition in determining the net sputter yield. This will be combined with the 3D PIC simulation of impingement ion spatial and angular distribution to study the variation in sputter yield from that at normal incidence. These tools will be used with the available test data to define the variation in net sputter yield.

The magnitude and macroscopic distribution of impingement current also need to be updated to reflect more recent data. The ratio of impingement current to beam current for in-space conditions was originally specified in the model using data from tests with mercury propellant at very low tank pressures, with the expectation that the data for xenon would be similar. Subsequent tests of the NSTAR engine at very low facility pressures indicate values which are almost two times higher [17]. The radial variation of erosion in space was assumed to be the same as the beam current density radial profile, which was specified based on previous experience. Actual measurements of the beam profile for the NSTAR engine are now available that can be used to improve the characterization of this parameter. However, data show that the erosion profile can be significantly broader than the beam profile in ground tests [14]. A PIC model simulating the interaction of the entire beam with neutrals from the engine and ambient facility gas was used to study the distribution of charge exchange ion backflow. Initial results confirm qualitatively that the broadening is largely an artifact of ground based testing [14,18]. This code is being further developed to quantitatively calculate the erosion profile broadening to assist in interpreting ground test data and to help specify the driver distribution for the probabilistic failure analysis.

The model currently assumes that the incident ions strike the grid with an energy corresponding to the grid potential. The energy should be characterized by a distribution to incorporate two effects. First, ions created upstream of the neutralization plane will have somewhat lower energies. This effect is modeled in the ERC code and can easily be incorporated as a distribution on energy for a given operating condition. Second, preliminary results of plume modeling indicate that the potential distribution in the beam depends strongly on the ambient electron density and may be much higher in space where the ambient plasma density is lower. Charge exchange ions generated in the beam may therefore start from a potential which is several tens of volts higher than the space plasma potential and strike the grid with more energy than is observed in ground tests. The variation in beam potential will be modeled with the PIC simulation of the entire beam.

**Cathode Orifice Plate and Heater Erosion**

Cathode orifice plate and heater wear was only recently recognized as a potential engine failure mecha-
anism, so it is not as well understood as the accelerator grid erosion problem. The analysis of this failure mode is still at the very early stages of information gathering to understand the erosion mechanisms. In the initial NSTAR engine design, the main cathode had no keeper and was ignited by coupling directly to the anode. During an interruption of the first NSTAR endurance test after 876 hours of operation severe erosion of the cathode assembly was noted. While the orifice and chamfer appeared to be undamaged, the outer edge of the orifice plate was eroded to the extent that the electron beam weld bead on the periphery was no longer visible. Damage to the heater coil outer sheath on the portion of the first turn facing downstream was also observed. The entire cathode assembly was replaced and the 2000 hour test was completed. Similar erosion was found on the cathode assembly used in the second test segment. Subsequent examination of the cathode used in a 900 hour test of a laboratory model ring cusp thruster at 5 kW revealed similar damage to the orifice plate, but not the heater coil. Surprisingly, no orifice plate or heater erosion was observed on four cathodes which had each been tested for 2000 hours at a current of 12 A in a test setup which simulated the discharge, but with no beam extraction [19]. A cathode with a keeper which was tested in a similar geometry for 5000 hours at a current of 25 A also showed no significant erosion on the downstream face of the orifice plate or the heater [20]. Recent tests of 3 cathodes with enclosed keepers in 15 cm ion engines run at 0.5 kW also revealed no damage.

These test data indicate that orifice plate and heater erosion is very sensitive to operating conditions or geometry. No signs of melting were found in the damaged regions, so the erosion is evidently due to sputtering. There must, therefore, be a source of high energy ions downstream of the cathode orifice plate. The erosion pattern on the heaters from the 2000 hour is consistent with a distributed source on the cathode axis extending as far as 10 mm downstream of the cathode with most of the ions generated within 1 mm of the orifice plate face. Brophy and Garner proposed the existence of a jet of high energy ions emanating from the cathode to explain severe erosion of downstream components which had been observed in several tests [20,10]. Subsequent tests confirmed the existence of this jet and characterized the energy distribution of the ions downstream of the cathode. The same process that generates high energy ions directed downstream may also produce the flux of ions responsible for the orifice plate and heater erosion. Two mechanisms for ion acceleration to energies greater than the discharge potential have been proposed. Friedly and Wilbur suggested that the ions are accelerated from a potential hump close to the orifice [21]. Latham proposed that the current density in the orifice may be sufficiently high to create a self magnetic field that results in magnetogasdynamic acceleration of the cathode plasma [22]. However, no attempt has yet been made to develop a quantitative model of the ion acceleration process based on either of these mechanisms.

Although this failure mechanism is not well understood, several design changes have been implemented to mitigate cathode and heater erosion based on the available evidence. The weld which joins the orifice plate to the cathode tube has been relocated to the side of the cathode so that it is shielded from the high energy ion flux by the orifice plate. This design change will clearly extend the life of this critical joint. The cathode heater has also been retracted so that the downstream surface is 1.7 mm upstream of the cathode tip. In long duration tests prior to the 2000 hour test the heaters were retracted and experienced no erosion, presumably because they were also shielded by the orifice plate. Finally, the cathode assembly has been enclosed within a cylindrical keeper. The gap between the downstream face of the cathode orifice plate and the upstream side of the keeper orifice plate is as small as possible in an attempt to shield the orifice plate from the high energy ions. However, if the ions are actually generated within 1 mm of the orifice, the keeper will not act as a physical shield. There is some evidence that a keeper can increase the current density of high energy ions in the downstream jet [23], and no orifice plate erosion has been observed in tests in which a keeper was used. However, given the apparent sensitivity of this phenomenon to operating conditions and test geometry, the applicability of these test data to the NSTAR cathode design is not clear.

The effectiveness of these design changes is being evaluated in a 1000 hour engine test which is now being conducted. If successful, this test will provide confidence in the new design. However, without un-
derstanding this wear process no conclusion can be drawn about its impact on engine reliability. The only way to guarantee that this failure mode will not pose an unacceptable risk over the service life required for planetary mission applications is to understand the physical process and how sensitive it is to geometric, environmental, and operating parameters which may vary from those in the test.

To quantitatively assess the impact of cathode and heater wear on engine lifetime a model of the physical processes must be developed. Several experiments are recommended to guide the development of this model. The highest priority is to determine the mechanism of ion acceleration. Several experiments can be conducted to distinguish between acceleration from a potential hump and magnetogasdynamic acceleration. As Friedly has observed, a potential hump should produce ion flow which is more or less spherically symmetric [21]. In contrast, magnetogasdynamic processes would generate a more collimated jet. The erosion patterns observed in [24] indicate that the ion current density is peaked on the centerline, but the actual ion current density and energy distribution should be mapped spatially with a retarding potential analyzer (RPA) or an electrostatic energy analyzer (ESA). In addition, a potential hump will accelerate multiply charged ions to velocities which are proportional to their charge state, while magnetogasdynamic processes will result in velocities which are largely independent of charge state. This can be resolved with an ESA coupled with a mass spectrometer to distinguish charge state. If insufficient multiple ions are generated in the cathode plasma, the xenon could be seeded with a more easily ionized gas to generate higher charge states for this diagnostic technique. Finally, rapid probing of the downstream plasma with small floating probes or nonintrusive optical measurements of the electric field distribution could be used to verify the existence of a potential hump.

The structure of the acceleration zone should then be determined to aid in model development. The radial and axial extent of the zone could be measured with probing or by using an ESA with a very narrow acceptance angle. The current to the orifice plate should also be measured, perhaps by coating the downstream face with a thin layer of ceramic and a thin metal layer that could be biased to collect the ion current. The ion energies measured in the cathode jet are typically less than 100 eV. Because there are currently no experimental data available on sputter yields for xenon at such low energies, an experiment is now being planned to measure the yields for a number of materials of interest to the NSTAR program. Finally, the effect of an enclosed keeper must be evaluated. In particular, it is important to establish that the cathode erosion problem does not simply become a cathode keeper erosion problem. In addition, if the keeper does increase the current density of high energy ions on the centerline, as suggested in [23], the impact of this on screen grid erosion should be evaluated.

Experiments such as these should permit identification of the acceleration mechanism and help characterize how the physical processes vary with geometry and operating conditions. In parallel, efforts to develop a quantitative model of these processes that can be used in risk assessment should be pursued.

**Grid Shorting by Metal Flakes**

Grid shorts caused by flakes of sputter-deposited metal have long been recognized as a potential engine failure mechanism. This was a particular concern for the NSTAR ion engine after the discovery of relatively thick films and large flakes of sputter-deposited molybdenum in the discharge chamber after the 2000 hour wear test [25]. Thick molybdenum films, evidence of film formation were also found on the downstream surface of the screen grid after the test-to-failure [12].

The general physical processes involved in flake formation are more well-understood than those in cathode erosion and models of film failure have been developed for other applications, but there has been as yet no effort to model these processes in ion engines. There are two primary sources of sputtered material in the engine that contribute to the growth of films in the interior. When the pits in the accelerator grid wear completely through the electrode, material sputtered from the walls of the pit or from the upstream chamfer may be deposited on the screen grid or inside the discharge chamber. This was undoubtedly the source responsible for the films found on the screen grid in the test-to-failure. In the 2000 hour test of the NSTAR engine the primary source of ma-
terial for the films formed in the discharge chamber was the upstream surface of the screen grid, which was subject to severe erosion in this test. This erosion process is not well-understood, but the pattern of mass loss suggests that it is related to the structure of the sheath that forms upstream of the apertures. Evidently some ions that enter the sheath are not focussed through the grids and strike the edges of the holes. This process was aggravated in this test because the screen grid was electrically isolated and floated several volts negative of the cathode potential. In addition, the propellant efficiency may have been higher than expected during some fraction of the test, resulting in the generation of a high percentage of double ions.

Material sputtered from the accelerator grid most likely follows line-of-sight trajectories to the downstream face of the screen grid. The same was thought to be true of grid material that entered the discharge chamber. However, after the 2000 hour test grid material was found on a surface of the downstream magnet retainer which faces upstream and has no direct line-of-sight to the grids. It therefore seems probable that molybdenum atoms from the grids may form negative ions in the discharge and follow trajectories to anode potential surfaces which are influenced by the magnetic field. After thousands of hours, deposition of erosion products can form films which are many microns thick.

These films do not represent a danger unless they crack and become delaminated from the substrate. Intrinsic stress or thermal stresses can cause cracks to initiate at voids or defects in the film. The magnitude of thermal stresses depends on the difference in the coefficient of thermal expansion of the film and the substrate, while intrinsic stresses depend on the microstructure of the sputter-deposited film. The microstructure, in turn, is determined by the deposition conditions. For relatively low temperature deposition on surfaces which are in general not subject to ion bombardment such as the downstream face of the screen grid and anode potential surfaces the films are likely to be porous and have residual tensile stress [26]. Tensile films are most likely to fail by crack formation through the film, along the film-substrate interface, or by cracks which propagate through the film or substrate parallel to the interface. Crack formation is governed by the stresses, film thickness and the fracture resistance of the materials [27].

The film failure mechanism responsible for flake formation in ion engines has not been identified. The two photographs of the screen grid from the test-to-failure in Figs. (7) and (8) suggest a possible film failure process, however. The first photograph in Fig. (7) shows the thick deposits on the downstream face and walls of the grid apertures. Delamination appears to initiate on the sharp edges of the holes, and is particularly advanced where the webbing is thinnest between two holes. In virtually all of these locations the film has debonded and curled upward. In the center of the photograph is a region where a flake has formed and fallen off. The second photograph in Fig. (8) is an enlarged view of another flake site. The film evidently fails first at the thinnest part of the webbing and then becomes debonded, forming a triangular flake. This process dictates the scale of the flakes, and it is clear that they are large enough to bridge the gap between the electrodes. This photo also shows evidence of delamination in the films deposited on the walls of the holes. In other locations on this grid these films have peeled back and protrude into the aperture, forming structures which could de-focus the beamlet and cause increased accelerator
of the engineering model thruster.

To properly evaluate the failure risk associated with flake formation, the following combination of testing and analysis is recommended. A quantitative description of the sputtering rate must be developed. A model of accelerator grid erosion will be developed as a part of the analysis of that particular failure mode. Sputtering of the screen grid is itself a potential failure mode, and should be modeled to assess that risk as well as specify the material generation rate for flake modeling. A PIC code simulating the structure of the sheath upstream of the screen grid apertures and the ion acceleration process should be developed to determine the trajectories of double and single ions. This could be coupled with data on sputtering yields at low energies to calculate a mass loss rate.

To understand the film growth rate the deposition process must be studied. For material transport from the accelerator grid to the screen grid line-of-sight trajectories can be assumed. The difficulty is determining the initial direction of atoms sputtered from the accelerator grid. The distribution of deposits on grids used in long duration tests should be examined in more detail to determine if simple approximations to the sputtered material flux are justified. Otherwise more detailed models of sputtering in the walls of the pits and upstream chamfers will be required. Similar analyses of films from the discharge chambers of engines used in long duration tests should be performed. As a first approximation the flux of material from the screen grid can be assumed to follow a cosine distribution with a magnitude that varies radially in proportion to the beam current. A more detailed model would require a description of the negative ion generation rate and how they are affected by the magnetic field. It may also be possible to develop accelerated wear test methods, perhaps based on biasing the screen grid negative, that could be used to study the erosion and film deposition rates. However, the erosion processes must be sufficiently well understood to interpret the mass loss and deposition data obtained under accelerated wear conditions and apply it to normal engine operation.

Finally, the mechanisms of film cracking and delamination must be identified and modeled. The first step is to study the microstructure of flakes found on the screen grid and discharge chamber surfaces of en-
gines used in long duration tests. This may reveal the distribution of voids and defects, provide clues to crack initiation and propagation mechanisms, and yield data on the film properties which determine fracture resistance. Further experiments with films deposited under conditions like those expected in the engine should be performed to measure the intrinsic stress, further characterize the film microstructure and material properties and determine how the films behave on textured surfaces. Once the fracture mechanism is identified, a model of the process that can be incorporated in the probabilistic analysis framework must be developed to address the risk of flake formation. This effort could leverage an on-going program at JPL to apply fracture mechanics to the analysis of crack growth and propagation in thin-film based multichip module substrates [28]. Many of the analytical and experimental tools developed in this program are directly applicable to the problem of flake formation in ion engines.

Conclusions

Under the constraints on information availability typically encountered in assessing failure risk for aerospace systems, a meaningful evaluation of failure risk must be based on both test experience and engineering analysis. The probabilistic failure analysis methodology is a prescribed statistical structure for evaluating failure risk and identifying risk control actions that incorporates test experience and engineering analysis. This approach can be applied at any time in the design, development, or operational phases of a program to evaluate failure risk based on what is known about critical failure modes and to rank alternative actions to control risk, thus enabling the more effective allocation of limited financial resources.

This methodology is being used to provide planetary mission planners with the assurance that the NSTAR ion engine will operate reliably over the required service life as a part of NASA's NSTAR program. A full probabilistic analysis of accelerometer grid structural failure has been completed and refinements to the analysis are being implemented. Cathode orifice plate and heater erosion and flake formation represent two additional failure modes that have not yet been quantitatively analyzed. Recommendations on ways to gather sufficient information to quantitatively assess the failure risk associated with these phenomena have been outlined. This same general approach will be used on additional failure mechanisms identified in subsequent wear tests to minimize the failure risk.

Acknowledgements

The authors would like to acknowledge the participation of M. Patterson in useful discussions of cathode orifice plate and heater wear. The research described in this paper was conducted at the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the National Aeronautics and Space Administration.

References


