

DISSERTATION

PERFORMANCE AND LIFETIME SIMULATION OF ION THRUSTER OPTICS

Submitted by

Cody Coffman Farnell

Department of Mechanical Engineering

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2007

UMI Number: 3266391

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3266391

Copyright 2007 by ProQuest Information and Learning Company.

All rights reserved. This microform edition is protected against unauthorized copying under Title 17, United States Code.

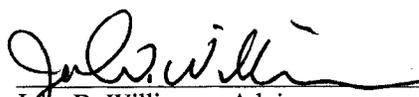
ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

COLORADO STATE UNIVERSITY

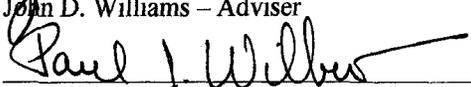
April 4, 2007

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY CODY C. FARNELL ENTITLED PERFORMANCE AND LIFETIME SIMULATION OF ION THRUSTER OPTICS BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work



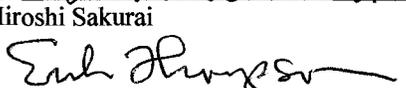
John D. Williams – Adviser



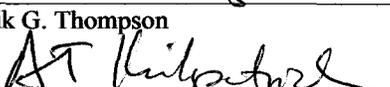
Paul J. Wilbur – Co-Adviser



Hiroshi Sakurai



Erik G. Thompson



Allan T. Kirkpatrick – Department Head

ABSTRACT OF DISSERTATION

PERFORMANCE AND LIFETIME SIMULATION OF ION THRUSTER OPTICS

A simulation code, ffx, was developed to study various aspects of ion thruster optics. Information concerning sheaths, impingement limits, perveance, electric potential, charge exchange, and sputtering is covered. Electron backstreaming and pit and groove wear are discussed in detail as two grid failure mechanisms.

The code was used to study the effects of several parameters on grid performance and lifetime, including grid spacing, aperture diameter, and grid thickness. An evolutionary algorithm was used with the ffx code to design grid sets, utilizing net accelerating voltage and current density as primary input parameters.

Validation of the ffx code was accomplished through comparison to other ion optics codes and to experimental data obtained from both gridlet and full thruster testing. Gridlet test comparisons included simulations of finite aperture grid sets. The NSTAR thruster was studied in detail with regard to lifetime. The methods used for accurate and efficient optics simulation are discussed, including the multigrid method for solving for electric potential.

Cody C. Farnell
Department of Mechanical Engineering
Colorado State University
Fort Collins, CO 80523
Spring 2007

ACKNOWLEDGEMENTS

I would like to thank Professors John Williams and Paul Wilbur for their guidance in my research. I would also like to thank my family. Much of this work has been supported through grants from the NASA Glenn Research Center.

TABLE OF CONTENTS

1	Introduction.....	4
2	Nomenclature	5
3	Ion Thrusters.....	7
3.1	Ion Thruster Operation	7
3.2	Ion Optics	8
3.2.1	Discharge Chamber and Beam Plasmas	9
3.2.2	Grids	9
3.2.3	Voltages	11
3.2.4	Currents	13
3.2.5	Sheath and Neutralization Surfaces	14
3.2.6	Impingement Limits.....	17
3.2.7	Grid Transparency To Ions	20
3.2.8	Perveance Fraction.....	22
3.2.9	Propellant Flow.....	24
3.2.10	Charge Exchange	26
3.2.11	Potential	33
3.2.12	Electron Backstreaming	35
3.2.13	Divergence Angle and Thrust Factor	41
3.2.14	Exhaust Velocity.....	43
3.2.15	Thrust.....	44
3.2.16	Specific Impulse.....	45
3.2.17	Sputtering.....	46
3.3	Grid Lifetime.....	49
3.3.1	Grid Failure Mechanisms	50
3.3.2	A Useful Lifetime Parameter	51
3.3.3	Erosion as a Function of Current	52
3.3.4	Pit and Groove Erosion Rates	60
3.3.5	Screen Grid Film Thickness	65
3.4	Thruster Properties	66
3.4.1	Number of Apertures	66
3.4.2	Beam Current.....	67
3.4.3	Grid Area	67
3.4.4	Flatness Parameter	68
3.4.5	Propellant Utilization Efficiency	69
3.4.6	Lifetime	71

3.4.7	Perveance Limit	72
4	Variable Effects.....	73
4.1	The Example Grid Set	73
4.2	Grid Design Goals	75
4.3	Geometry Effects.....	75
4.4	Parameter Effects	76
4.4.1	Grid Spacing, l_g	77
4.4.2	Accel Grid Hole Diameter, d_a	85
4.4.3	Accel Grid Thickness, t_a	88
4.4.4	Screen Grid Thickness, t_s	91
4.4.5	Accel Grid Voltage, V_a	99
5	Special Investigations.....	107
5.1	One, Three, and Seven Aperture Grids.....	107
5.2	Grid Cusps and Linear Grids	111
5.3	Offset Apertures	116
5.4	Charge Exchange Ion Space Charge.....	120
5.5	Charge Exchange Ion Initial Energy	121
5.6	Fixed Emission Surface	122
5.7	Thruster Current Density	129
6	Grid Design.....	137
6.1	Evolutionary Algorithms	137
6.2	Previous Genetic Algorithms.....	153
6.3	Algorithm For Grid Design	154
6.3.1	Algorithm Guidelines	154
6.3.2	Design Goal	155
6.3.3	Input Values.....	155
6.3.4	Variables.....	156
6.3.5	Variable Ranges.....	156
6.3.6	Fitness Value	159
6.3.7	Algorithm Results.....	160
7	Code Validation.....	170
7.1	Gridlet Tests	170
7.2	Carbon Based Ion Optics (CBIO) Grid Set	170
7.3	NSTAR.....	174
7.4	NEXT	193
8	ffx Ion Optics Simulation Code.....	210
8.1	Particle Tracking	216

8.2	Solution of the Poisson Equation.....	219
8.2.1	Gauss-Seidel and Jacobi Methods	221
8.2.2	Newton's Method	222
8.2.3	Successive Overrelaxation.....	223
8.2.4	Linearization of the Electron Equations	224
8.2.5	Multigrid.....	225
8.2.6	Linear Multigrid Method	225
8.2.7	Nonlinear Multigrid Method.....	229
8.3	Symmetry	233
8.4	Distributions	234
8.4.1	Cosine Distribution	234
8.4.2	Distribution Sampling.....	238
8.4.3	Maxwell-Boltzmann Distribution	241
8.5	Neutral Density.....	243
8.6	Random Number Generators	247
8.6.1	Pseudo-Random Number Generators	247
8.6.2	Quasi-Random Number Generators.....	248
9	References.....	251

1 Introduction

This work expands upon the body of knowledge gained by previous research in the area of ion thruster optics. One of the great utilities of ion optics simulation is that it aids in visualizing the physics behind ion beams. Another advantage of simulation is convenience. After validating a code, it can be used to analyze hundreds of grid designs relatively inexpensively. Simulation and laboratory research are complimentary. Simulation predictions are used to improve physical models, and experimental data is used in turn to improve computational models.

The ffx code is one of many ion optics simulation codes. Each code has advantages and disadvantages, with tradeoffs often being made among speed, accuracy, complexity, and flexibility. One of the advantages of the ffx code is its ability to simulate three dimensional grid erosion. Full lifetime erosion simulations can be performed over the course of a day, resulting in predictions of grid failure due to the onset of electron backstreaming or pit and groove wear for example. These simulations are accurate given a correct set of input conditions, which unfortunately is not usually a simple matter to obtain.

First, ion optics are described with regard to the way in which ions are accelerated to produce thrust. Topics related to this are electric potential, plasma sheaths, and perveance. Then, the life limiting process of grid erosion is described in terms of neutral propellant flow, charge exchange, and sputtering.

In the second section, a closer examination is made to determine how several design parameters affect grid performance and lifetime. Individual parameters, such as grid spacing or grid thickness, can be optimized to increase grid lifetime.

The third section describes new ways in which simulation is being used to study ion optics. Included in this are simulations of grids with finite numbers of apertures, which can be used to study the performance of edge apertures. Another study investigates the performance of, either intentionally or unintentionally, misaligned apertures.

In the fourth section, evolutionary algorithms are described as one way to improve grid design. These algorithms apply the ideas of evolution to evolve a set of, initially random, solutions over many generations to arrive at a better set of solutions under a collection of imposed constraints.

Next, the ffx code is validated against data from other ion optics codes in current use and against data from both small and large scale grid tests. This includes comparisons to the NSTAR and NEXT thrusters, with an interesting analysis of the erosion resulting over the course of the NSTAR Extended Life Test.

Finally, the methods of simulation used in the ffx code are explained. Particular attention is given to the multigrid method as an efficient way to solve the Poisson equation, a nonlinear elliptical partial differential equation, to obtain mesh potential values.

2 Nomenclature

Symbol	Units	Description
α	degrees	Divergence angle
amu	kg/molecule	Atomic mass unit: $1 amu = 1.6605 \cdot 10^{-27} \frac{kg}{molecule}$
A_T	m ²	Grid active area
d_a, d_s	m	Accel and screen grid hole diameter
e	C	Electron charge: $e = 1.602 \cdot 10^{-19} C$
ϵ_0	F/m	Permittivity of free space: $\epsilon_0 = 8.854 \cdot 10^{-12} F/m$
f_p	-	Perveance fraction
ft	-	Thrust factor
F	N	Thrust
I_{sp}	s	Specific impulse
j	A/m ²	Current density
J_a, J_s	A	Accel and screen grid current
J_b, J_B	A	Beamlet and beam current
k	J/K	Boltzmann constant: $k = 1.38065 \cdot 10^{-23} J/K$
l_{cc}	m	Aperture center-to-center spacing
l_e	m	Effective ion acceleration length
l_g	m	Grid spacing
m_e	kg	Electron mass: $m_e = 9.109 \cdot 10^{-31} kg$
m_i, m_n	kg	Ion and neutral molecule mass
\dot{m}	kg/s	Propellant flow rate
n_e, n_i, n_n	m ⁻³	Electron, ion, and neutral number density
P	A/V ^{3/2}	Perveance
ϕ_a, ϕ_s	%	Accel and screen grid open area fraction
ϕ	%	Transparency to ions
ϕ	V	Potential

η	-	Efficiency
q	C	Ion charge
ρ_e, ρ_i	C/m ³	Electron and ion space charge
R	-	R value
t_a, t_s	m	Accel and screen grid thickness
T_e, T_i	K or eV	Electron and ion temperature
v	m/s	Velocity
V_a, V_s	V	Accel and screen grid voltage
V_N, V_T	V	Net and total accelerating voltage
$V_{\text{SaddlePoint}}$	V	Centerline saddle point potential
Y	atoms/ion	Sputter yield

3 Ion Thrusters

Ion thrusters are a subclass of propulsion devices in the area of electric propulsion. Ion thrusters, and other electric propulsion devices in general, operate with better efficiency than chemical propulsion devices. Electric propulsion devices differ from chemical propulsion devices in that the energy used to accelerate the propellant does not come from the propellant itself, but from a separate power source.

3.1 Ion Thruster Operation

Ion thrusters are described as electrostatic thrusters because they use electric fields, rather than magnetic fields or pressure differences, to accelerate the propellant to produce thrust. The overall synopsis of what an ion thruster does is to ionize a propellant to form a plasma, and then accelerate ions from the plasma to high velocities using the ion optics in order to produce thrust.

A general schematic of an ion thruster is shown in Figure 3.1. In this figure, a propellant, typically xenon which is in a gaseous form, is fed into the ion thruster discharge chamber both directly (the main propellant flow) and through a device commonly used in ion thrusters called the hollow cathode. The hollow cathode is used to provide a stream of electrons into the discharge chamber. These electrons are considered monoenergetic, and are drawn into the discharge chamber as a result of the potential difference between the cathode and the anode, which is typically 25 to 30 Volts. These electrons have enough energy to strike the neutral xenon gas atoms inside the chamber and knock an electron away from the atom, which forms an ion and two electrons. Once this process is started, the collection of neutral gas atoms, electrons, and ions is called the discharge chamber plasma.

Magnet rings placed at various locations around the discharge chamber are used to provide magnetic fields that help to confine the electrons within the chamber. This is done to increase the collision frequency of the atoms and electrons in order to produce ions with less input power.

The ion thruster grids, which are also called the ion optics, are attached to one side of the ion thruster and are used to accelerate ions from the discharge chamber to produce thrust. Two grids are shown in Figure 3.1, but three or more grids can be used depending on the application of the ion source. The grid closest to the discharge chamber plasma is called the screen grid, and the next grid is called the accelerator grid, commonly referred to as the accel grid for short.

Thousands of holes are cut into each grid so that ions can pass through them. Normally, the discharge chamber plasma is at a potential close to, or a few volts above, the potential of the anode. The screen grid is normally electrically connected to the potential of the cathode. Therefore, the screen grid is at a potential that is 25 to 30 volts lower than the discharge chamber plasma. Because of this setup, ions from the discharge chamber are slightly attracted toward the screen grid, and electrons are repelled away from the screen grid.

Outside the spacecraft downstream of the ion thruster there is a plasma called the beam plasma that is close to or a few volts positive of the potential of the spacecraft. The accel grid is held at a potential that is negative relative to the entire spacecraft. The accel grid is biased negative to both attract ions out of the discharge chamber and to repel electrons away from the thruster that are in the downstream beam plasma.

The final thing to notice from Figure 3.1 is a device called a neutralizer. The neutralizer is used to expel electrons from the spacecraft at exactly the same rate that ions are expelled. This is done to maintain a steady spacecraft charge relative to the beam plasma. The neutralizer can be a second hollow cathode that expels electrons into the beam plasma.

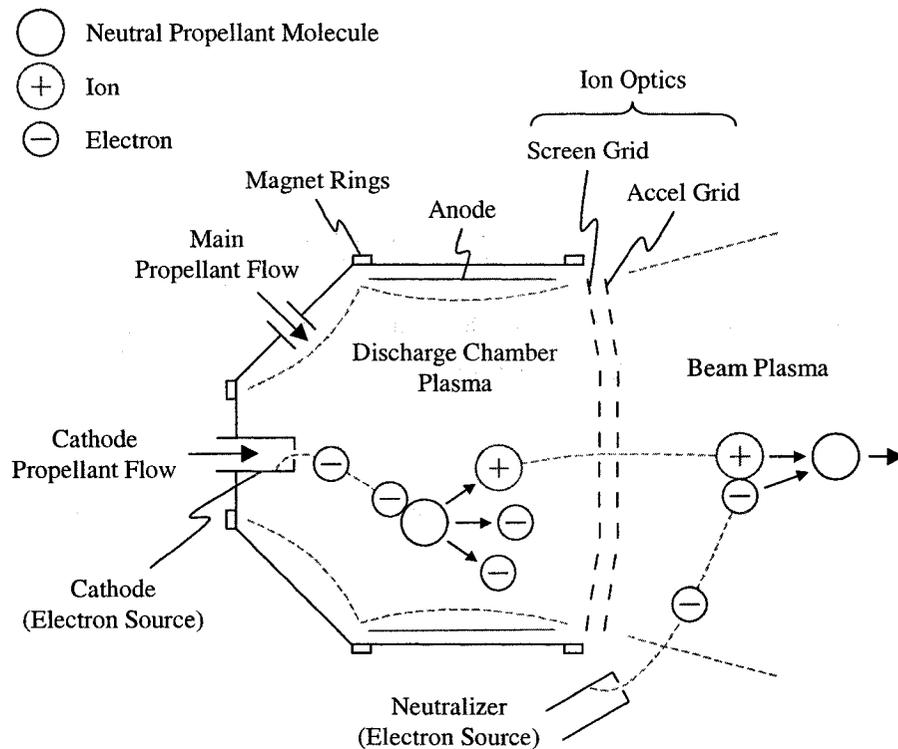


Figure 3.1 Schematic of an ion thruster.

3.2 Ion Optics

The purpose of the ion optics, or grids, is to accelerate ions from the discharge chamber to produce thrust. An example grid set has been created in order to investigate many different aspects of ion optics. This grid set is first used to provide an introduction to ion optics. Later, the creation of the example grid set will be explained in greater detail.

3.2.1 Discharge Chamber and Beam Plasmas

Within the discharge chamber, a plasma is created independently of the ion optics system. A plasma, simply put, is a collection of ions, electrons, and neutral particles. The discharge chamber plasma is characterized in part by a plasma potential, ϕ_u , an electron temperature, T_e , an ion temperature, T_i , and a neutral gas temperature, T_n . Downstream of the ion optics, a beam plasma exists which has a different potential, ϕ_d , and temperatures, $T_{e,d}$, $T_{i,d}$, and $T_{n,d}$, than the upstream discharge chamber plasma.

In typical ion thruster plasmas, most of the ions are singly-charged, meaning the molecules or atoms have one less electron than they do protons, while a small fraction of the ions are doubly-charged, meaning that the atoms have two fewer electrons than protons. The ratio of doubly-to-singly charged ions might on average be 10 % or so, but this ratio can vary greatly inside the discharge chamber. The plasma could also have smaller fractions of even higher multiply-charged ions, but these fractions are normally so small that they are neglected. Also, the discharge chamber plasma might typically be 10 % ionized, meaning 10 % of the propellant atoms are ionized while the rest of the atoms exist as un-ionized neutral atoms.

Within the plasma, there is approximate charge neutrality, meaning the charge density of ions, ρ_i , is equal to the charge density of electrons, ρ_e , as shown in Eq. 3.1. The charge density of an individual species is its number density multiplied by its charge state. Therefore, the charge neutrality approximation can also be written as Eq. 3.2 assuming only singly-charged and doubly-charged ions are present along with electrons. The discharge chamber and beam plasmas can be characterized by bulk (average) ion and electron densities.

$$\rho_i = \rho_e \quad \text{Eq. 3.1}$$

$$en_i^+ + 2en_i^{++} = en_e \quad \text{Eq. 3.2}$$

3.2.2 Grids

For most of the ion thrusters discussed here, two grids are used in the ion optics. The grid furthest upstream, which sits next to the discharge chamber plasma, is called the screen grid. The next grid downstream of the screen grid is called the accelerator grid, or simply the accel grid for short. Sometimes, a third grid is also used in the optics which is called the decelerator grid, or decel grid for short. Figure 3.2 shows an example a) two-grid system and b) three-grid system.

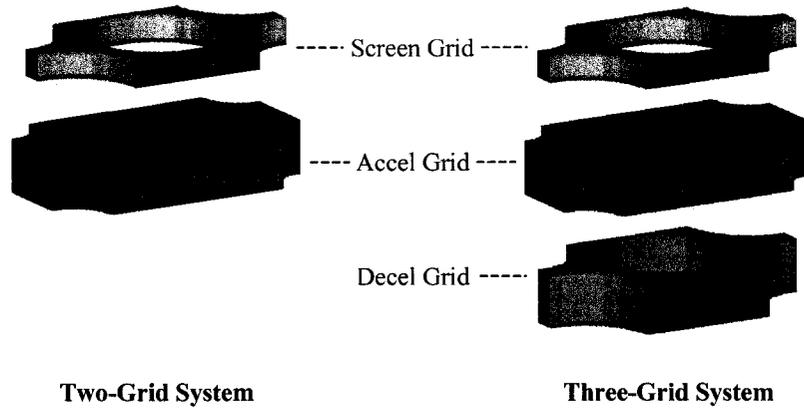


Figure 3.2 Example two-grid and three-grid systems.

A schematic of the two-grid system is shown in Figure 3.3 that shows some of the grid geometry labeling conventions. Aligned holes are made in the screen and accel grids to allow ions to pass through them. The screen grid hole diameter is given the symbol d_s , and the screen grid thickness t_s . Similarly, the accel grid hole diameter is d_a and the accel grid thickness is t_a . Finally, the spacing between the grids is l_g .

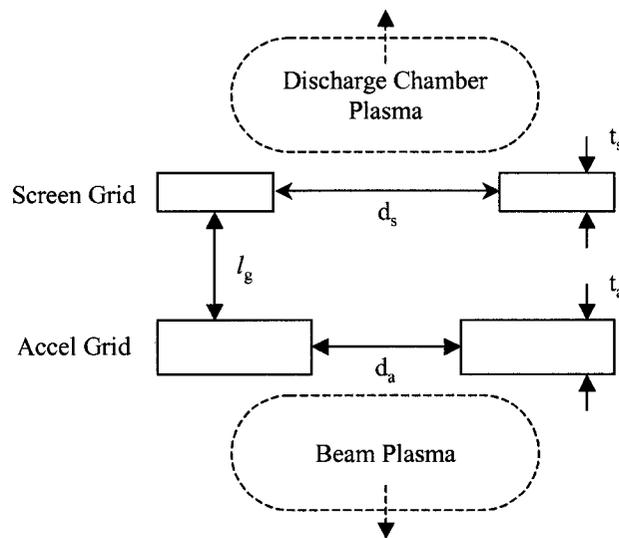


Figure 3.3 Schematic of a two-grid system showing definitions.

The holes made in the grids can in theory be any shape and can be arranged in any pattern over the grid face. Traditionally, circular holes are used in what is referred to as a hexagonal aperture layout. Figure 3.4 shows a small section of a full grid set viewed from the upstream side. The screen grid holes are typically larger than the accel grid holes, thus the inner parts of the accel grid holes, in blue, can be seen through the screen grid holes.

The ffx code analyzes a three dimensional, rectangular, volume. For the standard hexagonal aperture layout, the ffx code rectangular analysis volume, shown outlined in black in Figure 3.4, encompasses two quarter-sized sections of two separate apertures. Symmetry conditions are applied on all four sides of the analysis volume. Hexagons, shown outlined in green, can be drawn around each aperture.

These hexagons represent the smallest repeating unit on the grid face, hence the aptly named hexagonal aperture layout.

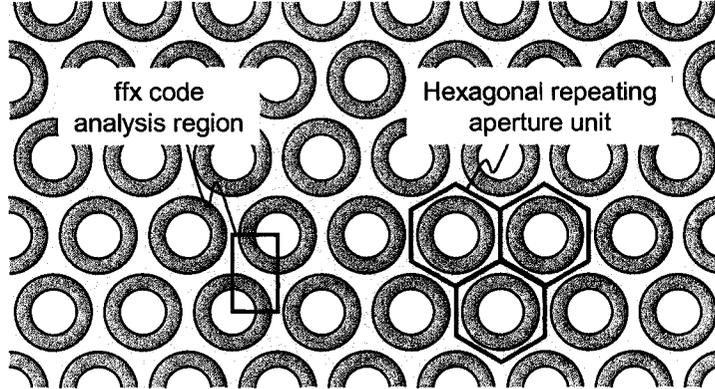


Figure 3.4 Typical ion thruster circular hole shape and hexagonal aperture layout pattern.

The spacing between the apertures in the grids is called the hole center-to-center distance, l_{cc} . The hole center-to-center distance is usually only slightly larger than the screen grid hole diameter, d_s , as shown in Figure 3.5.

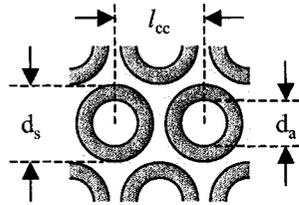


Figure 3.5 Definition of the hole center-to-center distance, l_{cc} .

3.2.3 Voltages

The discharge chamber plasma potential, ϕ_u , is typically close to, or a few volts above, the potential of the anode within the discharge chamber. Additionally, the screen grid is normally electrically connected to the cathode, thus the cathode and the screen grid are at the same applied potential, V_s . The difference between the anode and cathode potential is called the discharge voltage, V_d , given in Eq. 3.3. Common discharge voltages range from 25 to 30 V.

$$\begin{aligned}
 \phi_u &\approx V_{anode} \\
 V_s &= V_{cathode} \\
 V_d &= V_{anode} - V_{cathode} \\
 &\approx \phi_u - V_s
 \end{aligned}
 \tag{Eq. 3.3}$$

The net accelerating voltage, V_N , is the difference between the discharge chamber and beam plasma potentials, as given in Eq. 3.4. It is the net voltage that the ions will “fall” through when being

accelerated in going from the discharge plasma to the beam plasma. Generally, net voltages might range from 500 to 10,000 V in ion thrusters.

$$V_N = \phi_u - \phi_d \quad \text{Eq. 3.4}$$

The beam plasma potential, ϕ_d , is normally close to zero volts. As a result, the net accelerating voltage is commonly taken as being equal to the potential of the discharge chamber plasma. With this assumption, the net accelerating voltage is also the same as the screen grid voltage plus the discharge voltage, as shown in Eq. 3.5.

$$\begin{aligned} \phi_d &\approx 0 \text{ V} \\ V_N &\approx \phi_u \\ V_N &= V_s + V_d \end{aligned} \quad \text{Eq. 3.5}$$

The beam plasma potential has been seen to be as high as 20 V in ion thruster measurements. For a discharge chamber plasma potential of 1000 V and a beam plasma potential of 10 V, the error in the net accelerating voltage is only 1 % where the actual net accelerating voltage is 990 V and the approximate net accelerating voltage is taken to be 1000 V.

The applied accel grid voltage is almost always negative. More specifically, it is negative relative to the beam plasma potential. This is done to prevent electron backstreaming, which is discussed in greater detail later on. Essentially, the accel grid voltage is held negative to prevent electrons within the beam plasma from traveling upstream into the discharge chamber.

The total accelerating voltage, V_T , is defined as the difference of the net accelerating voltage, V_N , and the applied accel grid voltage, V_a , as in Eq. 3.6. The total accelerating voltage is approximately equal to the maximum potential change that an ion will experience in passing through the grids. In other words, each ion begins its acceleration through the grids at a potential close to the upstream plasma potential, which is also equal to the net accelerating voltage, V_N . Near the accel grid hole, the ion reaches its greatest velocity where the potential is lowest, where this potential is normally less than the beam plasma potential and greater than the accel grid voltage.

$$V_T = V_N - V_a = V_N + |V_a| \quad \text{Eq. 3.6}$$

The ratio of the net accelerating voltage to the total accelerating voltage is called the R value, shown in Eq. 3.7. In general, one typically wants to operate with a high R value, near unity, where the accel grid voltage magnitude is low relative to the net accelerating voltage. Typical ion thruster R values range from 0.85 to 0.90.

$$R = \frac{V_N}{V_T} = \frac{V_N}{V_N + |V_a|} \quad \text{Eq. 3.7}$$

Unless otherwise specified, the grid geometry and operating conditions for all examples in this chapter will be those given in Figure 3.6.

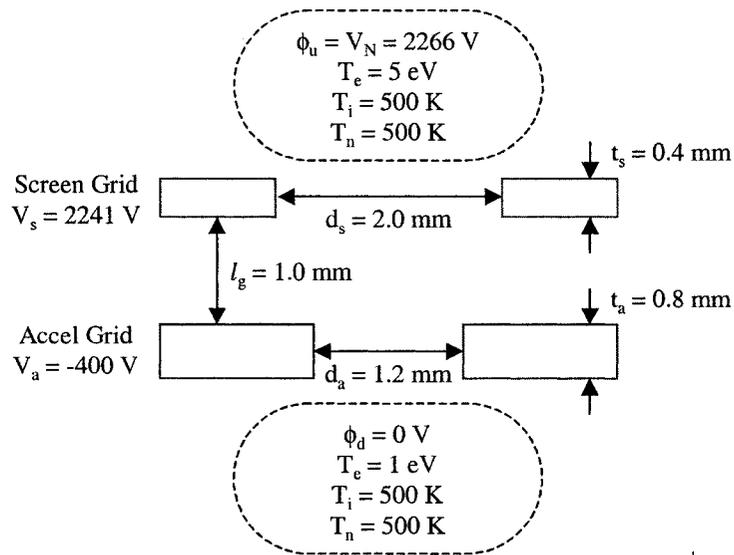


Figure 3.6 Two-grid system with example voltages.

3.2.4 Currents

After the discharge chamber plasma is created and the screen and accel grid potentials are applied, a beam of ions begins passing through each aperture pair. The beam of ions passing through a single aperture pair is referred to as an ion beamlet, whereas the beam of ions coming from an entire ion source is referred to collectively as the ion beam. An example ion beamlet is shown in Figure 3.7a. In this figure, the shape of the ion beamlet can be seen from the plot of the ion density, in ions/m³, within the simulation region.

There are several useful currents that can be measured. The rate of ions arriving at an aperture is dependent on the discharge chamber plasma density. The ions that arrive at an aperture can either a) strike the upstream surface of the screen grid, b) pass through the screen grid hole and strike the accel grid, or c) pass through both the screen and accel grid holes and go into the beam plasma.

The current of ions that passes through both the screen and accel grid holes is called the beamlet current, J_b . The current of ions that strike the screen grid is the screen grid current, J_s , and the current of ions that strike the accel grid is the accel grid current, J_a . Figure 3.7b shows a picture of an ion beamlet with these three currents labeled. In this case, J_a is zero as there are no ions impinging upon the accel grid.

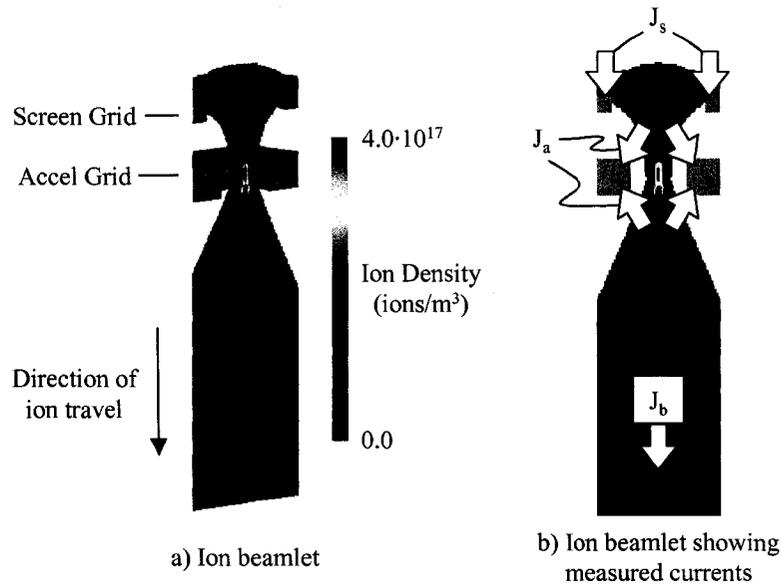


Figure 3.7 An example of an ion beamlet. $J_b = 0.139$ mA.

3.2.5 Sheath and Neutralization Surfaces

Because the screen and accel grids are held at potentials different than both the discharge chamber and beam plasmas, two plasma sheaths form that isolate the plasmas from the grids.

One of these sheaths forms upstream of the screen grid, and is simply called the sheath. Upstream of this sheath, all ions within the plasma are at a potential very close to the assigned bulk plasma potential, ϕ_u . Within the discharge chamber plasma, ions and electrons have equal charge densities, meaning that there are roughly equal numbers of ions and electrons. Electric fields upstream of the sheath can be considered to be small relative to the electric fields that exist downstream of the sheath closer to the grids.

Downstream of the sheath, strong electric fields exist due to the potential differences between the discharge chamber plasma and the screen and accel grids. Because the screen grid is negative relative to the discharge chamber plasma, ions that arrive at the sheath are attracted toward the screen grid while electrons are repelled back into the discharge chamber plasma. The electron density drops off rapidly downstream of the sheath.

Downstream of the grids, a second sheath exists which is called the neutralization surface. This sheath works in a similar way to the sheath upstream of the grids. Downstream of the neutralization surface, ions and electrons are considered to be of equal charge density and at the same potential, ϕ_d . Upstream of the neutralization surface, the potential becomes lower when moving closer to the accel grid. As a result, electrons are repelled back downstream into the beam plasma while ions are attracted toward the accel grid.

The region downstream of the sheath and upstream of the neutralization surface is called the intra-grid region. The entire ion acceleration process takes place within this region where the electric fields are strong. Figure 3.8 shows the sheath and neutralization surfaces for an example ion beamlet.

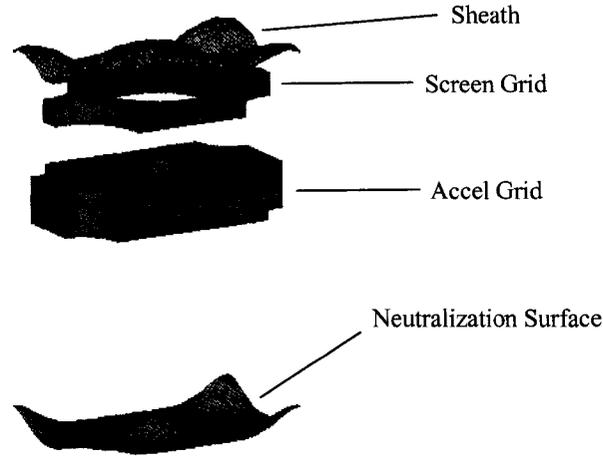


Figure 3.8 Sheath and neutralization surfaces for an example ion beamlet. $J_b = 0.139$ mA.

Weak electric fields can exist upstream of the sheath, and downstream of the neutralization surface. There is a region immediately upstream of the sheath called the pre-sheath region. Within an idealized plasma, ion and electron velocities are distributed equally in all directions. In the pre-sheath region, weak electric fields tend to make the ion velocities more or less perpendicular to the sheath. The number of ions per unit area arriving at any location on the sheath surface is approximately the same.

Ions moving through the pre-sheath region acquire a certain velocity called the Bohm velocity [Brown]. Eq. 3.8 gives the Bohm velocity where f_{Bohm} is a non-dimensional adjustment factor sometimes taken to be unity, q is the electric charge state of the ion, T_{e_eV} is the electron temperature of the bulk discharge chamber plasma in electron volts, and m_i is the mass of the propellant atom.

$$v_{Bohm} = f_{Bohm} \sqrt{\frac{qT_{e_eV}}{m_i}} \quad \text{Eq. 3.8}$$

The electron temperature of the bulk discharge chamber plasma can be expressed in Kelvin instead of eV by using the relation given in Eq. 3.9, where e is the elementary charge value and k is Boltzmann constant which relates temperature to energy.

$$T_{e_K} = T_{e_eV} \frac{e}{k} \quad \text{Eq. 3.9}$$

An important thing to remember is that the total energy of an ion is conserved as it moves. The ion's total energy is composed of potential and kinetic energy, as given in Eq. 3.10.

$$\begin{aligned} E_{Total} &= E_{Potential} + E_{Kinetic} \\ &= q_i\phi + \frac{1}{2}m_iv_i^2 \end{aligned} \quad \text{Eq. 3.10}$$

Within a plasma, the mean speed of the ions is given by Eq. 3.11. For xenon ions at an ion temperature of 500 K, the mean ion speed is 284 m/s, which corresponds to a kinetic atom energy of 0.055 eV. For a discharge chamber bulk electron temperature of 5 eV, the Bohm velocity for a singly charged xenon atom is 1917 m/s, which corresponds to a kinetic atom energy of 2.5 eV, which is about 46 times larger than the kinetic energy within the bulk plasma.

$$v_{i \text{ Mean}} = \sqrt{\frac{8kT_i}{\pi m_i}} \quad \text{Eq. 3.11}$$

The potential at the sheath can be calculated by using the conservation of ion energy equation applied between the discharge chamber plasma and the sheath, as shown in Eq. 3.12. Because the kinetic energy of the ion at the sheath is usually much greater than its energy within the plasma, the speed of the ion within the plasma can be neglected. Inserting the expression for the Bohm velocity into this equation leads to Eq. 3.13.

$$q_i \phi_u + \frac{1}{2} m_i v_{i \text{ Mean}}^2 = q_i \phi_{\text{Sheath}} + \frac{1}{2} m_i v_{\text{Bohm}}^2 \quad \text{Eq. 3.12}$$

$$\phi_u = \phi_{\text{Sheath}} + \frac{T_e}{2} \quad \text{Eq. 3.13}$$

The current density of ions arriving at the sheath, j_{Sheath} , can be calculated using Eq. 3.14, where $n_{i \text{ Sheath}}$ is the number density of ions at the sheath, v_{Bohm} is the Bohm velocity (the speed of the ions at the sheath), and e is the elementary charge. The barometric equation, Eq. 3.15, relates the electron density at any potential to the electron density in the bulk plasma farther upstream [Boerner]. The number density of ions at the sheath, $n_{i \text{ Sheath}}$, is calculated using Eq. 3.16, which is derived from the barometric equation, where $n_{i \text{ Bulk}}$ is the ion number density within the discharge chamber plasma. The ion and electron densities are considered equal within the bulk plasma and at the sheath, for instance $n_{i \text{ Sheath}} = n_{e \text{ Sheath}}$ and $n_{i \text{ Bulk}} = n_{e \text{ Bulk}}$.

$$j_{\text{Sheath}} = n_{i \text{ Sheath}} v_{\text{Bohm}} e \quad \text{Eq. 3.14}$$

$$n_e = n_{e \text{ Bulk}} \exp\left(\frac{\phi - \phi_u}{T_{e u}}\right) \quad \text{Eq. 3.15}$$

$$n_{e \text{ Sheath}} = n_{e \text{ Bulk}} \exp\left(\frac{\phi_{\text{Sheath}} - \phi_u}{T_{e u}}\right) \quad \text{Eq. 3.16}$$

$$n_{i \text{ Sheath}} = n_{i \text{ Bulk}} \exp\left(-\frac{1}{2}\right) \approx 0.607 n_{i \text{ Bulk}}$$

For a given plasma temperature and density, there is a certain distance over which charged particles within a plasma can interact. For a certain charged particle, any other charged particles within that

distance feel the effect of the particle, while all the particles outside of that distance do not know the particle is there. This distance is called the Debye length, λ_D , named after Peter Debye.

In ion thruster plasmas, the electrons are usually at a much greater temperature than the ions, and thus they have much greater mobility. For such plasmas, the Debye length is given in Eq. 3.17, where ϵ_0 is the permittivity of free space, k is Boltzmann's constant, T_e is the electron temperature within the plasma, e is the charge of an electron, and n_e is the number density of electrons within the plasma.

$$\lambda_D = \frac{\epsilon_0 k T_e}{e^2 n_e} \quad \text{Eq. 3.17}$$

The Debye length is the distance that electrons can shield out external electric fields. Thus, the distance between a plasma sheath and a surface held at a different potential than the plasma is roughly equal to the Debye length. In the case of the discharge chamber plasma and the screen grid for example, the Debye length is the distance over which electrons in the plasma can feel the effect of the screen grid. Electrons that go closer to the screen grid than the Debye length feel the effect of the screen grid and are repelled back into the plasma. At distances greater than the Debye length away from the screen grid, ions and electrons within the plasma do not feel the effect of the screen grid.

3.2.6 Impingement Limits

The ideal way to operate an aperture is to choose the upstream discharge chamber plasma density and grid voltages such that there is no direct impingement of ions onto the accel grid. Ions that originate from within the discharge chamber may have thousands of electron volts of energy when they reach the accel grid. If these ions strike the accel grid directly, erosion of the accel grid can take place rapidly.

It is often desirable to determine an appropriate beamlet current range for a given set of grid voltages. Increasing the discharge chamber plasma density naturally forces more current to be pushed toward the grids. With more ions passing through an aperture, the positive ion charge within the beamlet increases. This causes the ion beamlets to expand in diameter as a result of the ions repelling one another. Increasing the upstream ion density further eventually reaches the point where ions directly strike the accel grid. This is called the perveance limit, and it is often specified in terms of a beamlet current, J_b perveance. At the perveance limit, ions that enter the edge of the screen grid hole are accelerated along the outer edge of the beamlet where they impinge upon the accel grid.

On the other hand, decreasing the discharge chamber plasma density too far can also cause direct accel grid impingement. In this case, the electric fields at low beamlet currents can set up such that ions entering the edge of the screen grid hole can crossover the beamlet centerline and strike the accel grid on the opposite side of the hole from which they started. This is called the crossover limit, and it is also often specified as a beamlet current, J_b crossover.

It is best to operate all individual aperture pairs on the grid surface between the crossover and perveance limits where there is no direct impingement of high-energy ions onto the accel grid. Figure 3.9

shows five beamlets operating at beamlet currents from 0.035 to 0.558 mA. At the 0.035 mA beamlet current, ions entering the edge of the screen grid hole crossover the beamlet centerline and impinge upon the accel grid. At the 0.558 mA beamlet current, ions entering the edge of the screen grid hole are accelerated nearly axially downstream where they impinge upon the accel grid.

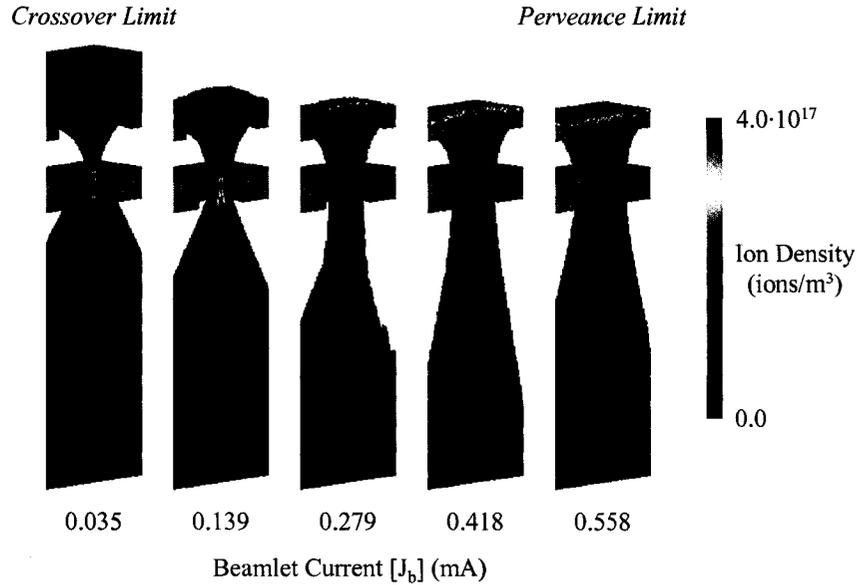


Figure 3.9 Beamlets operating at, and between, the crossover and perveance limits.

Figure 3.10 shows the trajectories of ions as they are released from the sheath as a function of beamlet current. Here, the trajectories of ions that crossover and strike the accel grid at the lowest beamlet current can be seen. Also, at the highest beamlet current, one can see the trajectories of the edge ions as they follow the beamlet surface axially and strike the accel grid.

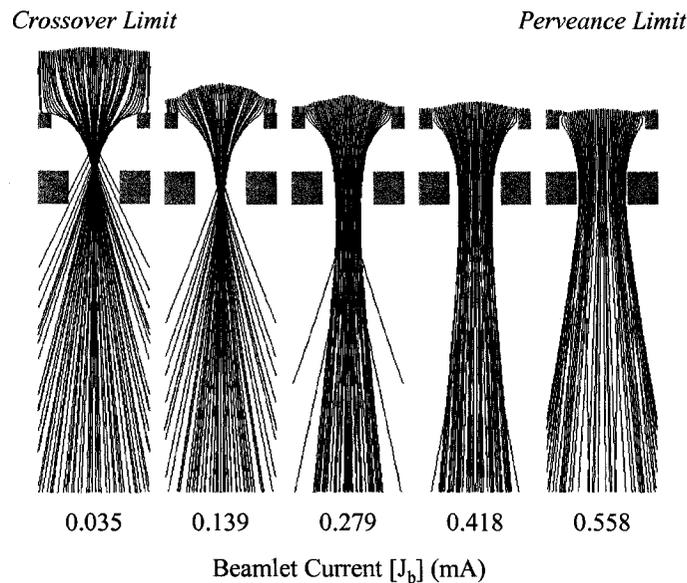


Figure 3.10 Trajectories of ions released from the sheath.

It is also interesting to view the sheath and neutralization surfaces as a function of beamlet current, as shown in Figure 3.11. At low beamlet currents, the sheath and neutralization surfaces are far away from the screen and accel grids. As the beamlet current increases, and the plasma density increases both upstream and downstream of the grids, the sheath and neutralization surfaces move closer to the grids. This is because the Debye length decreases as the electron density increases. As a result, the electrons shield the potential differences between the plasmas and the grids over shorter distances.

Another way to think about why the sheath and neutralization surfaces move closer to the grids as beamlet current increases is to think about the individual ion and electron interactions. An ion, for example, within a plasma is unlikely to interact with a grid unless the electric field between the ion and the grid is comparable to the electric field between the ion and the surrounding charged particles within the plasma. With more charged particles at greater plasma densities, the ions and electrons in the plasma have to move closer to the grid in order for the effect of the grid to be significant.

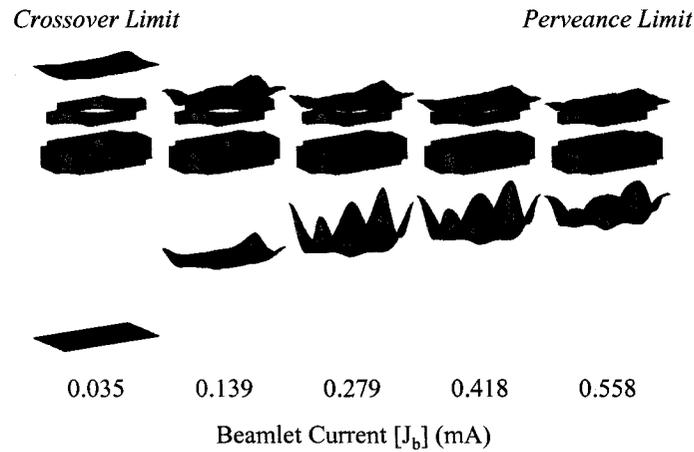


Figure 3.11 Sheath and neutralization surfaces.

The amount of impingement current to the accel grid, J_a , can be plotted as a function of the beamlet current, J_b . A slightly more useful thing to do is to plot the ratio of the impingement current to the beamlet current, J_a/J_b , versus the beamlet current. This is called an impingement limit curve because it allows the crossover and perveance limits to be found easily.

The impingement limit curve corresponding to the ion beamlets shown in Figure 3.9 is given in Figure 3.12. Experimentally, the crossover and perveance limits are sometimes defined as the points where the impingement ratio rises above some chosen value. In this case, if the limits are chosen as the points where the impingement ratio rises above 1 % for example, the crossover limit is about 0.054 mA and the perveance limit is about 0.508 mA.

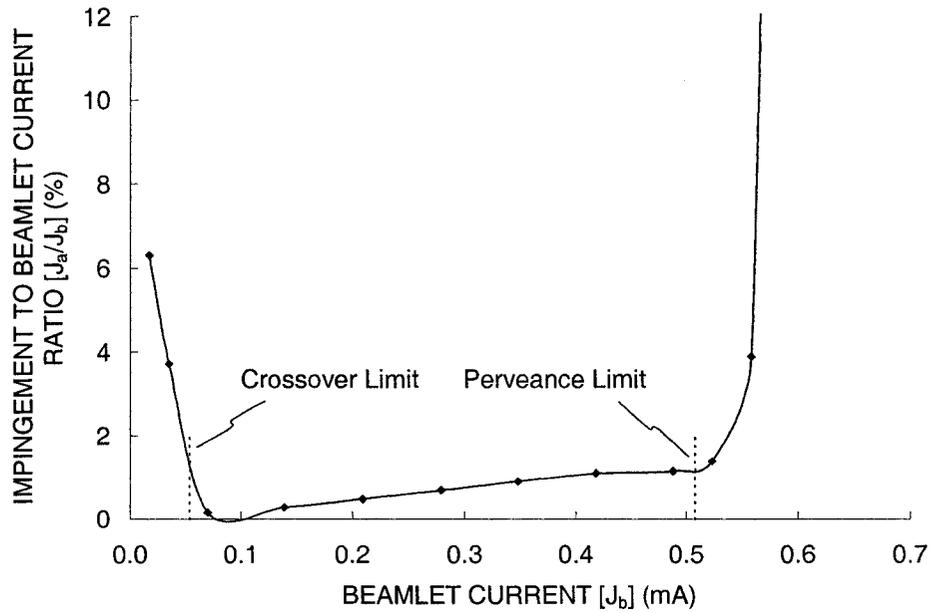


Figure 3.12 Impingement limit curve showing the crossover and perveance limits.

3.2.7 Grid Transparency To Ions

Ideally, all ions that arrive at an aperture from the discharge chamber would be accelerated through the grids to produce thrust. In reality, a small fraction of the ions that arrive at an aperture are accelerated into the screen grid, which is the current J_s . Recall that the screen grid is at a potential less than the discharge chamber plasma, thus it attracts ions. An ion that strikes the screen grid picks up an electron and returns to the discharge chamber as a neutral molecule.

The grid transparency to ions, ϕ , is defined in Eq. 3.18. It can be thought of as the current of ions that make it through the grids (J_b) divided by the total current of ions that arrive at the grids ($J_s + J_a + J_b$). It is also commonly called the screen grid transparency to ions because in normal operation, the accel grid current, J_a , due to ions from the discharge chamber is zero. In places, J_a may not be included in the definition. Additionally, the grid transparency to ions may or may not be expressed as a percentage through the factor of 100.

$$\phi = \frac{J_b}{J_s + J_a + J_b} \cdot 100 \quad \text{Eq. 3.18}$$

The actual grid transparency to ions calculated with Eq. 3.18 is often very close to the physical open area fraction of the screen grid, ϕ_s . The physical screen grid transparency for the hexagonal aperture layout is equal to the circular hole area divided by the area of the (imagined) hexagon that surrounds each aperture. This calculation is given in Eq. 3.19. Common screen grid open area fractions range from 65 to 70 %.

$$\phi_s = \frac{\text{OpenArea}}{\text{TotalArea}} = \frac{\frac{\pi}{4} d_s^2}{\frac{\sqrt{3}}{2} l_{cc}^2} = \frac{\pi}{2\sqrt{3}} \frac{d_s^2}{l_{cc}^2} \quad \text{Eq. 3.19}$$

Figure 3.13 shows the actual grid transparency to ions, as calculated using Eq. 3.18, as a function of beamlet current. As is typical, the transparency to ions is greater near the crossover limit than it is near the perveance limit. This is mainly a result of the shape of the sheath changing at different beamlet currents. As was seen in Figure 3.11, the sheath is located farther away from the screen grid at lower beamlet currents than it is at greater beamlet currents. At the lower beamlet currents, the electric fields are such that more ions that enter the sheath have trajectories that take them through the grids rather than striking the screen grid directly.

This effect is also easily seen in Figure 3.10. At the lowest beamlet current, many of the ions that start out directly upstream of the screen grid turn into the beamlet before they strike the screen grid. When the sheath is located upstream away from the screen grid, the pull of the accel grid on the ions is greater, causing more of the edge ions to be attracted toward the accel grid rather than the screen grid. At the greatest beamlet current however, the ions that start out near the screen grid see only the effect of the screen grid and have no chance to enter the beamlet.

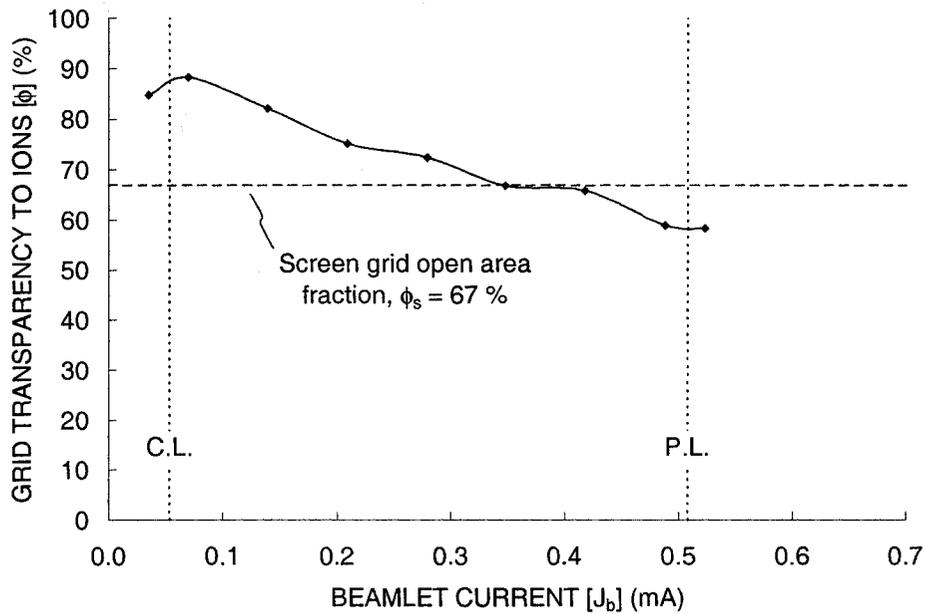


Figure 3.13 Grid transparency to ions, ϕ .

The actual grid transparency to ions in this case, which is typical, is generally slightly greater than or equal to the physical open area fraction of the screen grid. Below the crossover limit and above the perveance limit, both the accel grid current, J_a , and screen grid current, J_s , increase relative to the beamlet current, which causes the transparency, ϕ , to decrease.

3.2.8 Perveance Fraction

One measure of a beamlet is called the perveance fraction, f_p . It is a measure of the actual beamlet current being extracted to the maximum amount of beamlet current that could be extracted if conditions were ideal. The idea of the perveance fraction comes from the Child-Langmuir law, which gives the maximum current density of ions that can pass between two surfaces of different potential. Perveance fraction is defined in Eq. 3.20, where j is the current density of ions (which can be calculated several ways), P_{Max} is the maximum perveance which is a constant for a given propellant and given in Eq. 3.21, l_e is the effective ion acceleration length defined in Eq. 3.22, and V_T is the total accelerating voltage.

$$f_p = j \frac{1}{P_{Max}} \frac{l_e^2}{V_T^{3/2}} \quad \{Dimensionless\} \quad \text{Eq. 3.20}$$

$$P_{Max} = \frac{4\epsilon_0}{9} \sqrt{\frac{2e}{m_i}} \left\{ \frac{A}{V^{3/2}} \right\} \quad \text{Eq. 3.21}$$

$$l_e = \sqrt{(l_g + t_s)^2 + \frac{d_s^2}{4}} \quad \{m\} \quad \text{Eq. 3.22}$$

The perveance law was originally developed to apply between two parallel, perfectly transparent, infinite surfaces separated by a certain distance. In this situation, the maximum theoretical amount of current density that can be obtained is found by using a perveance fraction of unity. In practice, a current density less than or equal to this amount can be obtained. For xenon, P_{Max} is equal to $4.77 \cdot 10^{-9} \text{ A/V}^{3/2}$.

When the perveance fraction equation is applied to the ion optics, some assumptions have to be made. The effective acceleration length is thought of as being the distance from the sheath, which is at a potential more or less equal to the net accelerating voltage, to the upstream centerline point of the accel grid, which is at potential less negative than the accel grid itself due to the effects of ion space charge and the proximity of the positive screen grid. The sheath is typically curved upstream with the sheath being close to the screen grid at the hole edge, and as a result the effective acceleration length is more or less constant at all points on the sheath as shown in Figure 3.14.

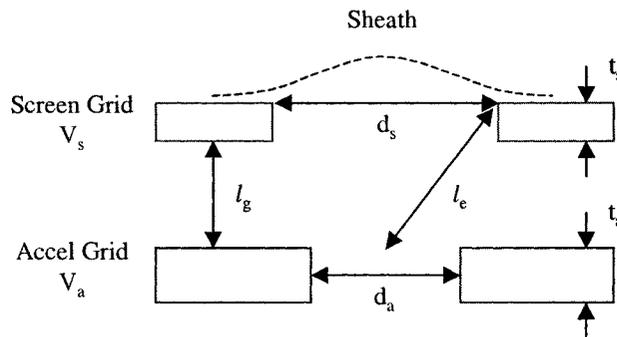


Figure 3.14 Definition of the effective acceleration length, l_e .

There are several ways to define the current density in the perveance equation. One way is to define current density as the beamlet current divided by the area of a single aperture, $A_{Aperture}$. Most commonly, a hexagonal aperture layout is used where the area of an aperture is defined as A_{Hex} in Eq. 3.23. Using this area, the maximum value of the perveance fraction, f_p , is unity. A different, more common definition is to use the circular area of the screen grid hole, as defined in Eq. 3.24. Using this definition essentially assumes that the ions that pass through an aperture all come from an area equal to the circular aperture area. However, if the sheath is curved upstream for example, more current can enter the sheath area and pass through the aperture than is assumed in using the circular hole area. In that case, the perveance fraction can actually turn out to be more than unity. Possibly the best definition would be to use the actual sheath area from which ions are extracted, given in Eq. 3.25.

$$A_{Hex} = \frac{\sqrt{3}}{2} l_{cc}^2 \quad \text{Eq. 3.23}$$

$$A_{Circle} = \frac{\pi}{4} d_s^2 \quad \text{Eq. 3.24}$$

$$A_{Sheath} = A_{Aperture} \phi \quad \text{Eq. 3.25}$$

Regardless of the current density definition, there is often a maximum beamlet current that can be reached before the onset of direct ion impingement onto the accel grid, which is called the perveance limit. This point is commonly reached at perveance fractions of 0.6 to 0.9. As a result, in normal thruster operation, all beamlet currents on the grid face must be less than or equal to this value. The peak beamlet current will likely occur on the centerline of the thruster where the ion production is greatest (typically when using a hollow cathode device to produce the primary electrons for instance).

The greatest utility of the perveance fraction equation is that it allows grid sets of different size to be compared directly through data reduction. Many results in this work are presented in terms of perveance fraction rather than current density or beamlet current, for example, as the results can be more readily compared to other grid sets of different size operating at different voltages and currents.

The impingement limit curve from Figure 3.12 can be put in terms of perveance fraction as well as beamlet current, the result of which is shown in Figure 3.15. Here, the current density has been defined as the beamlet current divided by the circular screen grid hole area, given in Eq. 3.26. Unless specifically stated, this will always be the way current density is defined when calculating the perveance fraction.

$$j = \frac{J_b}{\frac{\pi}{4} d_s^2} \quad \text{Eq. 3.26}$$

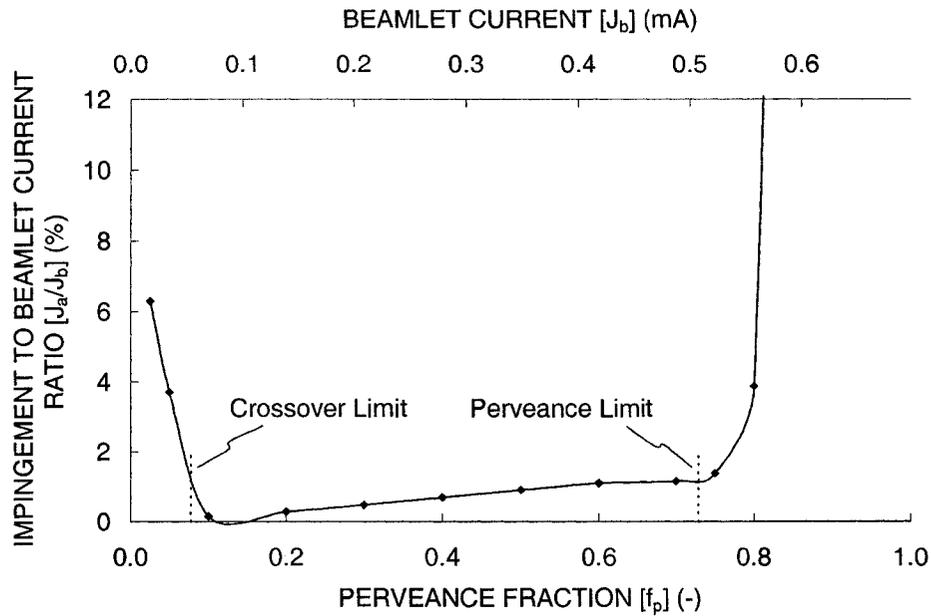


Figure 3.15 Impingement limit curve in terms of perveance fraction as well as beamlet current.

According to Figure 3.15, the crossover limit is at a beamlet current of 0.054 mA, which is equal to a perveance fraction of 0.078. Also, the perveance limit occurs at a beamlet current of 0.508 mA or a perveance fraction of 0.729.

When doubly charged ions are significant, an average ion charge, \bar{e} defined in Eq. 3.27, can be used in place of e in the perveance equation. As the double to single current ratio goes to infinity, the beamlet current at the perveance limit can be expected to increase by the square root of two.

$$\bar{e} = e \frac{1 + \frac{J_b^{++}}{J_b^+}}{1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+}} \quad \text{Eq. 3.27}$$

3.2.9 Propellant Flow

The discharge chamber plasma is a collection of ions, electrons, and neutral, un-ionized, propellant molecules. Ideally, all of the propellant that enters the discharge chamber would be ionized and accelerated through the ion optics to produce thrust. However, un-ionized neutrals can simply drift through the open holes in the grids. The ratio of the number of particles that are accelerated as ions to the number of particles that are lost as neutrals is called the propellant utilization efficiency, η_u , which is dimensionless.

The propellant utilization efficiency is approximately equal to the beamlet current divided by the propellant flow rate expressed in something called Amps equivalent, as shown in Eq. 3.28. Expressing the

flow rate as an equivalent amount of current is the idea that if every propellant molecule were ionized, $\dot{m}_{Aeq.}$ would be the current of ions coming out of the grids. But since every molecule is not ionized, only a fraction (η_u) of this maximum amount of current is actually extracted as ions.

$$\eta_u = \frac{J_b}{\dot{m}_{kg} \frac{e}{m_i}} = \frac{J_b}{\dot{m}_{Aeq.}} \quad \text{Eq. 3.28}$$

The above expression for the propellant utilization efficiency does not take into account the number of particles that are doubly-charged instead of only singly-charged. The expression in Eq. 3.29, which is developed later, takes this into account using the double to single current ratio, J_b^{++}/J_b^+ . The beamlet current is comprised of the current of ions that exits as singly-charged ions added to the current that exits as doubly-charged ions. The propellant utilization efficiency is really a number ratio, rather than a charge ratio, of the charged to uncharged particles exiting the grids.

$$J_b = J_b^+ + J_b^{++}$$

$$\eta_u = \frac{1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+}}{1 + \frac{J_b^{++}}{J_b^+}} \frac{J_b}{\dot{m}_{Aeq.}} \quad \text{Eq. 3.29}$$

Figure 3.16 shows an example cross sectional view of the neutral number density of xenon atoms through the grids for a propellant utilization efficiency of 90 %. This particular figure shows the neutral density variation calculated with an analytical model. In the analytical model, the neutral density at the exit of the accel grid is calculated such that it gives the desired propellant utilization efficiency. The neutral density increases going upstream through each grid, and is constant between the grids and upstream of the screen grid. The neutral density decline downstream of the grids has been set through comparison to experimental current measurements. This model is described in greater detail in a separate section.

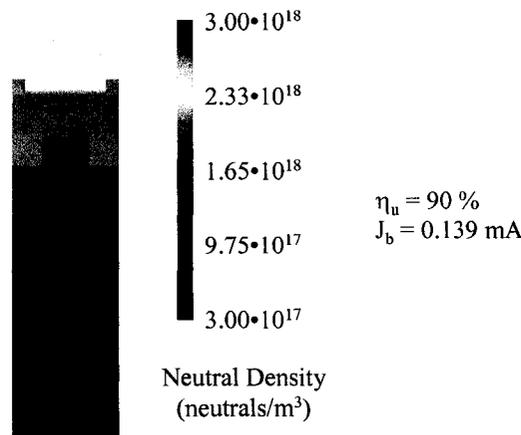


Figure 3.16 Neutral density variation through the grids.

Figure 3.17 shows a plot of the neutral density variation through the centerline of the beamlet. In this figure, the regions where the neutral density is constant, and where it is variable, can easily be seen.

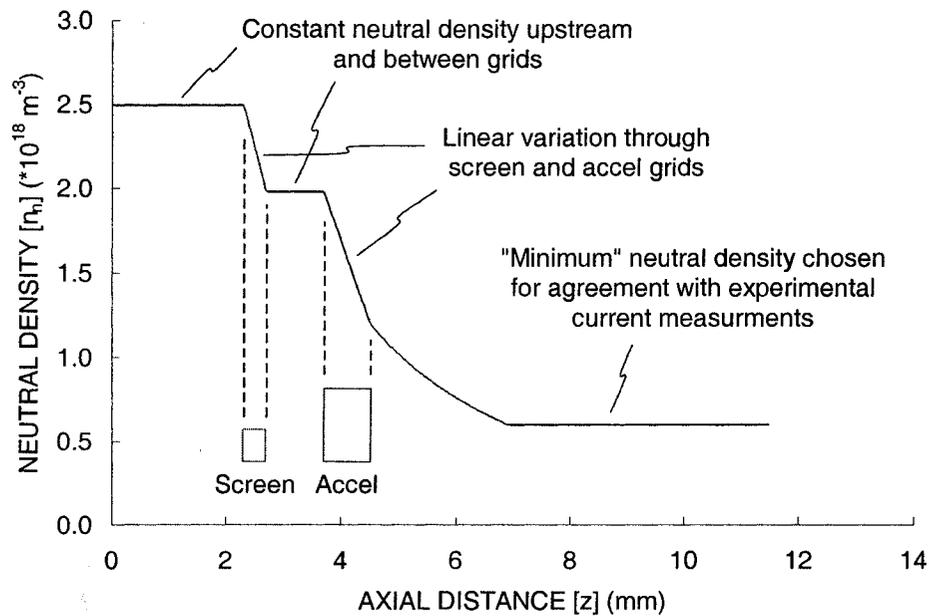


Figure 3.17 Axial neutral density variation.

3.2.10 Charge Exchange

A phenomenon called charge exchange, or charge transfer, takes place between the ions in the beamlet and the neutral atoms flowing through, and downstream, of the grids. The charge exchange process is illustrated in Figure 3.18a. When a fast moving ion in the beamlet passes very close to a relatively slower moving neutral, an electron from the neutral can transfer to the ion. This exchange of charge results in a fast moving neutral, which used to be the ion, and a slow moving ion, which used to be the neutral.

Charge exchange reactions are considered long distance interactions, meaning that no momentum is transferred between the ion and neutral particles. At the instant when charge exchange ions are created, they have the initial energy of the neutral atom which is small, often less than 1 eV. Because they are slow moving, they are accelerated essentially from rest by the local electric fields.

Figure 3.18b shows some example charge exchange ion trajectories. Many of the charge exchange ions that are created near the screen grid and far downstream of the grids, two of which are shown in blue in Figure 3.18b, are simply accelerated downstream away from the grids. These charge exchange ions are partially beneficial in that they do produce some thrust, although not as much as if no charge exchange reaction had occurred.

Charge exchange ions created in other parts of the beamlet can be accelerated into the screen and accel grids. These ions are undesirable because they cause erosion. The charge exchange ions that strike the

screen grid, an example of which is shown in light blue, usually have very little energy, less than 25 to 30 eV depending on what the plasma potential is relative to the screen grid potential, and do not cause much damage.

The charge exchange ion trajectories shown in red, orange, and yellow are cause for concern. These charge exchange ions can have gained several hundred electron volts of kinetic energy by the time they reach the accel grid from the point at which they were created. Charge exchange ions that are created downstream of the grids are generally accelerated toward the downstream side of the accel grid, shown in red. Charge exchange ions created near the accel grid hole can be accelerated into the inner part of the hole, referred to as the hole barrel, as shown by the orange trajectory. Finally, charge exchange ions created upstream near the screen grid at the outer surface of the beamlet can be accelerated into the upstream face of the accel grid, as shown by the yellow trajectory.

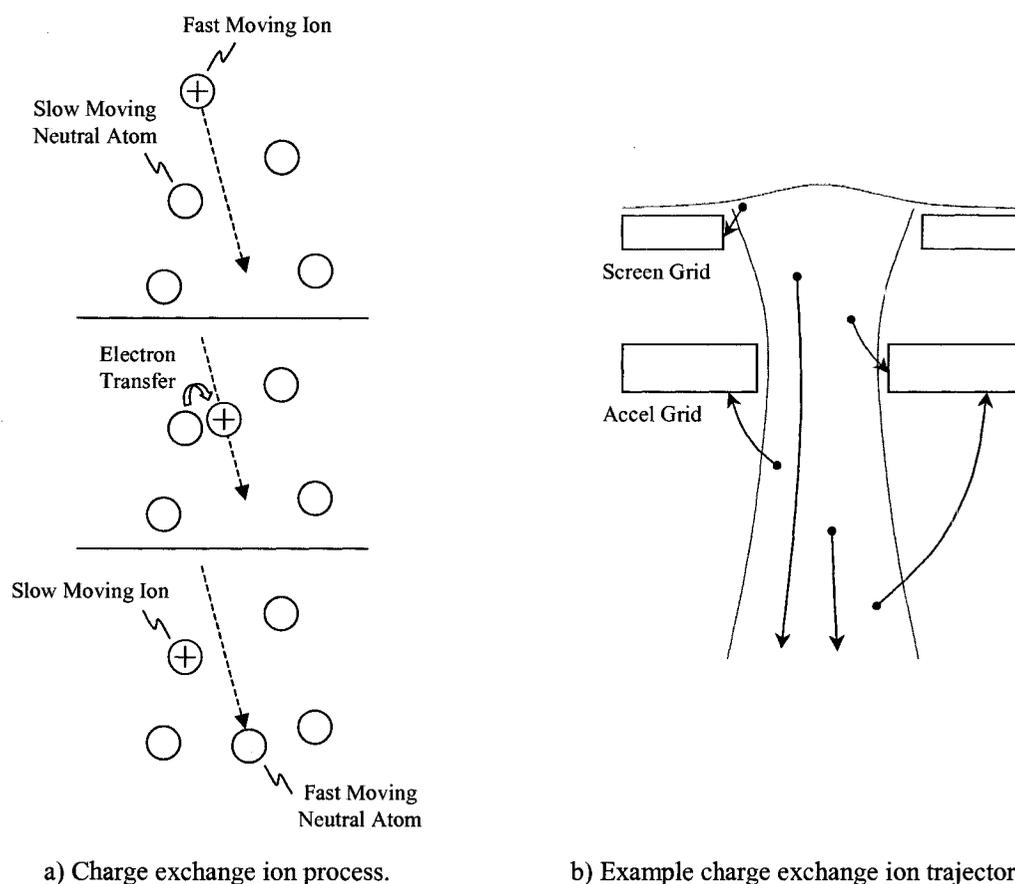


Figure 3.18 Charge exchange a) ion creation and b) example trajectories.

Figure 3.19 shows example ion beamlets with different regions of the beamlets colored according to the final termination surface where charge exchange ions end up going once they are created. The coloring scheme is the same as was used in Figure 3.18b. The charge exchange ions created in the blue regions of the beamlet end up traveling out the downstream surface of the simulation. The ions created in the light blue regions strike the screen grid. The charge exchange ions that are created in the yellow,

orange, and red regions end up striking the upstream, barrel, and downstream surfaces of the accel grid, respectively. In the green regions to the sides of the beamlet, no charge exchange ions are created because no beam ions are present to charge exchange. Charge exchange does take place in the green region upstream of the sheath, but these ions can be considered to be part of the group of ions coming from the discharge chamber as they have nearly identical trajectories.

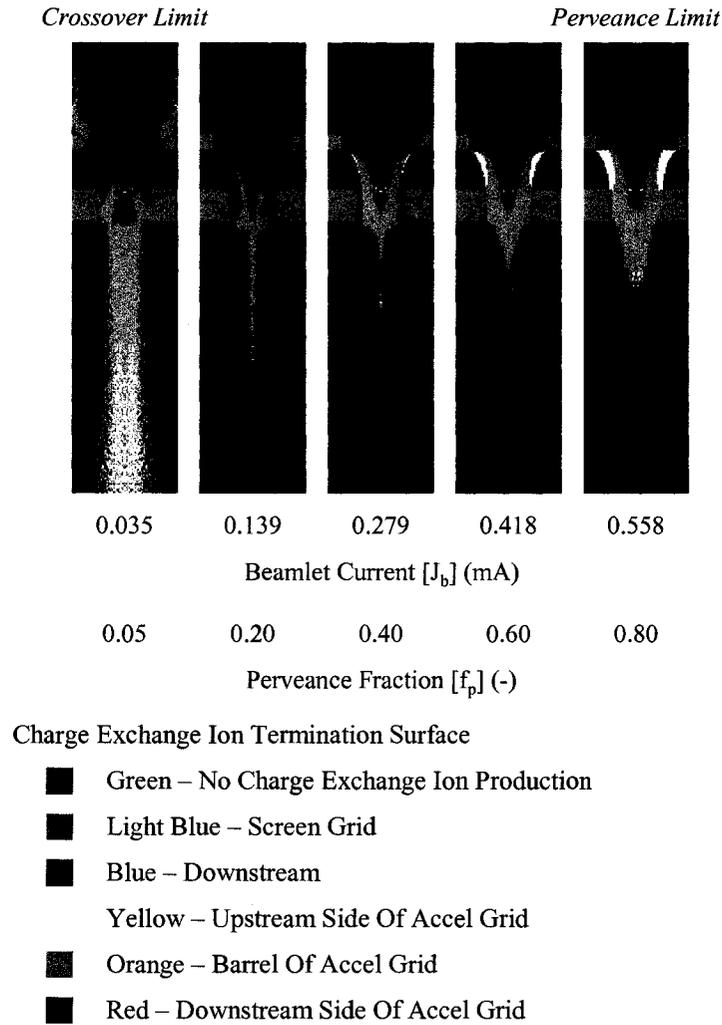
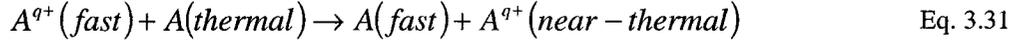


Figure 3.19 Ion beamlets colored according to the final termination surfaces of the charge exchange ions created in each region.

A simple equation is used to calculate the rate of charge exchange ion production, dn_{cx}/dt (in $m^{-3}s^{-1}$), at any point within the beamlet. The rate of charge exchange ion production at any point, given in Eq. 3.30, depends on the ion density, n_i (m^{-3}), the neutral density, n_n (m^{-3}), the speed of the ions as they pass by the neutrals, v_i (m/s), and the charge exchange cross section, $\sigma(v_i)$ (m^2), which is a function again of the ion speed relative to the neutral speed. The velocity used in this equation is the simply the ion velocity at the particular location because in almost all cases, the slow moving neutrals can be considered to be essentially stationary compared to the fast moving ions.

$$\frac{dn_{cx}}{dt} = n_i n_n v_i \sigma(v_i) \quad \text{Eq. 3.30}$$

The charge exchange reaction cross section, $\sigma(v_i)$, is different for each propellant type. There are also different types of charge exchange reactions. Symmetric charge exchange cross sections are often described by a chemistry type reaction equation given in Eq. 3.31. In this equation, A is the propellant species, Xe for instance, and q is a positive integer describing the charge state of the ion. In this case, only singly-charged ($q = 1$) and doubly-charged ($q = 2$) cross sections are relevant. The “symmetric” qualifier implies that the ion after the reaction has the same charge as the opposite ion before the reaction, or in other words q is the same for the ion on both sides of the reaction.



There can also be asymmetric charge exchange processes that might have to be accounted for. An example of a single electron asymmetric charge exchange reaction for xenon is given in Eq. 3.32. Here, it can be seen that two singly-charged ions are created from the reaction between a doubly-charged ion and a neutral xenon atom.

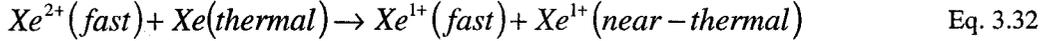


Figure 3.20 shows three charge exchange ion cross sections presented as a function of ion kinetic energy instead of ion velocity. For these data at least, it can be seen that the asymmetric cross section is several times smaller than the two symmetric cross sections. Combined with the fact that the doubly-charged ion component is normally small compared to the singly-charged ion component in the beamlet, the asymmetric charge exchange reaction can be neglected. The rates of singly-charged and doubly-charged charge exchange ion production are given in Eq. 3.33 and Eq. 3.34, respectively.

$$\begin{array}{l} \text{Singly-charged symmetric cross} \\ \text{section.} \end{array} \quad \frac{dn_{cx}^+}{dt} = n_i^+ n_n v_i^+ \sigma^+(v_i^+) \quad \text{Eq. 3.33}$$

$$\begin{array}{l} \text{Doubly-charged symmetric cross} \\ \text{section.} \end{array} \quad \frac{dn_{cx}^{++}}{dt} = n_i^{++} n_n v_i^{++} \sigma^{++}(v_i^{++}) \quad \text{Eq. 3.34}$$

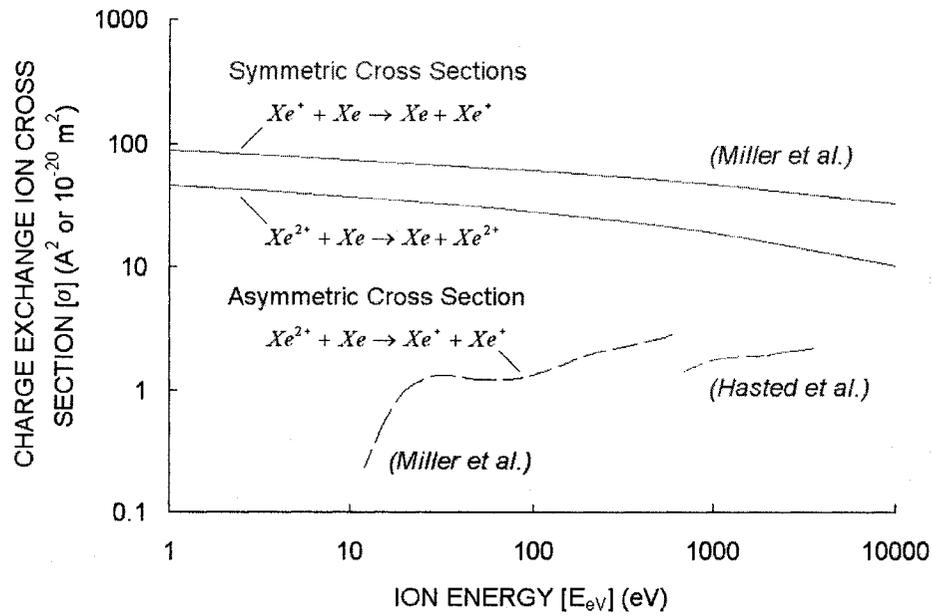


Figure 3.20 Charge exchange cross sections for xenon.

The two symmetric charge exchange cross sections for xenon shown in Figure 3.20 were calculated from equations reported by Miller, given in Eq. 3.35. These are probably the most commonly used cross sections in the ion thruster community. Sakabe and Izawa have reported many symmetric, single electron, charge exchange cross sections, the most relevant of which to ion thrusters are presented in Figure 3.21. Also shown in Figure 3.21 is the cross section for xenon given by Miller. Other commonly used cross sections are those reported by Rapp and Francis.

$$\begin{aligned}\sigma_{Xe^+}(E_{eV}) &= (87.3 \pm 0.9) - (13.6 \pm 0.6) \log(E_{eV}) \\ \sigma_{Xe^{2+}}(E_{eV}) &= (45.7 \pm 1.9) - (8.9 \pm 1.2) \log(E_{eV})\end{aligned}\tag{Eq. 3.35}$$

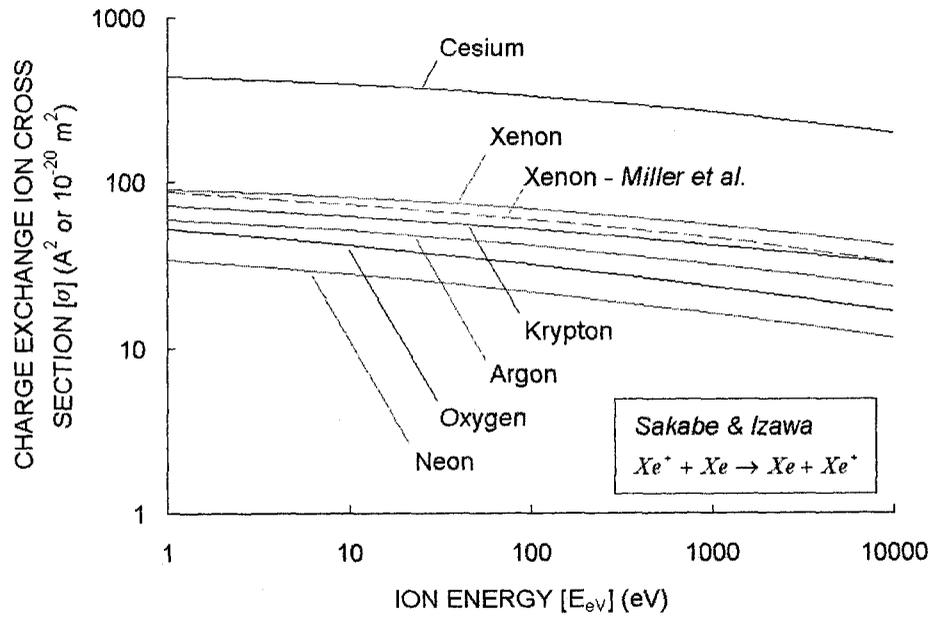


Figure 3.21 Symmetric single electron charge exchange cross sections for various propellants.

Figure 3.22 shows charge exchange production rates for several beamlet currents. Here, the production rates have been put in terms of milliamps per cubic centimeter, mA/cm^3 , by multiplying the number production rate, dn_{cx}/dt , by the charge of an ion. Note that although there is charge exchange ion production upstream of the sheath in some cases, these charge exchange ions are not tracked through the beamlet as they have nearly identical trajectories to the beam ions which also originate upstream of the sheath.

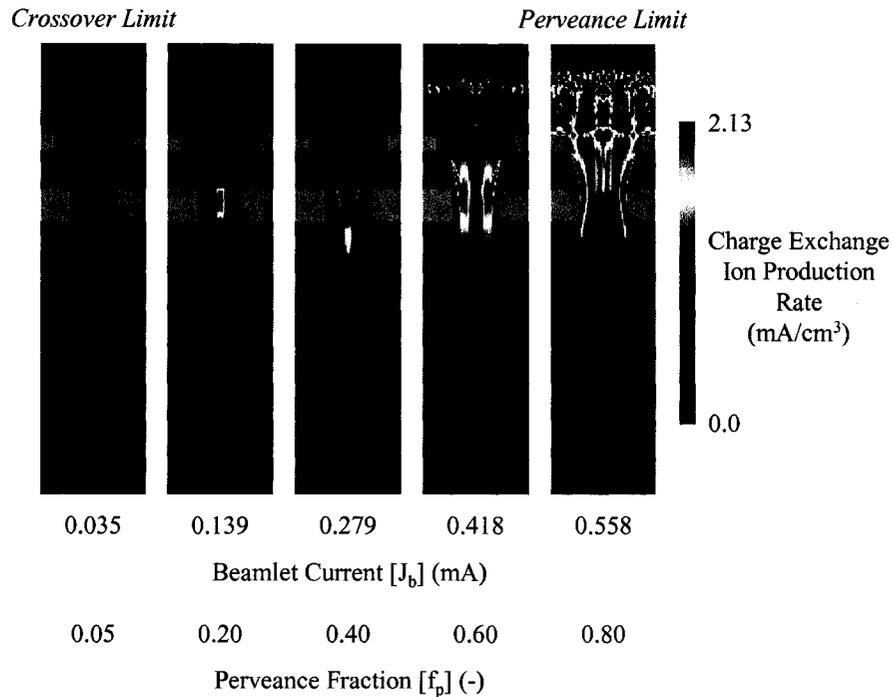


Figure 3.22 Charge exchange ion production rates.

There are several competing effects in the charge exchange ion production rate equation. Charge exchange ion production increases for higher ion densities, which is a function of the beamlet shape, and higher neutral densities, which is greatest furthest upstream. The production rate also increases for higher ion velocities, greatest near the accel grid, and high reaction cross sections, greatest at low ion velocities near the screen grid.

With typical propellant utilization efficiencies near 90 %, the number of charge exchange ions created in the beamlet is usually very small compared to the number of ions passing through the beamlet that originate from within the discharge chamber. To a good approximation, the space charge contribution from charge exchange ions can almost always be neglected relative to the space charge contribution from the larger collection of beam ions.

Charge exchange ions contribute to the screen (J_s), accel (J_a), and beamlet (J_b) currents in the same way ions from the discharge chamber do. Any contribution to the screen, accel, and beamlet currents due to ions from the discharge chamber, which begin upstream of the sheath, will be referred to as “direct” ion current. Similarly, the contribution to these currents from charge exchange ions will be referred to appropriately as “charge exchange” ion current.

In the impingement limit curve of Figure 3.15, it can be seen that the accel grid current, J_a , is non-zero between beamlet currents, J_b , from about 0.1 to 0.5 mA even though the direct current to the accel grid due to ions coming from the sheath is zero as seen in Figure 3.10. This “baseline” accel grid current is due to charge exchange ions striking the grid that were created within the beamlet.

Experimentally, a typical baseline impingement to beamlet current ratio, J_a/J_b , might be about 0.3%. The non-zero pressures inside vacuum tanks increase the neutral density in the region of the ion beamlets, which in turn leads to greater charge exchange ion production and increased accel grid impingement current.

3.2.11 Potential

The electric fields created near the grids, due to the positive and negative biasing of the screen and accel grids, are much larger than any other magnetic or pressure forces that might serve to act on the ions when they are accelerated. This is the reason ion thrusters are referred to as electrostatic thrusters, where static means that the electric fields are constant with time.

The trajectories of ions moving through the beamlet are dictated by electric fields. These electric fields are calculated from a potential function, ϕ . Potential is a scalar function of position, $\phi = \phi(x,y,z)$, and is measured in volts. The potential in the open space of the beamlet is affected by the potentials of the upstream and downstream plasmas and the applied potentials of the screen and accel grids.

The potential is also affected by the ion and electron densities at each location within the beamlet. The effect that ions and electrons have on the potential is called space charge. The potential increases in the presence of ions and decreases in the presence of electrons.

The equation that relates potential, ϕ , to ion and electron space charge is called Poisson's equation. Poisson's equation is a second-order differential equation that comes from Maxwell's first law which applies to electrostatics.

Maxwell's first law relates the electric field at a point with the density of charges at that point, and is given in Eq. 3.36. Here, E is the electric field at the point of interest, ρ is the combined ion and electron space charge at that point, and ϵ_0 is the permittivity of free space. The del operator, ∇ , is basically a partial derivative in all directions. The electric field, E , can in turn be defined in terms of the potential function, ϕ , given in Eq. 3.37. Combining these two equations to eliminate the electric field results in Poisson's equation, given in Eq. 3.38.

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon_0}$$

Maxwell's First Law Eq. 3.36

$$\epsilon_0 = 8.854 \cdot 10^{-12} \left\{ \frac{F}{m} = \frac{C^2}{Nm^2} \right\}$$

$$\vec{E} = -\nabla \phi$$

Electric Field Eq. 3.37

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0}$$

Poisson's Equation Eq. 3.38

Poisson's equation as it is written in Eq. 3.38 is general to any coordinate system. If a Cartesian coordinate system is used, the del and del squared operators (∇ and ∇^2) become those given in Eq. 3.39. Furthermore, the space charge term ρ in this case is the summation of the singly-charged and doubly-charged ion space charges, ρ_i^+ and ρ_i^{++} , and the electron space charge, ρ_e . The electric field equation and Poisson's equation expand to Eq. 3.40 and Eq. 3.41 respectively.

$$\nabla = \frac{\partial}{\partial x} \hat{i} + \frac{\partial}{\partial y} \hat{j} + \frac{\partial}{\partial z} \hat{k}$$

$$\nabla^2 = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

Eq. 3.39

Electric Field
(Cartesian Coordinates)

$$\vec{E} = -\frac{\partial \phi}{\partial x} \hat{i} - \frac{\partial \phi}{\partial y} \hat{j} - \frac{\partial \phi}{\partial z} \hat{k}$$

Eq. 3.40

Poisson's Equation
(Cartesian Coordinates)

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0}$$

Eq. 3.41

Note that the electric field equation results in an electric field vector, while Poisson's equation is a scalar relationship. For example, the electric field in the x-direction at a certain location, $E_x = -\partial\phi/\partial x$, only depends on the change of potential in the x-direction at that location. Also note that the electric field, potential, and space charge terms are all functions of position, i.e., $\phi = \phi(x,y,z)$, $\rho_i^+ = \rho_i^+(x,y,z)$, etc.

Figure 3.23a shows a three-dimensional view of surfaces of constant potential within an example beamlet. The sheath and neutralization surfaces are two examples of surfaces of constant potential. The surfaces in this figure are spaced every 250 V. A somewhat more useful depiction of the potential is obtained by viewing two-dimensional cross sections of the potential in the x-z or y-z plane. An example cross section is shown in Figure 3.23b, where the cut has been taken through the centerline of a beamlet.

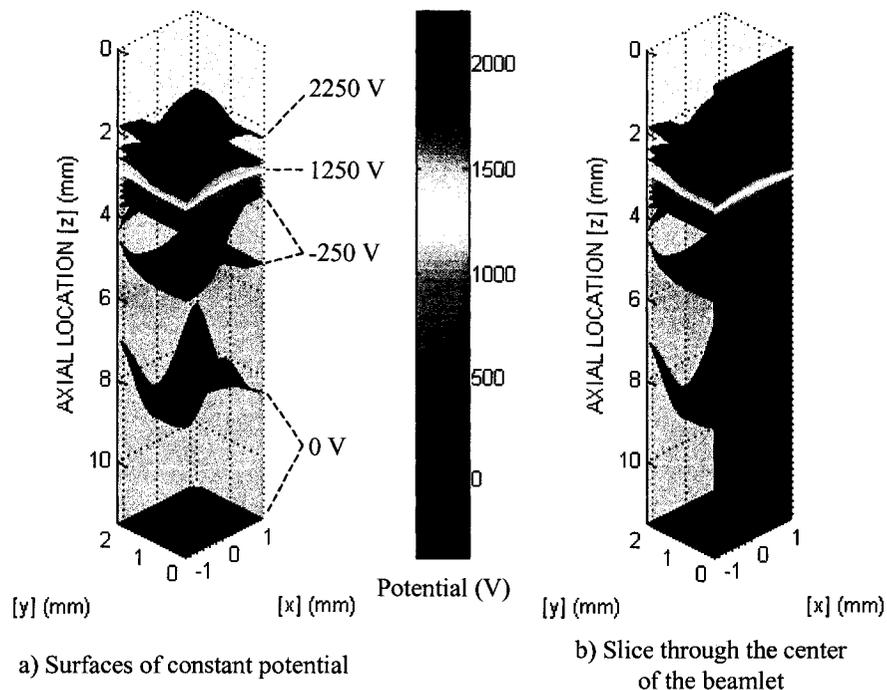


Figure 3.23 Surfaces of constant potential within an example ion beamlet.

Electric fields measure the rate of change of potential over distance. As a result, the distance between evenly spaced potential contours indicates the electric field strength at any location. The electric fields are strongest where the potential contours are closest together, roughly in the region between the grids.

3.2.12 Electron Backstreaming

The accel grid is biased negative relative to the downstream beam plasma potential in order to prevent electrons within the beam plasma from traveling upstream, or backstreaming, into the discharge chamber.

An easy way to see why electrons from the beam plasma would want to travel upstream is to view the potential cross section through the centerline of a beamlet three-dimensionally, as shown in Figure 3.24. Electrons in the beam plasma downstream of the accel grid are roughly at a zero volt potential, which is the beam plasma potential in this case. Electrons are attracted to higher potentials, hence they would travel readily up the hill leading toward the screen grid and discharge chamber plasma if they could. However, the negative bias on the accel grid creates a negative drop in the potential near the accel grid and provides a barrier to backstreaming electrons. In normal thruster operation, the accel grid always has to be kept at a voltage sufficiently negative, relative to the beam plasma potential, to prevent electron backstreaming.

The most likely path for electrons to backstream is usually on the centerline of the beamlet. This is because a) the centerline is the line furthest removed from the accel grid and b) the ion density (and space charge) is usually highest on the centerline. These two effects tend to make the potential along the centerline the most positive relative to any other path between the beam and discharge chamber plasmas.

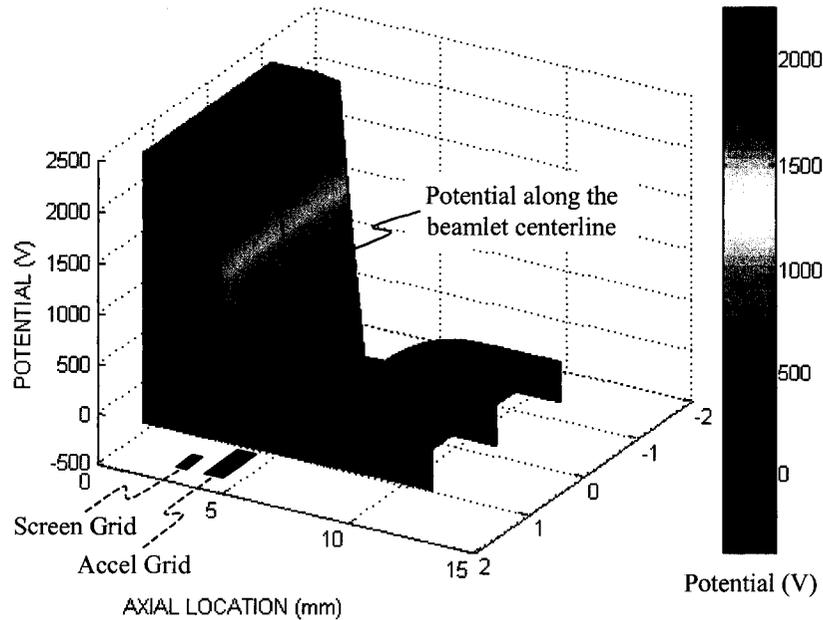


Figure 3.24 The most likely path for electron backstreaming.

The most negative path through the grids is the line that runs between three apertures (in a hexagonal aperture layout). This line is furthest removed from the three surrounding beamlets and the effects of the positive beam ions. The potential along the beamlet centerline and the potential along the line between three apertures are shown in Figure 3.25.

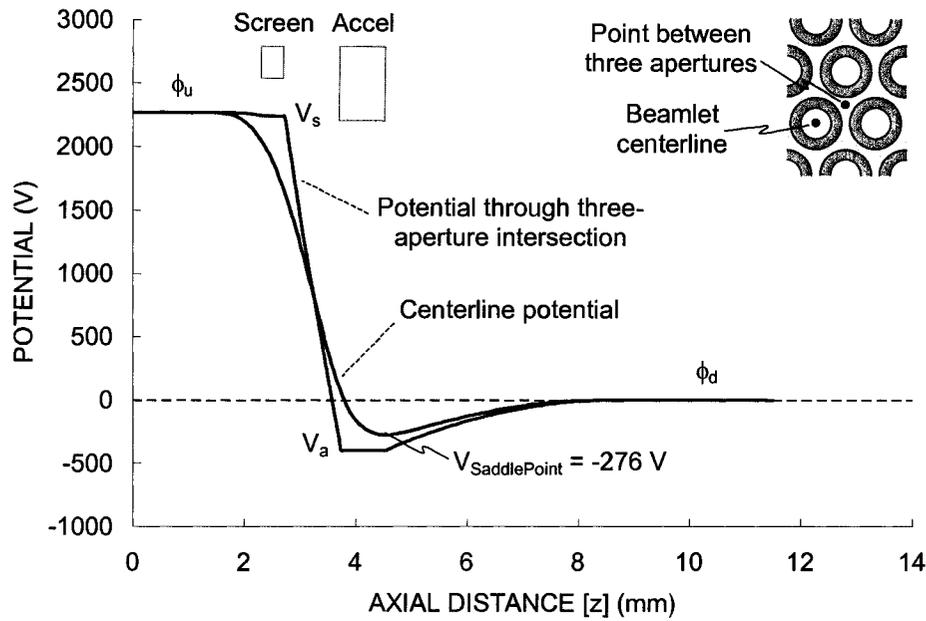


Figure 3.25 Potential along the beamlet centerline and the line that runs between three apertures.

In this case, the minimum potential along the centerline, called the saddle point potential, $V_{\text{SaddlePoint}}$, is -276 V. The beam plasma potential, ϕ_d , is 0 V. The Maxwellian distribution of electrons in the beam plasma is characterized by a temperature, $T_{e,d}$, that is normally a few electron volts or less. For an electron to backstream, it would need to have $\phi_d - V_{\text{SaddlePoint}}$ electron volts, or 276 eV in this example, of energy. For a significant number of electrons to backstream then, the saddle point potential along the centerline needs to be within a few volts of the beam plasma potential.

The saddle point potential, $V_{\text{SaddlePoint}}$, is affected by many variables including the accel grid voltage (V_a), beamlet current (J_b) (or perveance fraction [f_p]), grid spacing (l_g), and accel grid hole diameter (d_a) among others.

Figure 3.26 shows the variation in the saddle point potential as a function of perveance fraction, f_p , for two accel grid voltages, which is the same as a linear variation in the beamlet current. The saddle point potential becomes more positive as perveance fraction increases in general because there is more positive ion space charge within the beamlet. The focusing of the ion beamlet also affects the saddle point potential. Past a certain perveance fraction, the ion beamlet starts to become wider near the accel grid and the ion space charge spreads out, possibly decreasing the potential on the centerline even though the beamlet current is increasing. As a result, the least negative point on the saddle point curve often occurs at some intermediate perveance fraction. In Figure 3.26, the saddle point potential is least negative at a perveance fraction near 0.4.

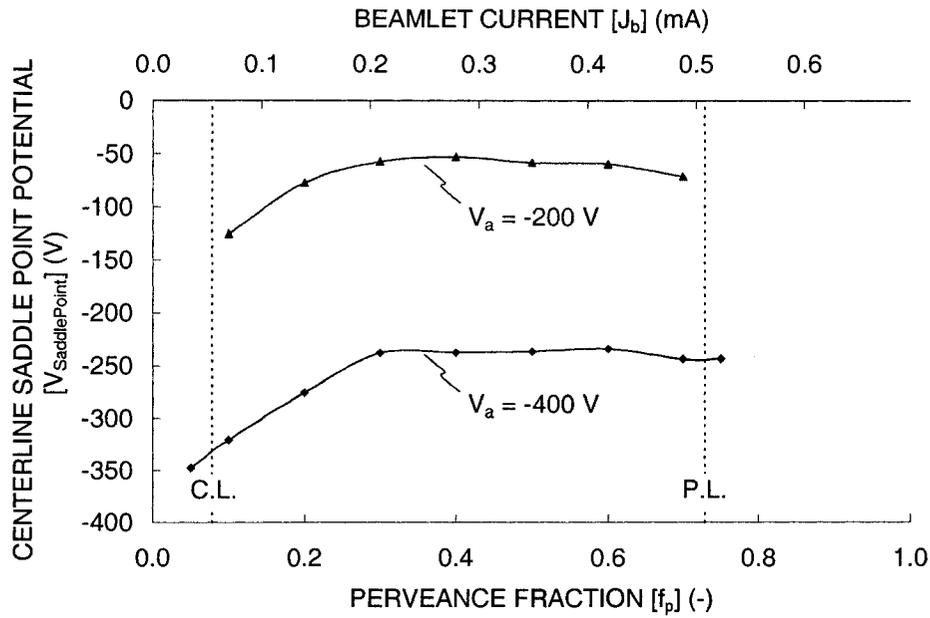


Figure 3.26 Effect of the applied accel grid voltage on the saddle point potential.

The applied accel grid voltage, V_a , tends to have a linear effect on the centerline saddle point potential, $V_{SaddlePoint}$. Figure 3.27 shows the relationship between the accel grid voltage and the saddle point potential for perveance fractions of 0.2, 0.4, and 0.6.

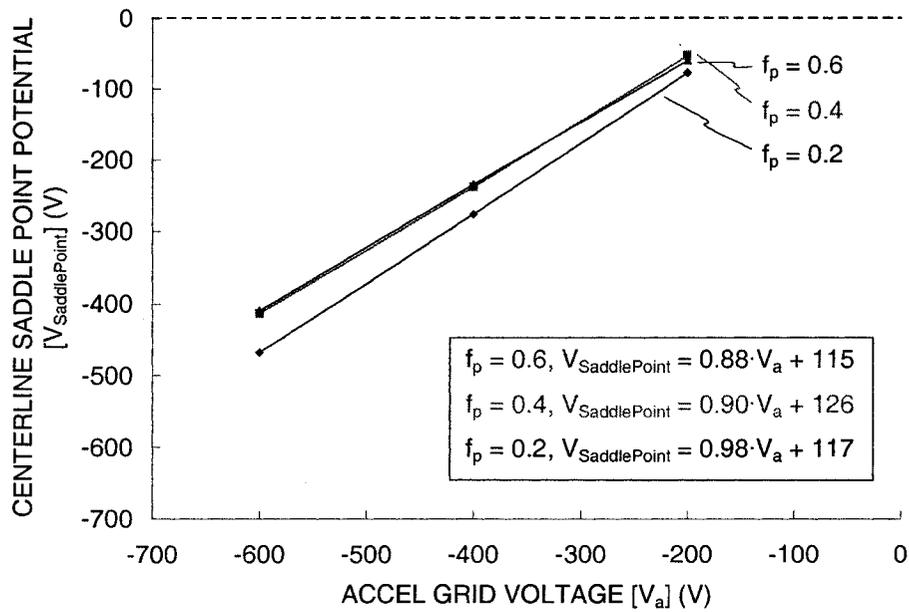


Figure 3.27 Relationship between the applied accel grid voltage and the centerline saddle point potential.

The accel grid voltage required to prevent electron backstreaming, $V_{a \text{ Req.}}$, is the accel grid voltage that makes the saddle point potential equal to, or a few volts less than, the beam plasma potential, ϕ_d , where electron backstreaming would just start to take place. Using the relationships between the saddle point potential and the accel grid voltage such as those developed in Figure 3.27, the accel grid voltage required to prevent electron backstreaming can be developed as shown in Eq. 3.42.

Here, S is the slope and I is the intercept of the linear curve fit equation at each beamlet current, and $V_{SP_Backstreaming}$ is the saddle point potential at which electron backstreaming just starts to begin.

$$\begin{aligned} V_{\text{SaddlePoint}} &= S \cdot V_a + I \\ V_{a \text{ Req.}} &= V_a \left(V_{\text{SaddlePoint}} = V_{SP_Backstreaming} \right) \\ &= V_a \left(V_{\text{SaddlePoint}} \approx \phi_d \right) \end{aligned}$$

Using the equation directly:

$$V_{a \text{ Req.}} = \frac{V_{SP_Backstreaming} - I}{S}$$

If the slope, S , and the saddle point potential at a single accel grid voltage are known:

$$\frac{dV_{SP}}{dV_a} = S \approx \frac{V_{SP} - V_{SP_Backstreaming}}{V_a - V_{a \text{ Req.}}} \quad \text{Eq. 3.42}$$

$$V_{a \text{ Req.}} = V_a - \frac{V_{SP} - V_{SP_Backstreaming}}{S}$$

Figure 3.28 shows the accel grid voltage required to prevent electron backstreaming, $V_{a \text{ Req.}}$, as a function of perveance fraction. At a perveance fraction of 0.3, the accel grid needs to be -141 V or more negative to prevent electron backstreaming. Also shown in this figure is the definition of the voltage margin against electron backstreaming, V_{Margin} . The margin against electron backstreaming is the amount that the accel grid voltage could be reduced from some set value to the point where electron backstreaming begins. For an accel grid voltage of -400 V for example, the accel grid voltage could be made more positive by 259 V before electron backstreaming begins at a perveance fraction of 0.3.

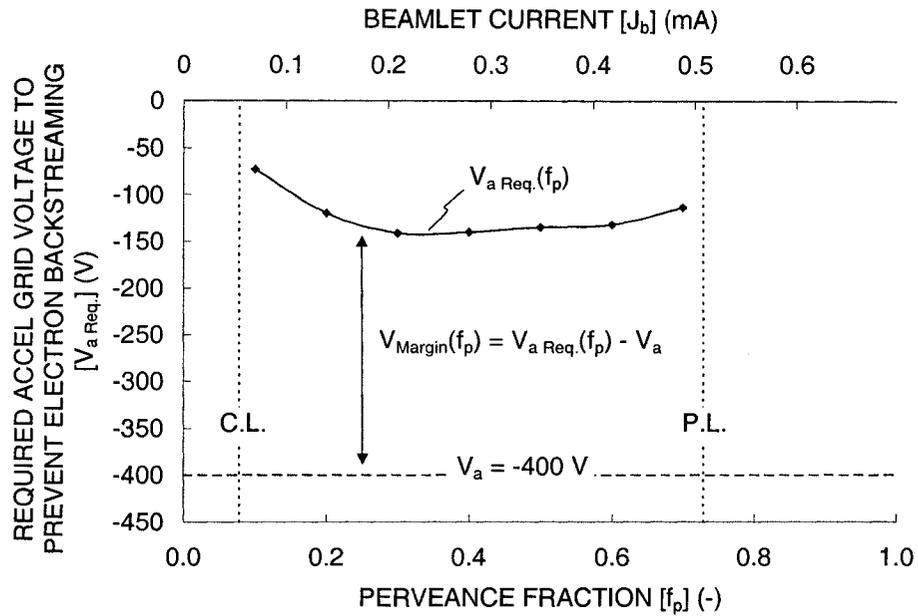


Figure 3.28 Accel grid voltage required to prevent electron backstreaming.

The margin against electron backstreaming, V_{Margin} , which should be a positive value during normal operation, is shown as a function of perveance fraction in Figure 3.29. In terms of individual beamlets, the margin against backstreaming is a function of the perveance fraction. When operating a full grid set however, the first aperture to start backstreaming determines the margin against electron backstreaming.

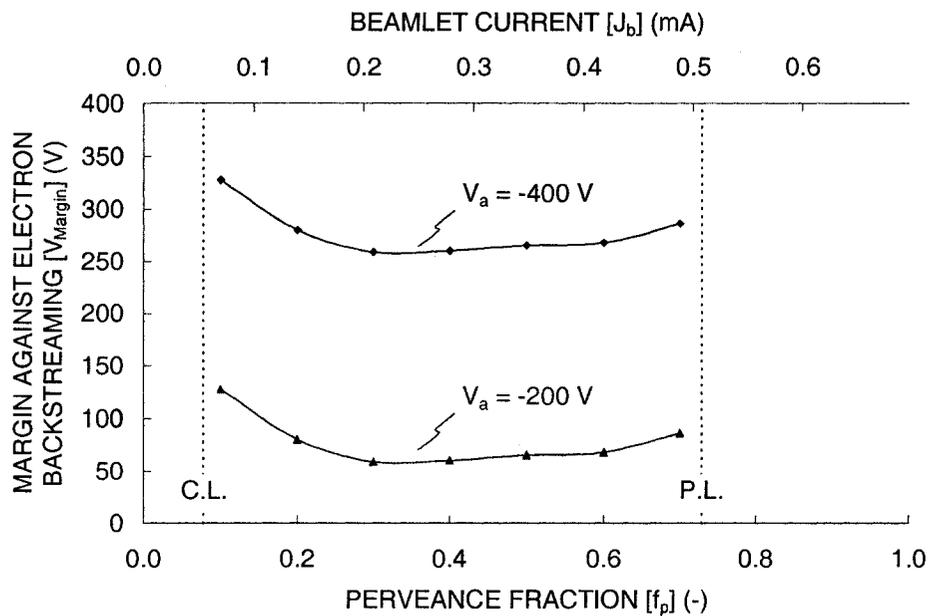


Figure 3.29 Margin against electron backstreaming.

The operating accel grid voltage is often set according to the desired margin against electron backstreaming. Once the voltage at which electron backstreaming begins is determined, $V_{a \text{ Req.}}$, the accel grid voltage might be set to a value that is more negative by 50 V for example.

3.2.13 Divergence Angle and Thrust Factor

There are several ways to quantify how well an ion beam is focused. One of the most common is the divergence angle, α , which is the half-angle that describes the change of the beam radius as a function of axial position.

Figure 3.30 shows the ion density at three axial locations within an example beamlet, one taken at the accel grid exit and the other two farther downstream at a spacing of about 0.9 mm.

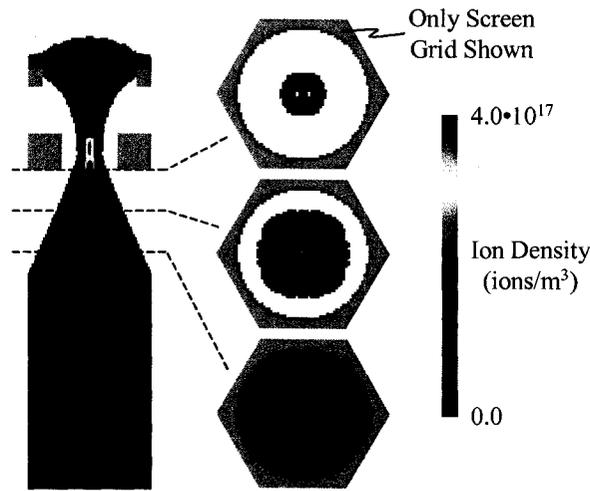


Figure 3.30 Cross sections of the ion density within a beamlet ($J_b = 0.139 \text{ mA}$, $f_p = 0.2$).

In an ion beam, or in an individual beamlet, there is a small fraction of ions that exit the grids at high angles which do not represent the true beam expansion rate. To take this into account, a cutoff point is set, often 95 %, where the edge of the beam is said to be at any axial location. The definition of the beam diameter could be the diameter that encompasses 95 % of the ions, or it could be the diameter where the ion density drops to 95 % of the centerline value for example.

Figure 3.31 shows the current density variation along lines passing through the centerline of the beamlet at four planes downstream of the accel grid. The beamlet diameter on a given plane was found by integrating the profile and determining the radius that encompassed 95 % of the ions on that plane.

The divergence angle can be found from the beamlet diameters. The electric fields that affect the ion trajectories become weaker the farther downstream the ions travel. As a result, it is best to use the measurements farther downstream to calculate the divergence angle. Using the beamlet diameters at the 0.9 and 2.7 mm locations downstream of the accel grid, the divergence angle, α , is about 16.9 degrees.

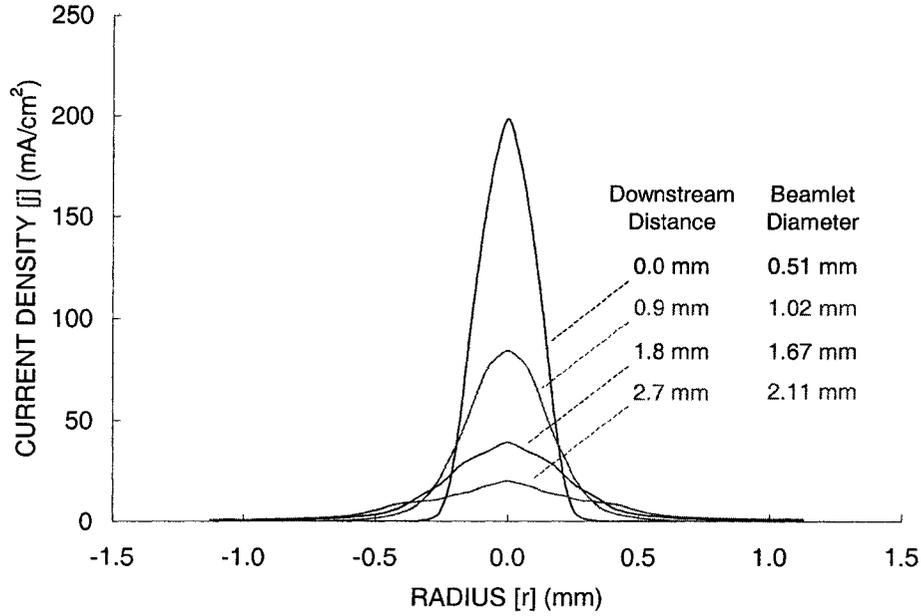


Figure 3.31 Ion density along lines passing through the centerline of the beamlet ($J_b = 0.139$ mA, $f_p = 0.2$).

Downstream of the neutralization surface, the electric fields are relatively small compared to the momentum gained by the ions after they have been accelerated through the grids. As a result, the exit angles of the individual ions are constant on any plane downstream of the neutralization surface. Each ion can be placed into a histogram according to its exit angle. The 95 % divergence angle can be defined as the angle in the histogram for which 95 % of the ion current exits at angles less than or equal to that angle. Figure 3.32 shows the 95 % divergence angle, α , defined in this way as a function of perveance fraction.

Another way to measure the beam divergence is through the thrust factor, ft . The thrust factor indicates how well the beam is aligned axially to the intended direction of thrust. The thrust factor can be found for a beamlet by knowing the exit angle of every ion. A summation is performed over all ions that compares the total ion axial velocity, v_z , with the total ion velocity magnitude, $|v|$, as given in Eq. 3.43. If the velocity magnitude of every ion is the same, which it should be if all of the ions go through the same potential difference, the thrust factor can be written as Eq. 3.44 where N is the total number of ions.

$$ft = \frac{\sum v_z}{\sum \sqrt{(v_r^2 + v_z^2)}} = \frac{\sum v_z}{\sum |v|} \quad \text{Eq. 3.43}$$

$$ft = \frac{\sum v_z}{N|v|} = \frac{1}{N} \sum \frac{v_z}{|v|} \quad \text{Eq. 3.44}$$

The thrust factor, ft , is shown in Figure 3.32 as a function of perveance fraction with the 95 % divergence angle, α . It can be seen that the thrust factor and divergence angle are related as expected. A low divergence angle implies that the thrust factor is high for instance. Typical divergence angles are

between 10 and 20 degrees while typical thrust factors are between 95 and 99 %. Ion beamlets tend to be better aligned at higher perveance fractions.

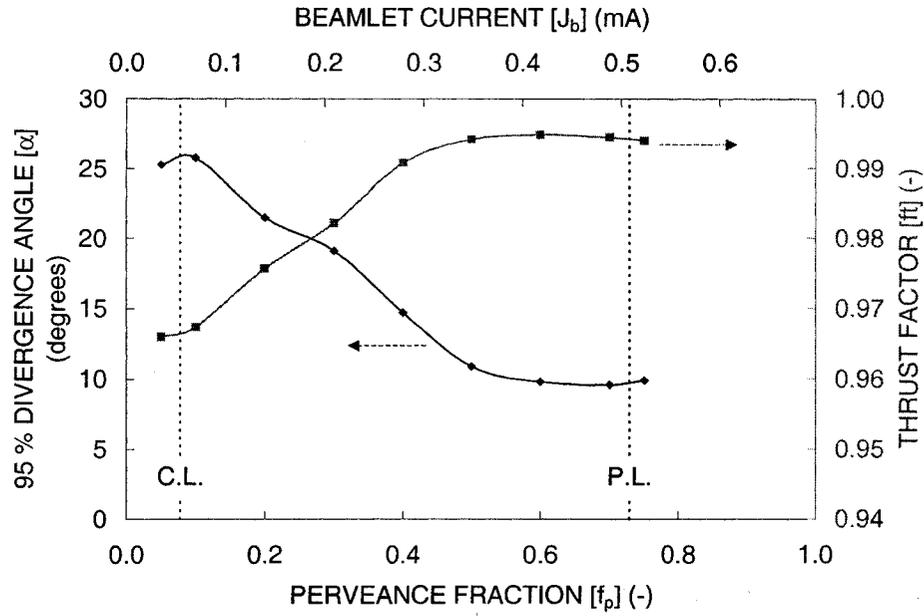


Figure 3.32 Divergence angle and thrust factor as a function of perveance fraction.

3.2.14 Exhaust Velocity

The exhaust velocity of the ions can be calculated from a conservation of energy equation applied as the ions move between the discharge chamber and the beam plasmas, as shown in Eq. 3.45. As stated before, the difference between the potentials of the discharge chamber and beam plasmas is the net accelerating voltage, V_N . Additionally, the mean speed of the ions within the discharge chamber is normally very small compared to the speed gained by the ions moving through the net accelerating voltage, thus it is neglected, which gives Eq. 3.46.

$$q_i \phi_u + \frac{1}{2} m_i v_{i Mean}^2 = q_i \phi_d + \frac{1}{2} m_i v_{ie}^2 \quad \text{Eq. 3.45}$$

$$v_{ie} = \sqrt{\frac{2q_i(\phi_u - \phi_d)}{m_i} + v_{i Mean}^2}$$

$$v_{ie} \approx \sqrt{\frac{2q_i V_N}{m_i}} \quad \text{Eq. 3.46}$$

3.2.15 Thrust

The force provided by an aperture is approximately equal to the mass flow rate of ions passing through the aperture multiplied by their exhaust velocity. Considering only singly-charged and doubly-charged ions, the force is given by Eq. 3.47. The thrust factor, ft , is included in this expression to take into account beamlet divergence, which reduces the actual amount of thrust that is obtained.

The exhaust velocity of the singly-charged and doubly-charged ions can be substituted into the force equation, the exhaust velocity of the doubly-charged ions being greater than the velocity of the singly-charged ions by the square root of two. Also, a simple current balance gives the singly-charged ion current in terms of the total beamlet current, J_b , and the double to single current ratio, J_b^{++}/J_b^+ . Substituting these into the force equation gives Eq. 3.48.

The propellant utilization efficiency, η_u , can be used to give the force in terms of the aperture propellant flow rate, \dot{m} , instead of beamlet current, J_b . Making this substitution gives Eq. 3.49. One more useful thing to do is to calculate the force per unit grid area, given in Eq. 3.50, obtained by dividing the force per aperture by the area of the aperture.

$$F_{Aperture} = ft \left(\frac{J_b^+}{e} m_i v_i^+ + \frac{J_b^{++}}{2e} m_i v_i^{++} \right) \quad \text{Eq. 3.47}$$

$$v_i^+ = \sqrt{\frac{2eV_N}{m_i}} \quad \text{and} \quad v_i^{++} = \sqrt{\frac{2(2e)V_N}{m_i}} = \sqrt{2}v_i^+$$

$$F_{Aperture} = J_b^+ \frac{m_i}{e} \left(1 + \frac{\sqrt{2}}{2} \frac{J_b^{++}}{J_b^+} \right) ft \sqrt{\frac{2eV_N}{m_i}}$$

$$J_b = J_b^+ + J_b^{++} \quad \text{rearranged gives} \quad J_b^+ = \frac{J_b}{1 + \frac{J_b^{++}}{J_b^+}}$$

$$F_{Aperture} = J_b \frac{m_i}{e} \frac{1 + \frac{\sqrt{2}}{2} \frac{J_b^{++}}{J_b^+}}{1 + \frac{J_b^{++}}{J_b^+}} ft \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 3.48}$$

$$J_b = \dot{m} \frac{e}{m_i} \eta_u \frac{1 + \frac{J_b^{++}}{J_b^+}}{1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+}}$$

$$F_{Aperture} = \dot{m} \eta_u \frac{1 + \frac{\sqrt{2} J_b^{++}}{2 J_b^+}}{1 + \frac{1 J_b^{++}}{2 J_b^+}} ft \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 3.49}$$

$$\frac{F_{Aperture}}{Area} = \frac{1}{A_{Aperture}} \dot{m} \eta_u \frac{1 + \frac{\sqrt{2} J_b^{++}}{2 J_b^+}}{1 + \frac{1 J_b^{++}}{2 J_b^+}} ft \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 3.50}$$

These equations were developed for an individual aperture, but they just as easily apply to the entire thruster. The beamlet current, J_b , and the per aperture flow rate, \dot{m} , can be replaced by the total thruster beam current, J_B , and flow rate to obtain the total thrust of the ion thruster, where the thrust factor and propellant utilization efficiency are overall thruster values.

3.2.16 Specific Impulse

Specific impulse is a quantity commonly used in the area of space propulsion because it is useful for describing thruster performance. Specific impulse is defined as the ratio of the thrust provided by the thruster to the Earth weight of the propellant flow. The thrust obtained from an ion thruster is roughly equal to the propellant mass flow rate, \dot{m} , multiplied by the ion exhaust velocity, v_{ie} . Essentially then, specific impulse is a measure of the exhaust velocity of a thruster. Using the expression for the force given in Eq. 3.49, the specific impulse is presented as Eq. 3.51.

$$Isp = \frac{Thrust}{\dot{m} g_E} \approx \frac{\dot{m} v_{ie}}{\dot{m} g_E} = \frac{v_{ie}}{g_E}$$

$$Isp = \frac{F_{Aperture}}{\dot{m} g} = \eta_u \frac{1 + \frac{\sqrt{2} J_b^{++}}{2 J_b^+}}{1 + \frac{1 J_b^{++}}{2 J_b^+}} \frac{ft}{g} \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 3.51}$$

As an example, the net voltage is 2266 V for the example grid set. For singly charged xenon ions, the ion exhaust velocity is thus about 57,700 m/s. For a combined efficiency, η_{uft} , of 0.85, the specific impulse, Isp , is 5000 s. As a comparison, a typical chemical rocket might have an exhaust velocity of 4,000 m/s, giving a specific impulse of 408 s.

3.2.17 Sputtering

Atoms from the surface of a target material can be ejected when the material is struck by high energy atoms or ions. This process is called sputtering. In a vacuum tank, or in space, the atoms from the surface of the target material go into a gaseous phase, where they can then deposit onto nearby surfaces. Sputtering is often used for thin-film deposition when manufacturing semiconductors for example.

With regard to ion thrusters, sputtering is seen most notably on the accel grid. The accel grid is biased negative to both prevent electron backstreaming and to extract current from the discharge chamber. In normal operation, beam ions that originate from within the discharge chamber do not strike the accel grid directly. However, when operating below the crossover limit or above the perveance limit, high energy ions strike the accel grid directly, often with enough momentum to cause sputtering. Charge exchange ions also strike the accel grid with enough momentum to cause sputtering.

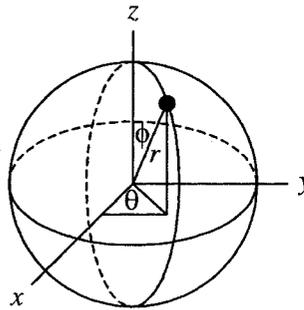
The number of target atoms that get ejected from the surface for every incident ion or atom is called the total sputter yield, or just the sputter yield, and is given the symbol Y . The sputter yield is a statistical value. As a result, the sputter yield could be something like 2.5 atoms per incident ion for example.

There are several factors that determine how much sputtering takes place. The weight of the incident ion (or atom) and the weight of the target atom both affect the sputter yield. The sputter yield is also a function of the incident ion velocity and the angle at which the ion strikes the target surface, called the angle of incidence. The sputter yield is also affected by the temperature of the target material.

The atoms that are ejected from the target surface leave with a distribution of angles. Some atoms may be ejected in a direction perpendicular to the surface while others may be ejected at an angle of 45 degrees away from the target normal vector. The distribution of atoms ejected from the target is described by the differential sputter yield.

The total sputter yield has units of atoms per ion (atoms/ion). The differential sputter yield takes into account the distribution of atoms in terms of a solid angle, hence it is measured in atoms per ion per steradian (atoms/(ion· Ω)). In a full sphere, there are 4π steradians, which is the surface area of a unit sphere, as shown in Eq. 3.52. The total sputter yield, Y , is the integral of the differential sputter yield, $y(\phi, \theta)$, over the solid angle above the target surface, as given in Eq. 3.53.

$$\Omega_{Sphere} = \int_0^{2\pi} \int_0^{\pi} \sin \phi \, d\phi \, d\theta = 4\pi$$



Eq. 3.52

$$Y = \iint_S y(\phi, \theta) \sin \phi \, d\phi \, d\theta \quad \text{Eq. 3.53}$$

The most common grid material is molybdenum because it has reasonable sputter characteristics and has a high density compared to many other metals. For two materials that have the same sputter yield, the higher density material will last longer because there are more atoms per unit volume available to be sputtered.

In the last few years, many grids intended for very long lifetime applications are being made out of various forms of carbon. Even though carbon is not very dense compared to molybdenum for example, the sputter rate is many times lower than molybdenum and the end result is that the carbon materials last longer. Carbon based materials are used within vacuum chambers to shield sensitive components and act as beam targets because the very low sputter rates result in the least amount of contamination. Two such carbon materials are pyrolytic graphite and carbon-carbon composites.

The total sputter yields, Y , for xenon ions impinging upon molybdenum (Mo), titanium (Ti), and carbon (C) target materials as a function of ion energy are shown in Figure 3.33. According to these data, for an ion energy of 300 eV the molybdenum sputter yield is about 15 times greater than that of carbon. The data in this figure are from Yamamura; recent work has been done by Doerner, Kolasinski, and Zoerb.

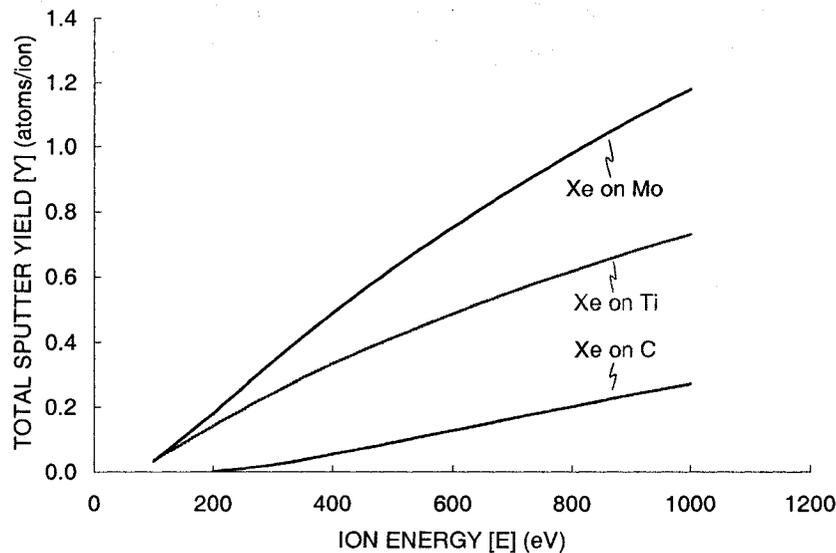


Figure 3.33 Total sputter yield for xenon ions impinging upon three target materials.

The sputter yields alone do not imply that carbon is the best material to use because they do not take into account the density of the materials. A better measure of the material is its volume erosion rate, dV/dt . An equation for the volume erosion rate, in $m^3/s/A$, is given as Eq. 3.54, where J is the current of impinging ions, $Y(E)$ is the sputter yield as a function of the impinging ion energy, m_a is the mass of an individual target atom, and ρ is the density of the target material.

$$\frac{dV}{dt} = J \frac{1}{e} Y(E) m_a \frac{1}{\rho} \quad \text{Eq. 3.54}$$

Figure 3.34 shows the volume erosion rate as a function of ion energy for the same three target materials as in Figure 3.33. The volume erosion rate has been normalized to the impinging ion current, resulting in a quantity that only depends on the sputter yield and material properties. For an ion energy of 300 eV, the molybdenum volume erosion rate per unit current is about 27 times greater than that of carbon.

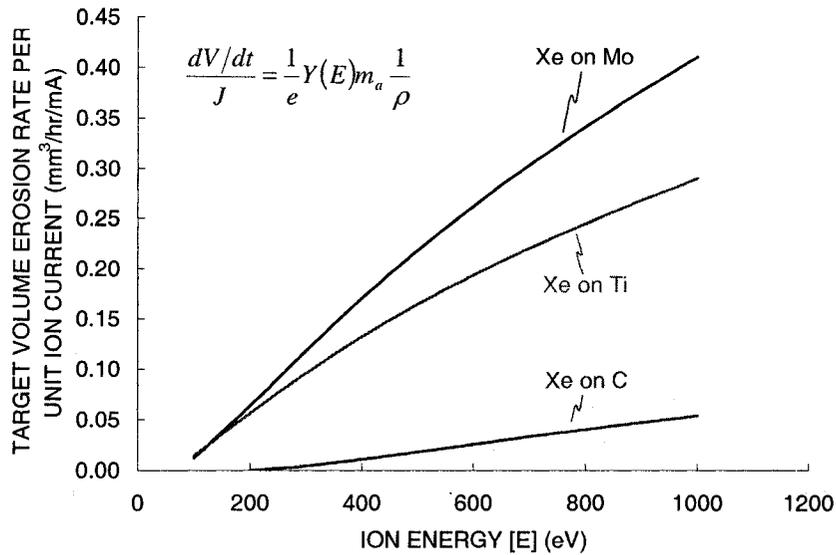


Figure 3.34 Volume erosion rate for xenon ions impinging upon three target materials.

The sputter yields in Figure 3.34 are for normally incident xenon ions. However, the total sputter yield changes with ion angle of incidence. One way to account for this is to use a sputter yield multiplier, such that the actual yield is the normal incidence yield multiplied by an adjustment factor. Figure 3.35 shows a sputter yield multiplier for xenon ions impinging a molybdenum target as a function of the ion angle of incidence at a xenon ion energy of 300 eV. In this case, the sputter yield is a maximum at an angle near 45 degrees where the actual sputter yield is about 84 % greater than the sputter yield calculated for normally incident ions. Similar trends have been observed for other ion and target combinations.

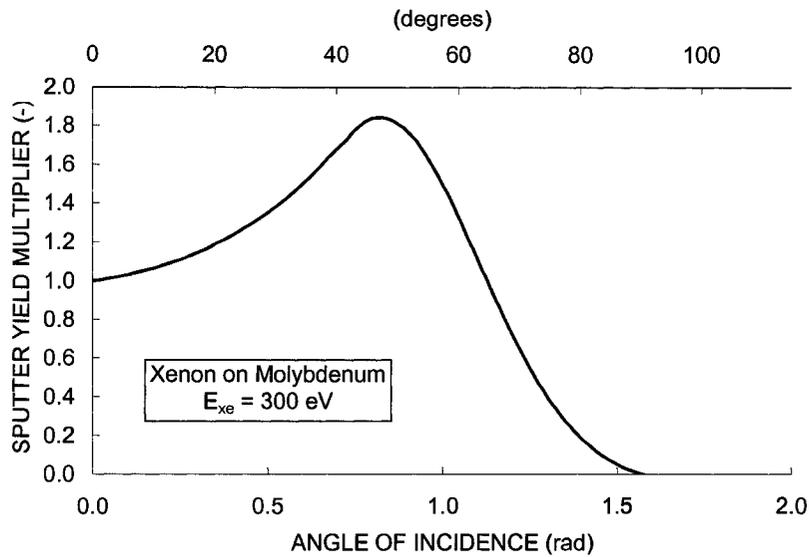


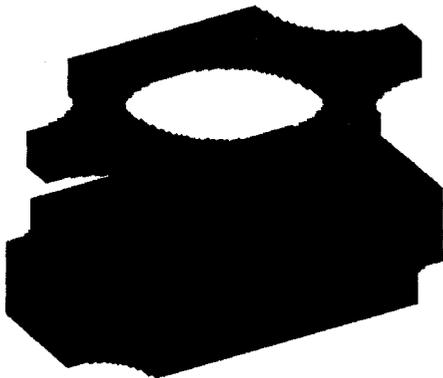
Figure 3.35 Total sputter yield multiplier for xenon ions impinging upon a molybdenum target.

A cosine distribution is used to model the ejection of sputtered grid material from grid surfaces. This model is described in a separate section.

3.3 Grid Lifetime

One of the greater concerns with long lifetime ion thruster operation is the failure of the grids as a result of ion erosion. Charge exchange ion impingement onto the accel grid causes sputter erosion which can lead to several possible grid failure mechanisms.

An example of grid erosion is shown in Figure 3.36. The beamlet in this figure was operated for 14,400 hours, or about 1.64 years. The cells that are colored in have been completely eroded away from the grid set. In this case most of the erosion takes place on the downstream side of the accel grid, with a small amount of erosion also seen on the downstream edge of the hole. This erosion pattern is typical of beamlets operating at low perveance fractions.



Grid material that has been eroded away (colored) shown with grid material that still remains (gray).



All of the colored cells have been completely eroded away from the grids. The color of the cell represents the axial location of the cell only.

Figure 3.36 An example of grid erosion. $J_p = 0.139$ mA ($f_p = 0.2$), $T = 14,400$ hours (1.64 years).

With a hexagonal aperture layout, a “pit and groove” erosion pattern is often seen. Figure 3.37 shows the pits and grooves labeled for the example erosion case. Six pits, and six grooves running between the pits, surround each aperture.



Another way to view the grids at any time is to color the surface cells according to axial location.

Note that the same depiction of the grid erosion is obtained by either a) looking downstream at the grid material that has been eroded away or b) looking upstream at the grid material that still remains.

(Viewing the grid cells that have been completely eroded away)

(Looking upstream at the grid material that remains)

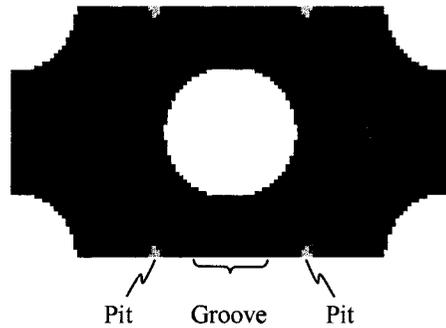
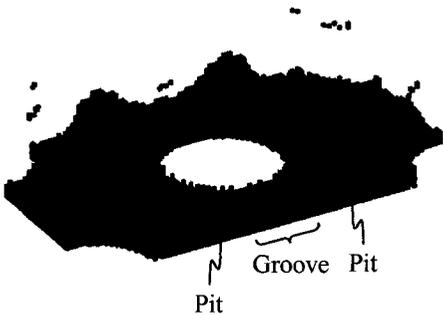


Figure 3.37 Pit and groove erosion pattern.

3.3.1 Grid Failure Mechanisms

There are several possible grid failure mechanisms. Erosion of the grids due to ion impingement can lead to a) electron backstreaming, b) structural failure, and c) grid shorting [Foster].

Electron backstreaming occurs when the negative potential applied to the accel grid no longer prohibits electrons within the downstream beam plasma from backstreaming toward the higher potential

discharge chamber plasma. These electrons can gain significant energy in traveling through the large potential difference between the plasmas and can deposit significant power into the discharge chamber and overheat components including the discharge chamber cathode.

Electron backstreaming can occur as a result of a combination of aperture diameter enlargement and grid thinning. The particular cause of the electron backstreaming is specific to the grid design. For example, aperture enlargement might be more important than grid thinning when operating at a high beamlet current where there is significant hole barrel impingement from charge exchange ions created between the grids.

The second failure mode is referred to as structural failure. Erosion of the accel grid by charge exchange ions in particular typically results in pit and groove erosion. When the grooves erode completely through the accel grid thickness, the aperture can detach from the rest of the accel grid. General weakening of the accel grid due to erosion can also lead to a change in grid deflection.

The third failure mechanism, which is related to structural failure, occurs when an electrical short between the grids cannot be cleared. Flakes that lodge between the grids can come from cathode erosion or accel grid erosion for example. For instance, if the grooves wear completely through an aperture, the unattached portion can end up causing an un-clearable short.

The onset of electron backstreaming is relatively easy to define in the ffx simulation code. Electron backstreaming is said to start to occur when the saddle point potential along the beamlet centerline becomes close to (within a few volts of) the downstream beam plasma potential.

Structural failure is harder to concretely define than electron backstreaming. The pit and groove erosion rates can be monitored and extrapolated to determine the point where the pits and grooves will completely wear through the accel grid thickness. Another convention is to take structural failure to occur when a certain percentage, 50 percent is used here, of the accel grid mass has been eroded away.

3.3.2 A Useful Lifetime Parameter

The goal of operating an ion thruster for a long period of time is really to create a significant change in the velocity of a spacecraft. A quantity useful for comparing the lifetime of grids of different size is now developed.

The change in momentum per unit area obtained from an aperture is the integral of the force per unit area over time. If the force per unit grid area is taken to be constant over the life of the aperture, the change in momentum per unit area simply becomes the force per unit area multiplied by the lifetime, L , as shown in Eq. 3.55. In this equation, the propellant utilization efficiency of the aperture, mass of an ion, electronic charge, and net accelerating voltage can all be considered as being the same when comparing two grid sets. Additionally, the thrust factor and double to single current ratio are likely to be similar. Thus, to maximize the change in momentum per unit area, one would want to maximize the product of the lifetime, L , beamlet current, J_b , and the inverse of the aperture area, $A_{Aperture}$, as shown in Eq. 3.56.

$$\frac{\Delta p}{A} = \int \frac{F}{A} dt$$

$$\frac{\Delta p}{A} = \frac{L}{A_{Aperture}} J_b \frac{m_i}{e} \frac{1 + \frac{\sqrt{2} J_b^{++}}{2 J_b^+}}{1 + \frac{J_b^{++}}{J_b^+}} ft \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 3.55}$$

$$\text{Max}\left(\frac{\Delta p}{A}\right) \approx \text{Max}\left(\frac{L \cdot J_b}{A_{Aperture}}\right) \quad \text{Eq. 3.56}$$

If a grid has fixed geometry (where $A_{Aperture}$ is fixed) and the goal is to find the best operating beamlet current for that geometry, one would maximize the product of the lifetime and the beamlet current. This could be very useful for deciding what perveance fraction to run at given a certain grid geometry.

Given this result, it is not immediately clear whether operating at a low or high perveance fraction is better. The answer lies in the complicated physics of things like the beamlet shape at a certain beamlet current, the charge exchange ion current that is created within the beamlet, the electric fields that accelerate the charge exchange ions, and so on.

The total propellant use at any time T, including unionized neutral atoms through the propellant utilization efficiency η_u , is given in Eq. 3.57.

$$m_{\text{Propellant}} = T \frac{1 + \frac{1 J_b^{++}}{2 J_b^+}}{1 + \frac{J_b^{++}}{J_b^+}} \frac{J_b m_i}{\eta_u e} \quad \text{Eq. 3.57}$$

3.3.3 Erosion as a Function of Current

Figure 3.38 shows the progression of the erosion pattern for the example case operating at a perveance fraction of 0.2. In the left set of pictures, the current accel grid is shown without the screen grid. In the right set of pictures, the grid material that has been eroded away is shown. A complete accel grid would result at any time step if the two parts were added together. Viewing the eroded material is often more visually useful than viewing the current grid because the erosion along the inside of the hole tends to stand out a little better.

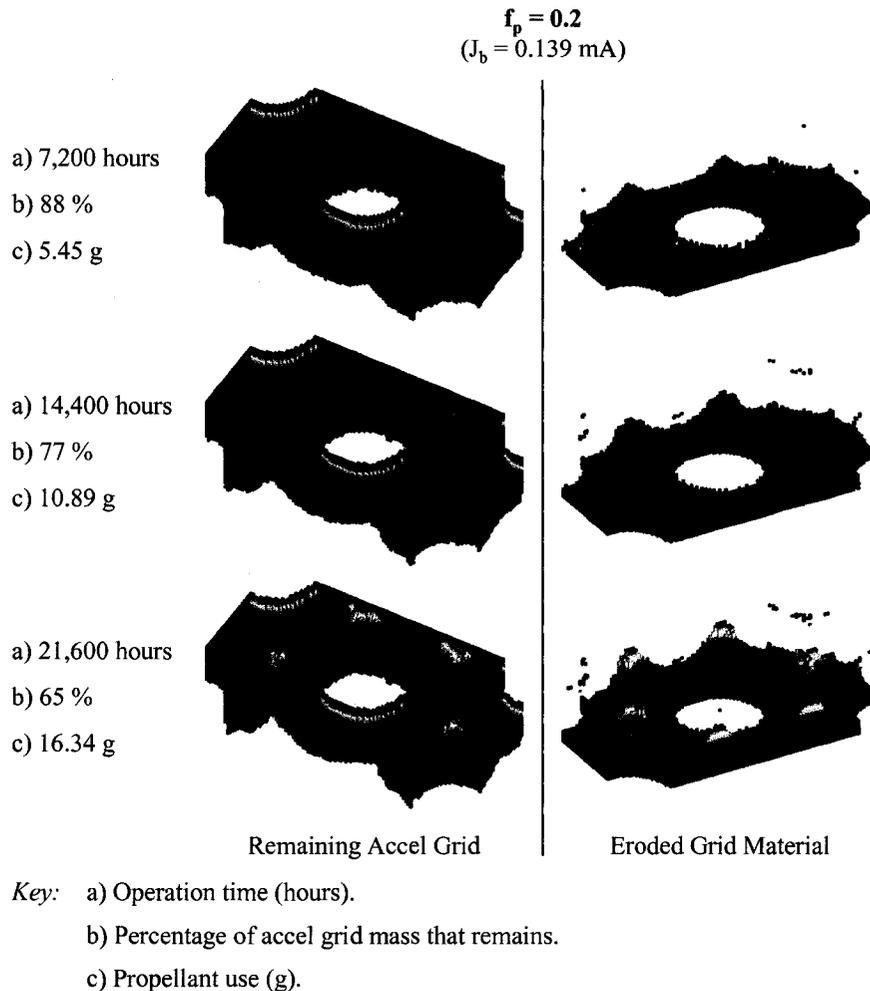


Figure 3.38 An example of erosion as a function of time.

The erosion in Figure 3.38 is shown at 7,200, 14,400, and 21,600 hours of grid operation, listed as part a) to the left of the pictures. Part b) lists the percentage of the accel grid mass that remains at each time. Structural failure of the grid is taken to occur here as the point when 50 percent of the accel grid mass has been eroded away. Part c) lists the total amount of xenon propellant that has passed through the aperture.

Erosion patterns are shown in Figure 3.39 for permeance fractions of 0.2, 0.4, and 0.6. Along each row of pictures, the percentage of accel grid mass remaining is approximately the same, listed in the lower left hand corner of each picture. Because the beamlets operate at different currents, and have different amounts of charge exchange ion impingement, the length of operation time varies to reach the same amount of accel grid mass loss. For instance, the beamlet at a permeance fraction of 0.2 is operated for a much longer period of time to reach 50 % mass loss than the beamlet at a permeance fraction of 0.6.

The other value listed is the total amount of propellant that has passed through each aperture. The grid operating at a perveance fraction of 0.6 processes only about 23 % as much propellant as the grid operating at a perveance fraction of 0.2 for similar amounts of accel grid mass loss. For reference, the grid operating at the lowest perveance fraction of 0.2 takes 21,960 hours to process 16.61 g of propellant.

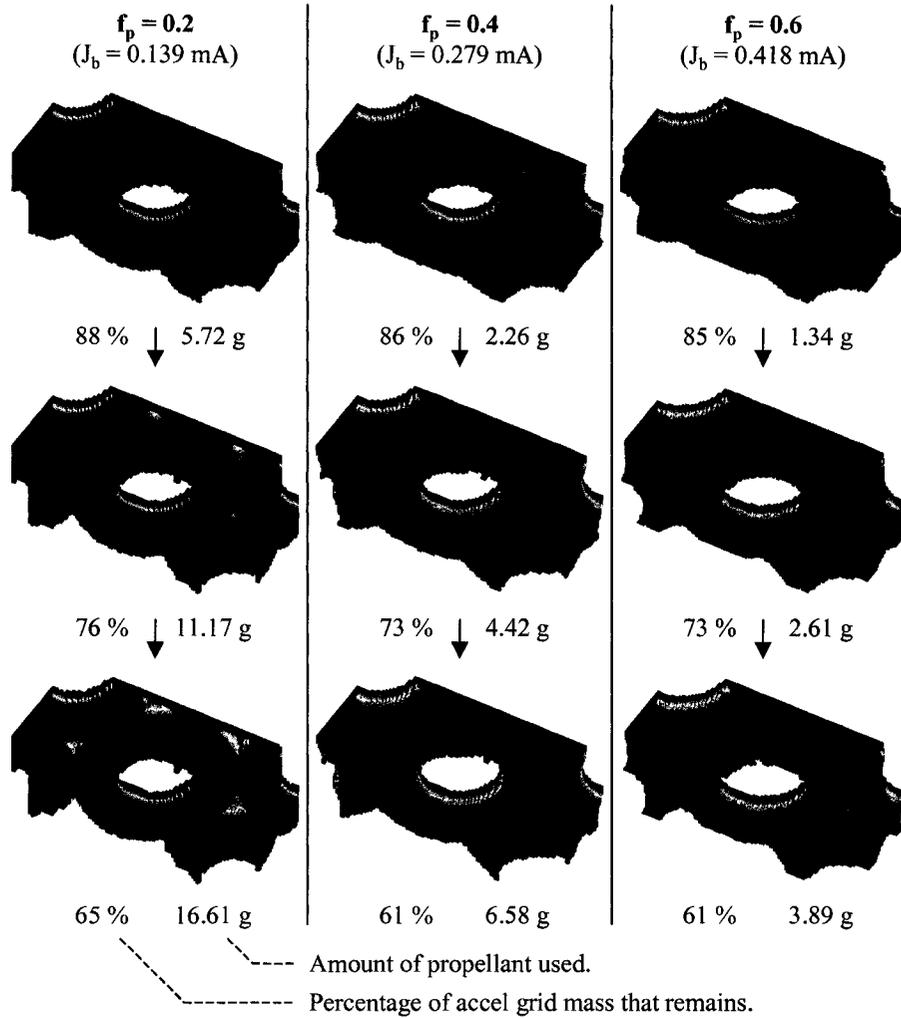


Figure 3.39 Accel grid erosion patterns as a function of time for three beamlet currents.

A second depiction of the erosion that occurs at the three beamlet currents is given in Figure 3.40 showing the accel grid mass that has been eroded away. In this figure, it is easy to see the differences in the pit and groove erosion. On average for this particular grid setup, the pits are about 1.5 times deeper than the grooves. Also notice that as perveance fraction increases, the grooves become narrower and the pit erosion is more concentrated.

Another thing to notice in Figure 3.40 is the difference in the amount of barrel erosion that takes place at the three perveance fractions. The amount of barrel erosion increases significantly at the greater perveance fractions. This erosion is caused largely by charge exchange ions created between the grids, upstream of the accel grid. Figure 3.19, given previously, shows charge exchange ion termination surfaces

as a function of the location where charge exchange ions are produced. As perveance fraction increases in this figure, greater fractions of charge exchange ions from farther upstream are directed into the accel grid upstream surface (shown in yellow) and barrel region (shown in orange). These ions have significant energy when they strike the accel grid because they are created in a region of very high potential. As a result, the accel grid hole barrel can be eroded away quickly, essentially enlarging the original hole size.

At the higher perveance fractions, much of the sputtered accel grid material can deposit onto the screen grid. Over time, this material can build up to the point where the code can add new screen grid cells (or accel grid cells, etc.) if the rate of re-deposition is great enough.

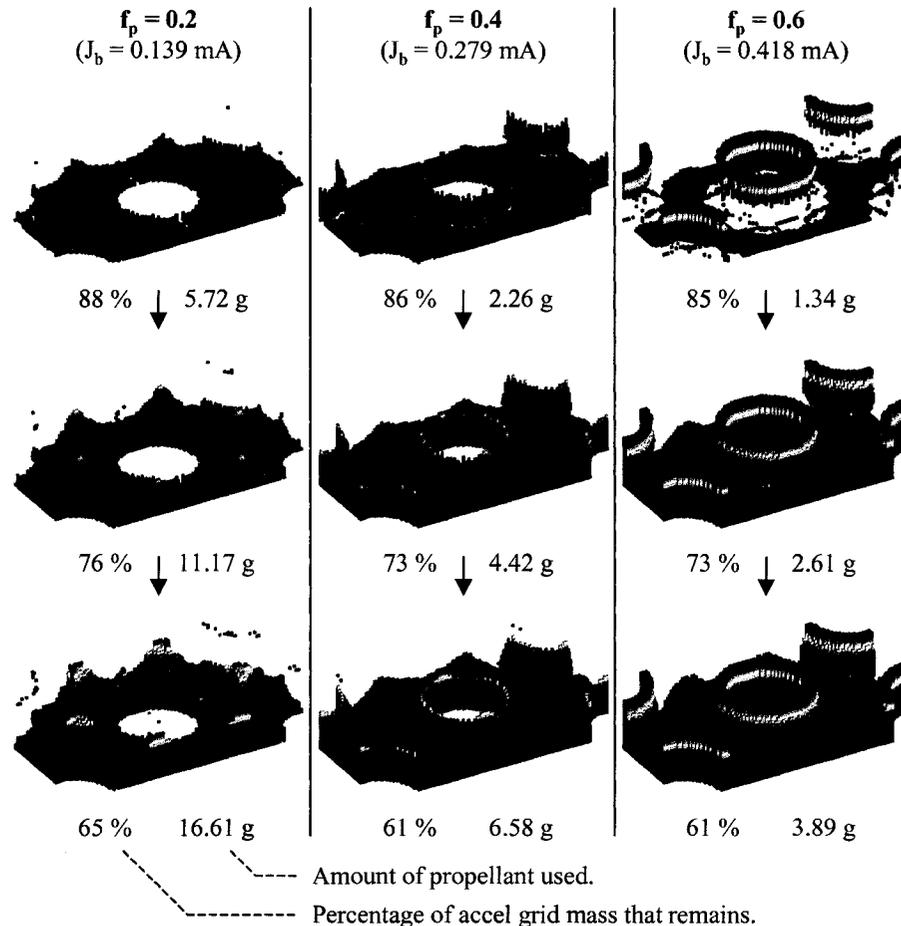


Figure 3.40 Eroded accel grid material for three beamlet currents.

Figure 3.41 compares the rates of accel grid mass loss for beamlets operating at perveance fractions from 0.1 to 0.7. Structural failure of the grid is taken to occur at 50 % accel grid mass loss. The independent variable on the x-axis is the total propellant throughput per aperture, which does use the propellant utilization efficiency to include any unionized propellant that flows through the aperture. Considering this quantity alone, it is better to operate the thruster at a lower perveance fraction in order to obtain the greatest change in momentum.

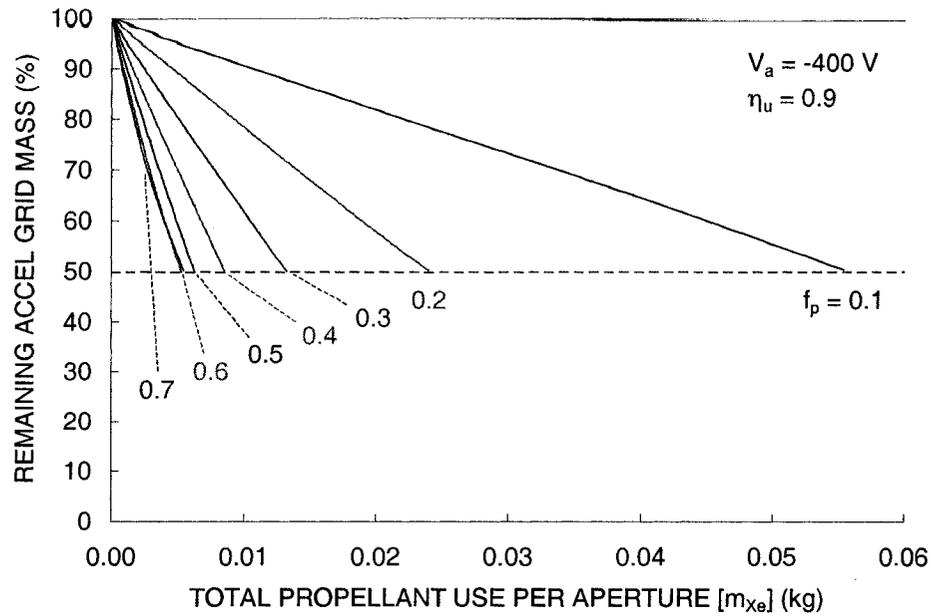


Figure 3.41 Accel grid mass loss rate for several beamlet perveance fractions.

Figure 3.42 shows how the saddle point potential along the beamlet centerline changes as a function of operation time for several perveance fractions. The saddle point potential increases with time because charge exchange ion erosion causes a) the accel grid hole diameter to increase and b) the accel grid thickness to decrease. Once the saddle point potential, of any beamlet on the thruster, reaches the beam plasma potential (or comes within a few volts) electron backstreaming will begin. This signals the electron backstreaming grid failure mechanism unless something is done to mitigate the backstreaming, such as further decreasing the accel grid voltage, increasing the grid spacing, or possibly decreasing the beam current for example.

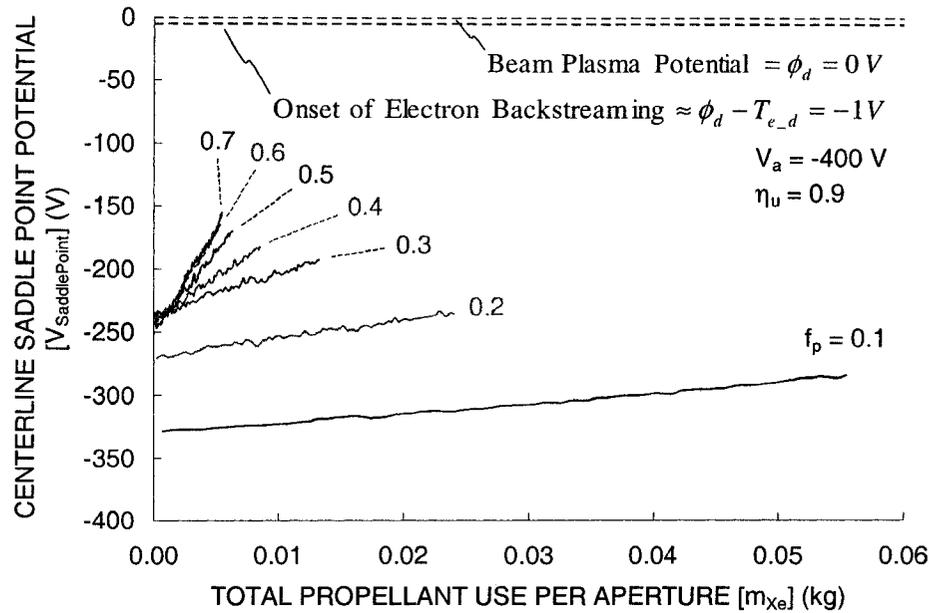


Figure 3.42 Saddle point potential for several perveance fractions.

In this case, a rather large accel grid voltage of -400 V was used. By comparing the charts for accel grid mass loss and saddle point potential, it is clear that structural failure of the accel grid will occur before electron backstreaming begins. Had the accel grid voltage not been so negative, electron backstreaming might have been the limiting grid failure mechanism.

The reason it is better to operate at a low perveance fraction is that the impingement to beamlet current ratio, J_a/J_b , is lower there than it is at higher perveance fractions. This means that for an equivalent amount of propellant throughput, the accel grid operating at the lower perveance fraction has less total ion impingement than an accel grid operating at a higher perveance fraction.

Figure 3.43 shows impingement to beamlet current ratios as a function of grid operation time. Note that all of the beamlets in this figure were operated with a propellant utilization efficiency of 0.9. There are several reasons why the impingement current ratios decrease with operation time. Charge exchange ion production current is linearly related to the neutral density near the grids. Charge exchange ion erosion causes a) the accel grid to become thinner and b) the accel grid hole diameter to increase. Both of these effects result in a decrease in the neutral density, in turn resulting in less charge exchange ion production. Additionally, at greater perveance fractions, charge exchange ions created farther upstream toward the screen grid start to miss the accel grid as the hole is enlarged from erosion, further reducing the accel grid impingement current.

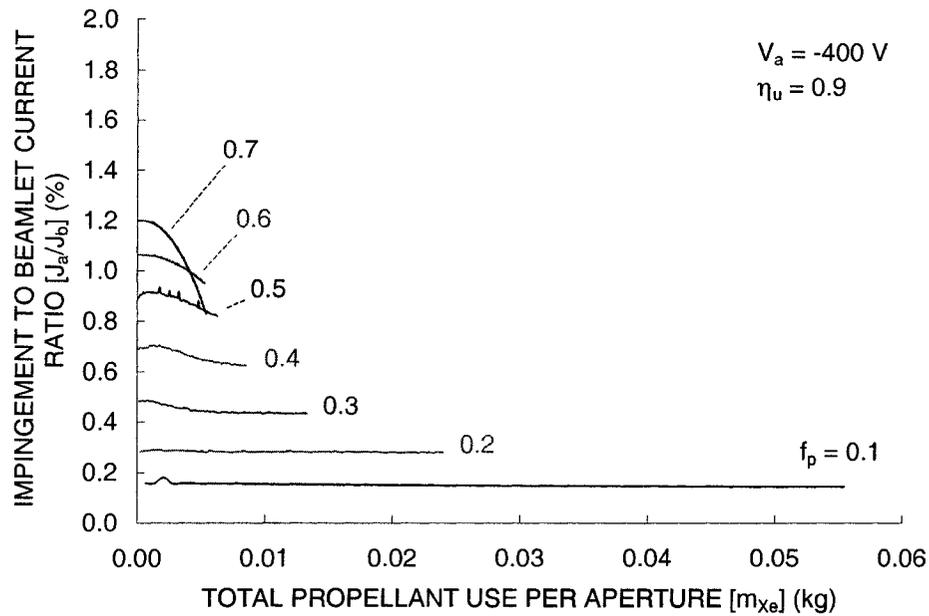


Figure 3.43 Impingement current variation over time.

There is a unique occurrence at perveance fractions near the perveance limit, where the impingement to beamlet current ratio, J_a/J_b , is seen to increase with operation time. Figure 3.44 shows an ion beamlet at three operation times: beginning of life at 0 hours, 120 hours, and 240 hours. This beamlet is operating at a perveance fraction of 0.75, or a beamlet current of 0.523 mA. At high perveance fractions, a significant amount of erosion can occur on the upstream side of the grid. In this case, once the grids eroded to a certain point, near 200 hours of operation, the electric fields changed to the point where the perveance limit decreased far below the operating beamlet current of 0.523 mA.

Erosion of the accel grid progresses rapidly when maintaining a constant beamlet current above the perveance limit. This is because the impingement current has to be very high in order to get the required current of ions through the aperture. In a real situation, the number of ions arriving at the aperture from the discharge chamber would be constant, resulting in a decrease in the beamlet current as some of the ions start to directly impinge upon the accel grid.

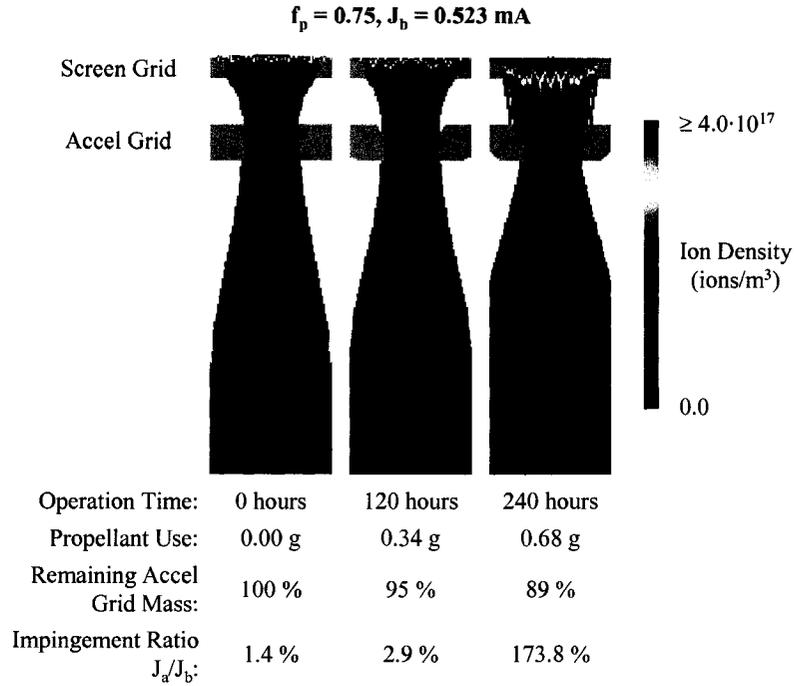


Figure 3.44 Impingement current increase with time at a perveance fraction of 0.75.

Figure 3.45 shows the erosion rates of the accel grid upstream, hole barrel, and downstream surfaces as a function of perveance fraction. The erosion rates of these surfaces have been divided by the beamlet current at each perveance fraction, giving erosion rates in mg/khr/mA. Per unit current, the erosion rates at the lower perveance fractions are lower. Instead of operating one hole at a beamlet current of 0.4 mA for example, it is better to operate two holes at 0.2 mA because the total erosion rate will be lower.

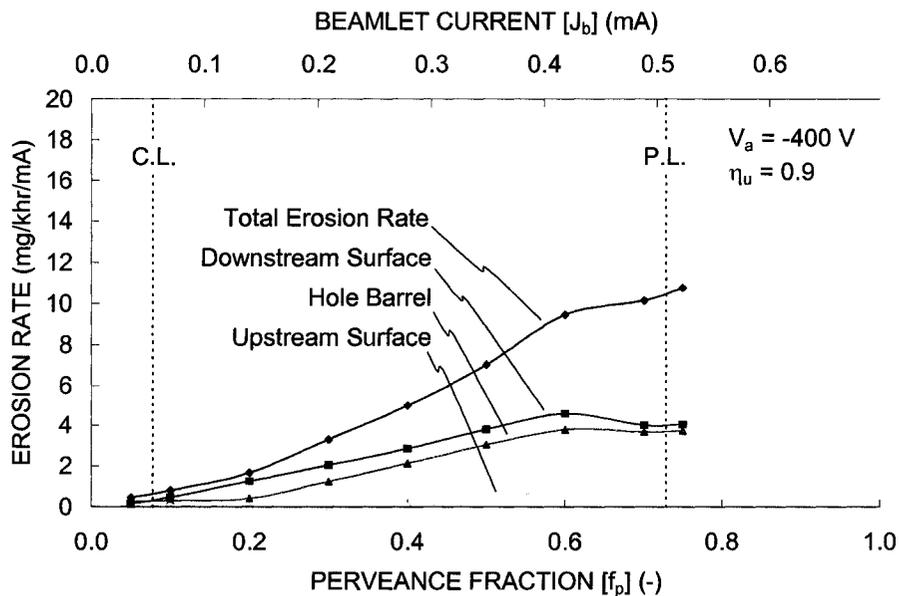


Figure 3.45 Accel grid surface erosion rates.

The erosion rate of the accel grid is approximately constant, decreasing only slightly over time in conjunction with the impingement current. Using this approximation, predictions of the operation time to 50 percent accel grid mass loss, and structural failure, can be made using only the initial erosion rates. Figure 3.46 compares the end of life predictions using a) the initial erosion rate extrapolations and b) the full lifetime simulations. The full lifetime simulations predict slightly greater propellant throughput capability as expected as a result of the slightly decreasing erosion rates over time.

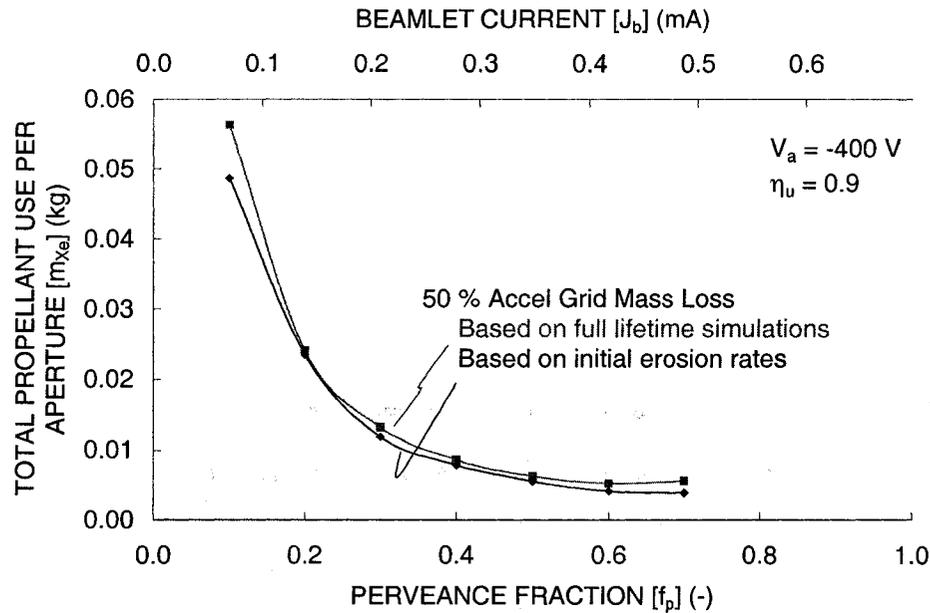


Figure 3.46 End of life predictions based on structural failure of the accel grid.

3.3.4 Pit and Groove Erosion Rates

The pit and groove depths can be found by looking at cross sections of the grids. Figure 3.47 shows four cross sections taken through the screen and accel grids. The pits and grooves are about 64 and 23 percent through the accel grid respectively in this particular case. The grids are likely to fail when the grooves wear completely through the accel grid, thereby dislodging the accel grid aperture and possibly causing a grid short for instance. For this case, the grooves will go through the accel grid after about 63,000 hours of operation.

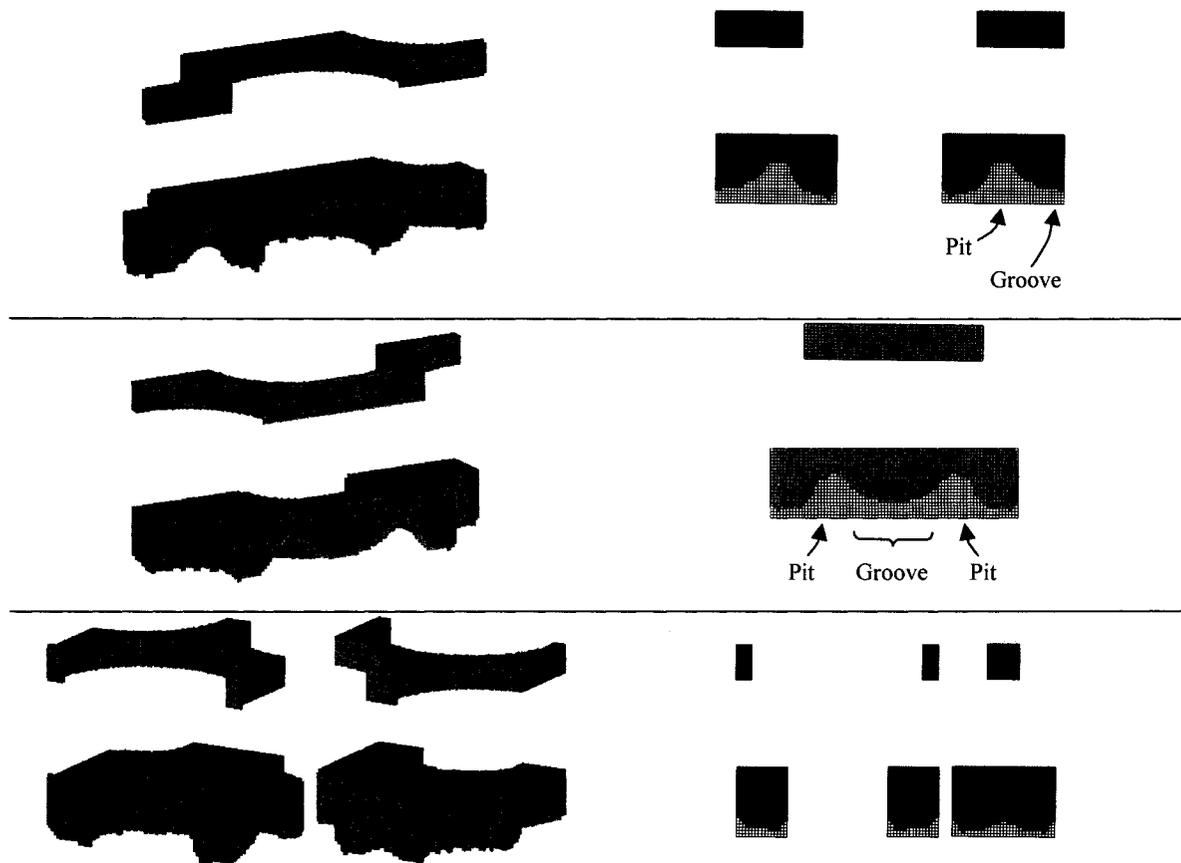


Figure 3.47 Cross sectional cuts through the screen and accel grids. $J_b = 0.139$ mA ($f_p = 0.2$), $T = 14,400$ hours (1.64 years).

In general, charge exchange ions created far downstream of the accel grid are vectored into the pits because the pits are the most negative locations on the downstream side of the accel grid. Alternatively, the charge exchange ions that originate just downstream of the accel grid are still within a highly non-zero electric field region, and many of these ions strike the accel grid in the groove regions before they reach the pits. For the charge exchange ions that strike the downstream side of the accel grid, Figure 3.48 shows the charge exchange ion originating locations colored red if those ions strike the pit regions and blue if those ions strike outside of the pit regions.

Downstream Side of the Accel Grid

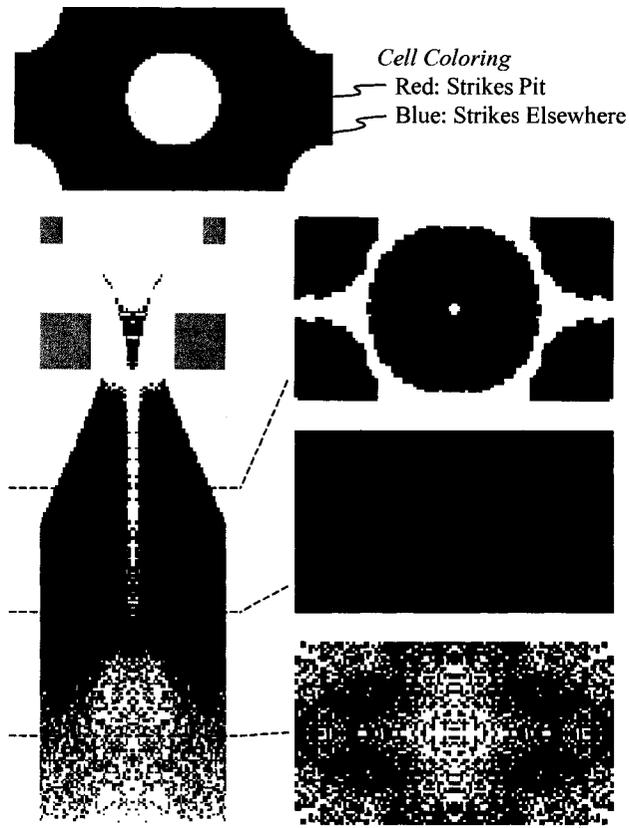


Figure 3.48 Charge exchange ions that strike the downstream side of the accel grid colored according to the pit or groove termination point.

The pit and groove erosion rates can be extrapolated to the points where the pits and grooves wear completely through the accel grid. The propellant throughput predictions at these points are compared in Figure 3.49 with the lifetime predictions using the 50 percent mass loss criteria. In this case, the 50 percent mass loss criteria is a conservative end of life predictor compared to the groove wear through predictions.

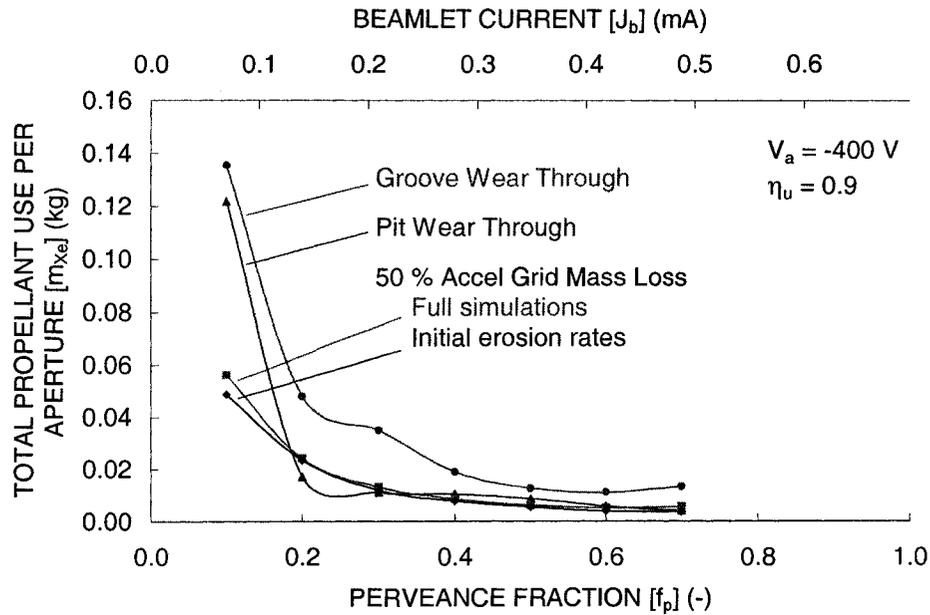


Figure 3.49 Lifetime predictions based on pit and groove depth erosion rates.

The pit and groove depth erosion rates, like the overall erosion rate, stay nearly constant over time. As the pits and grooves become deeper, more of the sputtered grid material gets re-deposited back onto the accel grid, causing the pit and groove erosion rates to decrease. An example of this is shown in Figure 3.50 for a perveance fraction of 0.2. Sputtered grid material can deposit onto the screen or accel grids, go through the screen grid and travel farther upstream, or travel downstream away from the grids.

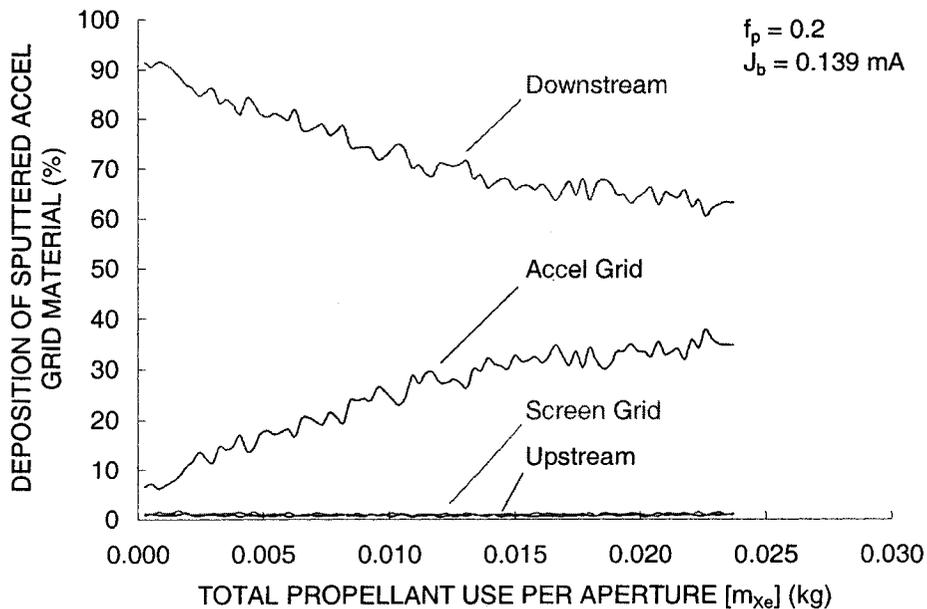


Figure 3.50 Deposition of sputtered grid material according to termination grid or surface.

Figure 3.51 shows the screen and accel grid surfaces for the beamlet operating at a perveance fraction of 0.2 colored according to mass loss rate, which results from sputtering, and mass gain rate, which is due to the re-deposition of sputtered grid material. At the beginning of life, all of the sputtered grid atoms on the downstream side of the accel grid travel downstream, away from the accel grid. At 21,600 hours of operation, a fraction of the sputtered grid material originating within the pits and grooves deposits back onto the accel grid.

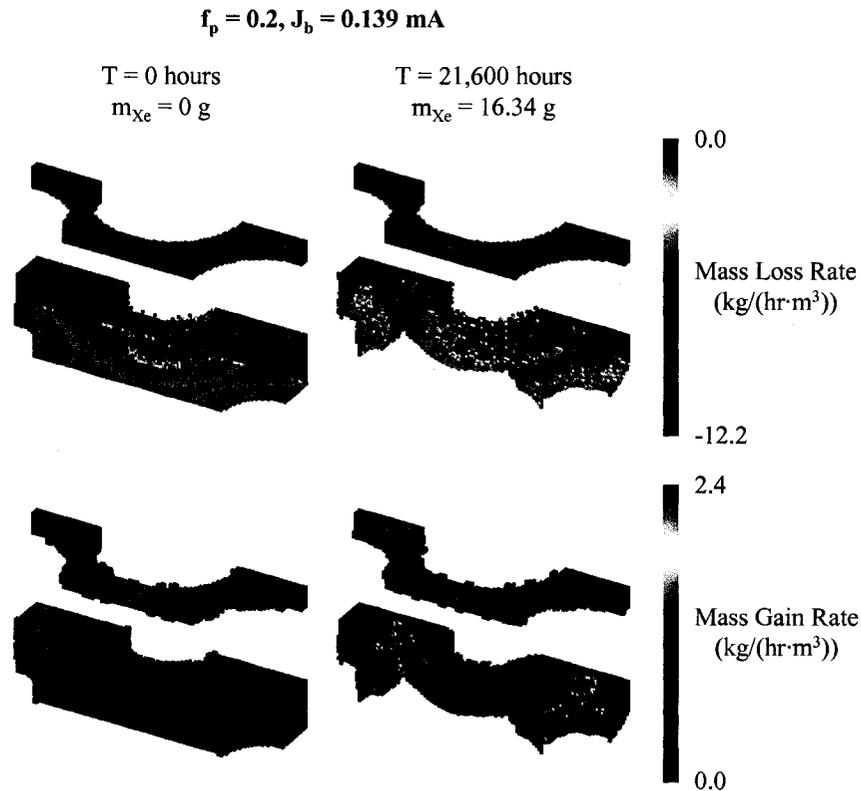


Figure 3.51 Mass loss and gain (re-deposition) rates.

The distribution of re-deposited material changes with beamlet current. Figure 3.52 shows the termination surfaces for the sputtered grid material as a function of perveance fraction. The values in this figure are averaged over the life of the grid at each perveance fraction. Deeper pits and grooves at the higher perveance fractions lead to more accel grid deposition. Also, a higher rate of hole barrel erosion leads to a higher percentage of sputtered material traveling upstream toward the screen grid or the screen grid hole.

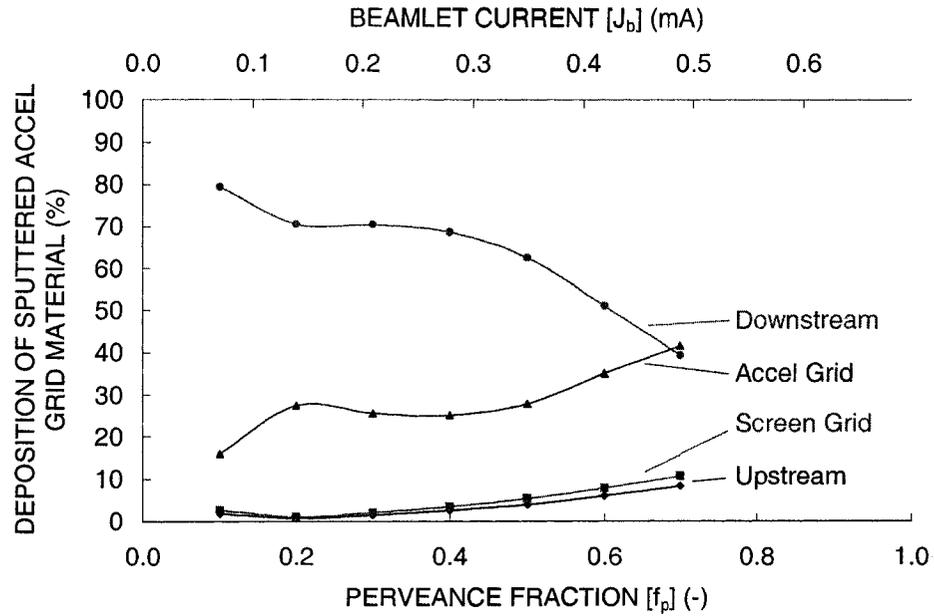


Figure 3.52 Deposition of sputtered grid material as a function of perveance fraction averaged over the life of the grid.

3.3.5 Screen Grid Film Thickness

The deposition of sputtered accel grid material onto the screen grid increases the screen grid mass over time. One way to characterize the amount of re-deposition is to calculate the thickness of the deposited material layer, called the film thickness.

The surface area of the downstream side of the screen grid is given in Eq. 3.58, and the surface area of the screen grid hole barrel is given in Eq. 3.59. The volume of deposited material will be taken as the film thickness multiplied by the sum of these two areas, as in Eq. 3.60. The mass of the deposited material, m , is the volume of the film multiplied by the density, ρ , of the film.

$$A_{downstream} = \frac{\sqrt{3}}{2} l_{cc}^2 - \frac{\pi}{4} d_s^2 \quad \text{Eq. 3.58}$$

$$A_{barrel} = \pi d_s t_s \quad \text{Eq. 3.59}$$

$$V = \frac{m}{\rho} = t_f (A_{downstream} + A_{barrel}) \quad \text{Eq. 3.60}$$

The screen grid film thickness build rate, in micro meters per gram of propellant ($\mu\text{m/g}$), is shown in Figure 3.53 as a function of perveance fraction. The screen grid film thickness at 50 % mass loss, which occurs at different times and amounts of propellant throughput for each perveance fraction, is also shown in

this figure. The amount of re-deposition is small, less than 6 % at most, compared to the grid spacing, which in this case is 1 mm.

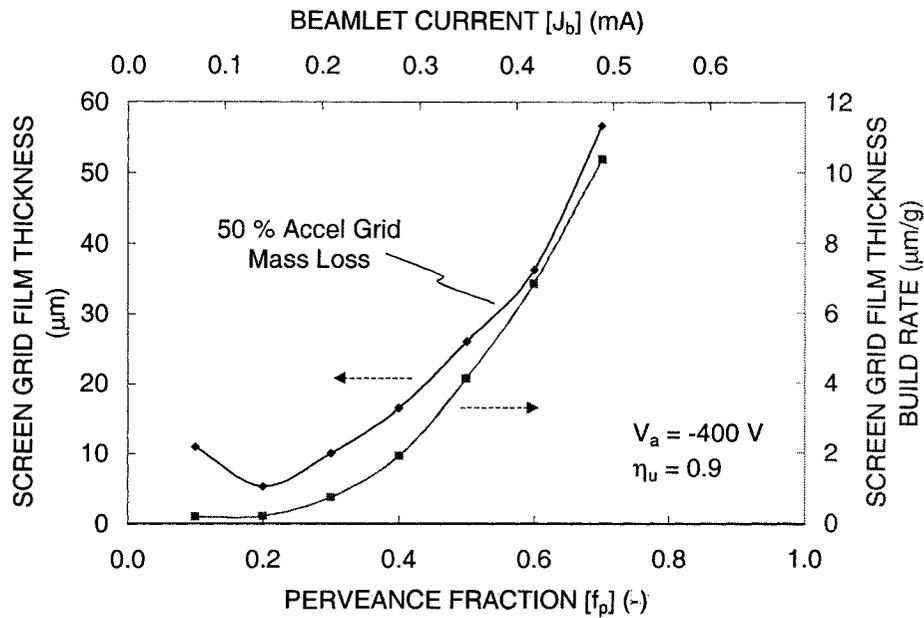


Figure 3.53 Screen grid film thickness as a function of perveance fraction.

3.4 Thruster Properties

A discussion of concepts related to the full thruster ensues. Topics include the number of apertures in the grid, total grid area, flatness parameter, and propellant utilization.

3.4.1 Number of Apertures

For a hexagonal aperture layout pattern, the area that a single aperture takes up in terms of the hole center-to-center spacing, l_{cc} , is given in Eq. 3.61. For a thruster face area A_{Total} , the number of apertures, H , is given as Eq. 3.62. The ion thruster active grid area, meaning the area with apertures, has been taken to be circular with a diameter d_B .

$$A_{Hex} = \frac{\sqrt{3}}{2} l_{cc}^2 \quad \text{Eq. 3.61}$$

$$H = \frac{A_{Total}}{A_{Aperture}} = \frac{\frac{\pi d_B^2}{4}}{\frac{\sqrt{3}}{2} l_{cc}^2} = \frac{\pi}{2\sqrt{3}} \frac{d_B^2}{l_{cc}^2} \quad \text{Eq. 3.62}$$

3.4.2 Beam Current

The total beam current is given the symbol J_B , while the beamlet current of any individual aperture is J_b . The average beamlet current is the total beam current divided by the number of apertures, as given in Eq. 3.63.

$$J_{b\ avg} = \frac{J_B}{H} \quad \text{Eq. 3.63}$$

The power going into the ion beam is P_B . It is related to the beam current and the net accelerating voltage through Eq. 3.64.

$$P_B = J_B V_N \quad \text{Eq. 3.64}$$

3.4.3 Grid Area

The beamlet current being extracted from a single hole is given by Eq. 3.65, where j is the current density of ions arriving at the aperture from the discharge chamber, $A_{Aperture}$ is the area that an aperture takes up on the thruster face, and ϕ is the actual grid transparency to ions. In the case of a hexagonal aperture layout, $A_{Aperture}$ is equal to A_{Hex} .

$$J_b = j \cdot A_{Aperture} \cdot \phi \quad \text{Eq. 3.65}$$

The average current density over the entire thruster face is the total beam current, J_B , divided by the total grid area, A_{Total} . The average beamlet current is this ratio multiplied the area of a single aperture, $A_{Aperture}$. This relation is shown in Eq. 3.66. The relation from Eq. 3.65 for the beamlet current can be substituted to find that the average current density over the thruster is equal to the current density of a single aperture multiplied by the transparency to ions. Additionally, the perveance equation can be substituted for the current density.

$$\begin{aligned} \frac{J_B}{A_{Total}} &= \frac{J_{b\ Average}}{A_{Aperture}} \\ &= j_{Average} \cdot \phi_{Average} \\ &= \left(f_{p\ Average} P_{Max} \frac{V_T^{3/2}}{l_e^2} \right) \cdot \phi_{Average} \end{aligned} \quad \text{Eq. 3.66}$$

This equation results in a useful expression for determining the total area of the grids from which ion extraction takes place, A_{Total} , based on the physical dimensions of the grids. Rearranging Eq. 3.66 for the total area gives Eq. 3.67. The average transparency to ions, $\phi_{Average}$, is more or less a constant regardless of the hole size, and is often slightly greater than the physical, area-based, transparency of the grids. This equation indicates that for the same average perveance fraction and ion transparency, decreasing the size of

the grids in terms of screen grid hole diameter (d_s) and grid spacing (l_g) will make the total area of the grids decrease.

$$\begin{aligned}
 A_{Total} &= \frac{J_B}{f_p \text{ Average } P_{Max} \frac{V_T^{3/2}}{l_e^2} \cdot \phi_{average}} \\
 &= \frac{J_B}{f_p \text{ Average } P_{Max} \frac{V_T^{3/2}}{\left[(l_g + t_s)^2 + \frac{d_s^2}{4} \right]} \cdot \phi_{average}}
 \end{aligned}
 \tag{Eq. 3.67}$$

Using the above expression for the total grid area, the total number of apertures in the grid is given in Eq. 3.68. Here, the individual aperture area has been put in terms of the physical open area fraction, ϕ_s . One result of this equation is that if the grid spacing, l_g , and the screen grid thickness, t_s , are scaled linearly with the screen grid hole diameter, d_s , while keeping everything else constant, the number of holes in the grid will stay the same even though the grid area is changing.

$$H = \frac{J_B \left[(l_g + t_s)^2 + \frac{d_s^2}{4} \right] \phi_s}{f_p \text{ Average } P_{Max} V_T^{3/2} \cdot \phi_{average} \frac{\pi}{4} d_s^2}
 \tag{Eq. 3.68}$$

3.4.4 Flatness Parameter

In a full ion thruster, there is a variation in the ion density throughout the discharge chamber. This results in a variation in the current density of ions at the grids. A plot of the current density measured downstream of the NEXT ion thruster is shown in Figure 3.54 [Soulas]. In using a hollow cathode as the source for primary electrons, the current density tends to be greatest on the thruster centerline where the highest rate of ionization takes place.

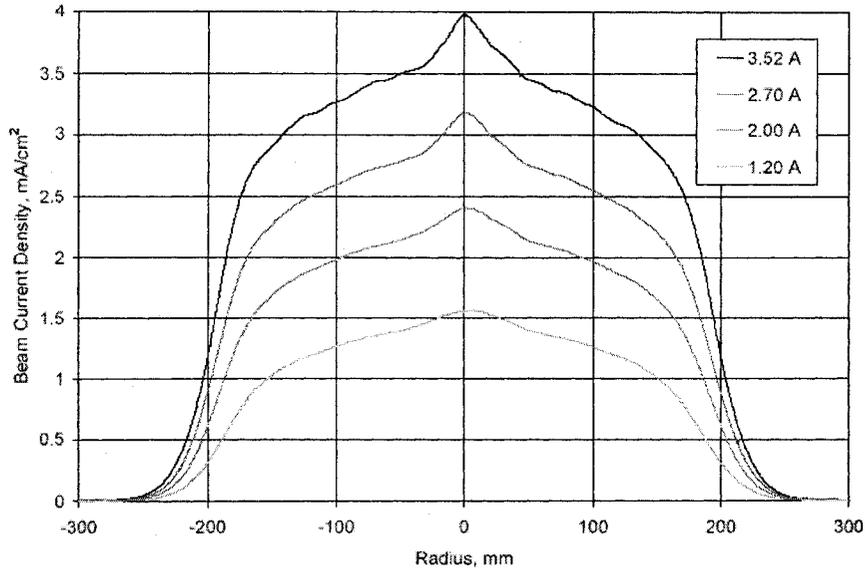


Figure 3.54 Current density variation downstream of an ion thruster.

The flatness parameter, F , is used to describe the ratio of the average current density over the entire thruster face, j_{avg} , to the maximum current density, j_{peak} . Additionally, this ratio is equal to the ratio of the average beamlet current, $J_{b avg}$, to the peak beamlet current, $J_{b peak}$, if the ion transparency is the same. The flatness parameter is found using Eq. 3.69, where the equation has partially been written in cylindrical coordinates for a circular ion thruster.

$$F = \frac{j_{avg}}{j_{peak}} = \frac{\int_0^{2\pi} \int_0^R j(r)r dr d\theta}{j_{peak} \int_0^{2\pi} \int_0^R r dr d\theta} = \frac{2\pi \int_0^R j(r)r dr}{j_{peak} \pi R^2} = \frac{1}{j_{peak}} \frac{J_B}{\pi R^2} \quad \text{Eq. 3.69}$$

$$F \approx \frac{J_{b avg}}{J_{b peak}} = \frac{J_B/H}{J_{b peak}}$$

3.4.5 Propellant Utilization Efficiency

Propellant is supplied mainly to the discharge chamber, but it can also be supplied to a neutralizer device, such as a hollow cathode, that expels electrons into the beam plasma at the same rate as ions leave the thruster in order to keep the spacecraft neutrally charged. The propellant flowing through the neutralizer is not used to produce thrust, thus it reduces the overall propellant utilization efficiency.

The overall propellant utilization efficiency is given in Eq. 3.70, where the propellant flow is comprised of the flow going through the discharge chamber, $\dot{m}_{discharge}$, and the flow going through the neutralizer, $\dot{m}_{neutralizer}$. When calculating the propellant utilization efficiency of the discharge chamber

only, the flow rate through the neutralizer is not included. The discharge chamber flow can be broken up into a “cathode” flow rate, which flows through the hollow cathode, and a “main” flow, which flows into the discharge chamber from elsewhere. The double to single current ratio, J_B^{++}/J_B^+ , is an average value for the entire ion beam.

$$\eta_u = \frac{1 + \frac{1}{2} \frac{J_B^{++}}{J_B^+}}{1 + \frac{J_B^{++}}{J_B^+}} \frac{J_B}{(\dot{m}_{\text{discharge}} + \dot{m}_{\text{neutralizer}}) \frac{e}{m_i}} \quad \text{Eq. 3.70}$$

It is thought that the unionized propellant that escapes through the grids does so evenly over the thruster face. This means that there is an equal number of neutral atoms leaving through the thruster centerline apertures, where the beamlet current can be relatively high, and through the edge apertures, where the beamlet current can be relatively low. As a result, the local propellant utilization efficiencies can be different across the grid face [Malone].

The discharge chamber propellant utilization efficiency, η_{ud} , is the ratio of atoms that go through the grids as ions to the total number of atoms. The flow rate of propellant atoms, $\dot{m}_{\text{neutrals}}$ in kg/s, that exit through the grids is given by Eq. 3.71, where the total flow rate is the discharge flow rate.

$$\dot{m}_{\text{neutrals}} = \dot{m}_{\text{discharge}} (1 - \eta_{ud}) \quad \text{Eq. 3.71}$$

After finding the neutral escape rate, one can then find the local propellant utilization efficiency, $\eta_{u \text{ local}}$, for an aperture operating at a beamlet current, J_b , and a local double to single current ratio, J_b^{++}/J_b^+ . This is given in Eq. 3.72, where the neutral escape rate per hole is simply $\dot{m}_{\text{neutrals}}/H$.

$$\eta_{u \text{ local}} = \frac{1}{1 + \frac{\frac{\dot{m}_{\text{neutrals}}}{H} \left(1 + \frac{J_b^{++}}{J_b^+} \right)}{J_b \frac{m_i}{e} \left(1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+} \right)}} \quad \text{Eq. 3.72}$$

The flow rate in a thruster is typically given in standard cubic centimeters per minute (sccm), which is a volume flow rate of propellant at a standard temperature and pressure. The flow rate can easily be converted to kilograms per second as shown in Eq. 3.73.

$$\dot{m} \{ \text{sccm} \} = \dot{m} \left\{ \frac{\text{cm}^3}{\text{min}} \right\} \quad \text{Eq. 3.73}$$

$$\dot{m} \left\{ \frac{\text{kg}}{\text{s}} \right\} = \dot{m} \left\{ \frac{\text{cm}^3}{\text{min}} \right\} \frac{1 \text{ min}}{60 \text{ s}} \frac{1 \text{ m}^3}{100^3 \text{ cm}^3} n_{\text{STP}} \left\{ \frac{1}{\text{m}^3} \right\} m_n \{ \text{kg} \}$$

Here, n_{STP} is the neutral density at standard pressure (1 atmosphere) and temperature (20 degrees Celsius), used to convert the volumetric flow rate to one that describes the actual number of atoms per

second. This neutral density is given in Eq. 3.74. For xenon, 1 sccm is about equal to 0.33 grams per hour, or about 0.067 Amps equivalent. Alternatively, 1 Amp equivalent of flow is equal to about 14.96 sccm.

$$n_{STP} = \frac{P}{kT} = \frac{101325 \left\{ \frac{N}{m^2} \right\}}{1.38 \cdot 10^{-23} \left\{ \frac{Nm}{K} \right\} 293.15 \{K\}} = 2.50 \cdot 10^{25} \{m^{-3}\} \quad \text{Eq. 3.74}$$

3.4.6 Lifetime

When evaluating the lifetime of a thruster, it is most useful to compare the grid erosion that takes place for different beamlet currents at equal times, taking into account the variation in the local propellant utilization efficiency.

As an example, the local propellant utilization efficiency was taken to be 0.9 at a perveance fraction of 0.4 (beamlet current 0.279 mA). From this, the neutral escape rate per hole was calculated, and new local propellant efficiencies of 0.818 and 0.931 were calculated for perveance fractions of 0.2 and 0.6, respectively. The accel grid material eroded away from apertures operating at these three perveance fractions is shown for two equal time steps in Figure 3.55.

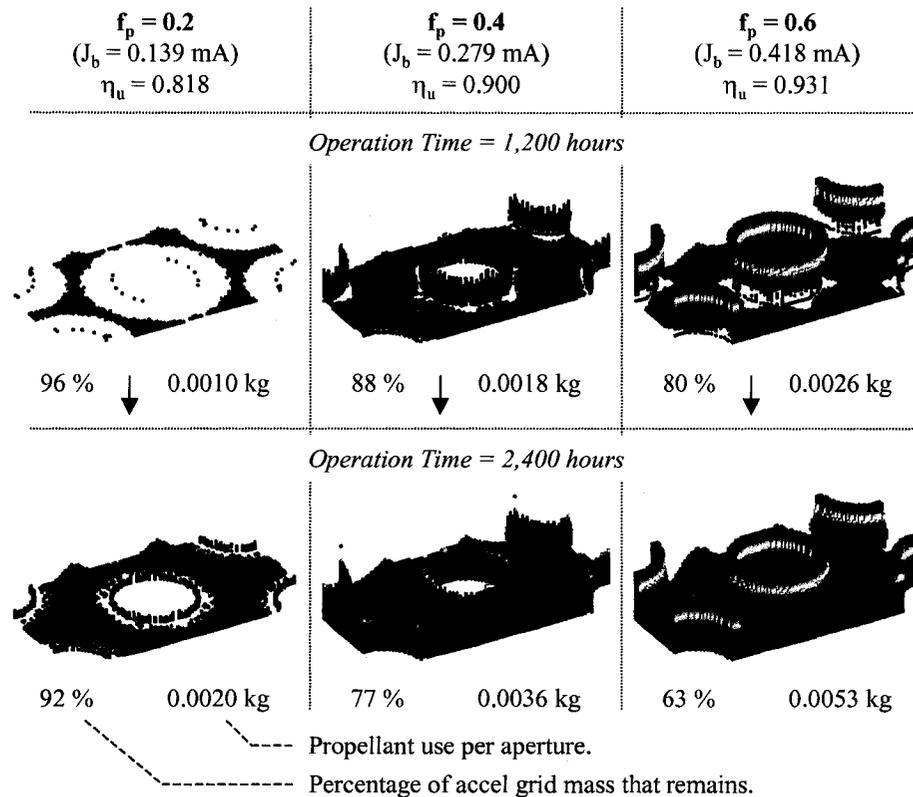


Figure 3.55 Erosion that would take place at equal times with varying propellant utilization efficiency on the face of an ion thruster.

Using the more realistic propellant utilization efficiencies increases the erosion at lower than average beamlet currents and decreases the erosion at higher than average beamlet currents when compared to using a single propellant utilization efficiency. Even with this effect, the erosion taking place at the higher beamlet currents is still life limiting. When using the same neutral flow rate at all beamlet currents, the neutral density is the same. As a result, the amount of charge exchange ion impingement current is proportional to the beamlet current only.

3.4.7 Perveance Limit

The perveance fraction equation is repeated below in Eq. 3.75. In looking at this equation, one way to find the perveance limit for a set net accelerating voltage, V_N , is to raise the beam current, J_B , until the accel grid impingement current, J_A , starts to rise. An alternative way to measure a perveance limit is to set the beam current and lower the net voltage until the accel grid current starts to rise. This second method is more commonly used in experimental testing.

$$f_p = \frac{\frac{J_B}{\frac{\pi}{4} d_B^2}}{P_{Max} \frac{l_e^2}{(V_N + |V_a|)^{3/2}}} \quad \text{Eq. 3.75}$$

4 Variable Effects

This chapter discusses the effects that geometric parameters and operating conditions have on grid performance. The knowledge gained through these investigations can then be used to improve grid designs to meet performance goals.

First, a simple procedure is presented by which the example grid set is created using roughly known operating conditions and existing grid designs. Subsequently, several parameters are varied to determine their effects on grid performance.

4.1 The Example Grid Set

An example grid set was used previously to introduce many aspects of grid operation. This section will explain how that grid set was constructed.

The example grid set was created with existing grid set designs already in mind. Figure 4.1 shows the geometrical relationships of some of these grid sets. The hole center-to-center spacing (l_{cc}), screen grid hole diameter (d_s), accel grid hole diameter (d_a), grid spacing (l_g), screen grid thickness (t_s), and accel grid thickness (t_a) have all been normalized to each grid's screen grid hole diameter, d_s .

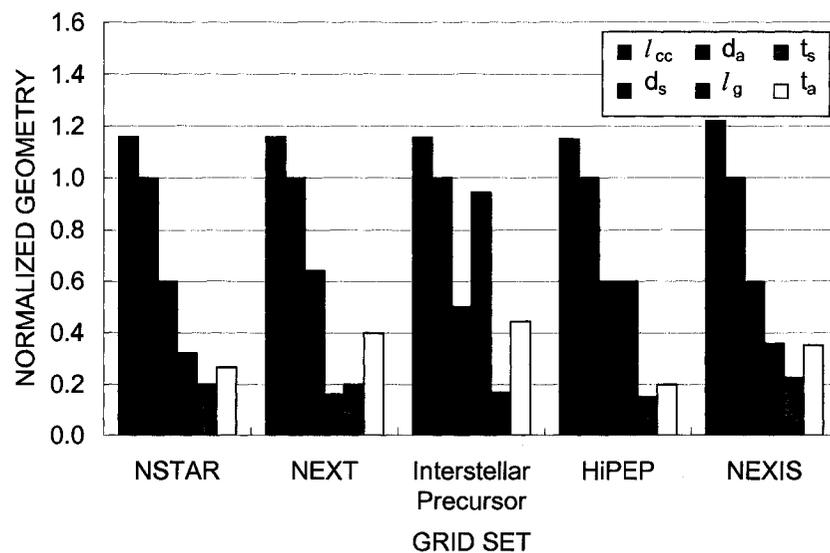


Figure 4.1 Geometrical relationships of several selected grid set designs.

These grid sets have been operated over a wide voltage range, from about 500 to 13,000 V. It is somewhat expected that all of these grid sets are fairly similar as they are probably all roughly related to the NSTAR grid set design.

Relationships between the geometrical parameters were chosen for the example grid set and are shown in Table 4.1. All values are normalized to the screen grid hole diameter, d_s . As a result, choosing a

value for d_s will set all other parameters. The ratio of the hole center-to-center spacing to the screen grid hole diameter, l_{cc}/d_s , was chosen to in turn give a physical screen grid open area fraction of 67 %.

Table 4.1 Geometrical parameter relationships in the example grid set.

Ratio	Value	Source of Value
l_{cc} / d_s	1.163	$\frac{l_{cc}}{d_s} = \sqrt{\frac{\pi}{2\sqrt{3}\phi_s}}$ $\phi_s = 0.67$
d_s / d_s	1	Identity
d_a / d_s	0.6	Typical
l_g / d_s	0.5	Typical
t_s / d_s	0.2	Typical
t_a / d_s	0.4	Typical

The first step in designing the example grid set was to choose a specific impulse, I_{sp} . Using xenon as the propellant, and taking the combined propellant utilization efficiency and thrust factor ($\eta_u \cdot ft$) to be roughly 0.85 as an estimate, the net accelerating voltage, V_N , is 2266 V for a specific impulse of 5000 seconds. This calculation is shown in Eq. 4.1.

$$I_{sp} = \eta_u \frac{ft}{g} \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 4.1}$$

$$5000 \{s\} = \frac{0.85}{9.81 \left\{ \frac{m}{s^2} \right\}} \sqrt{\frac{2 \cdot 1.6 \cdot 10^{-19} \left\{ \frac{J}{eV} \right\} \cdot 2266 \{eV\}}{131.3 \cdot 1.66 \cdot 10^{-27} \{kg\}}}$$

After choosing the net accelerating voltage, a value for R , the ratio of the net-to-total accelerating voltage V_N/V_T , was chosen to set the accel grid voltage. Typically, R values fall into the range from 0.85 to 0.90. To be conservative, a value of 0.85 for R was selected. This gives an accel grid voltage of -400 V as shown in Eq. 4.2.

$$R = \frac{V_N}{V_T} = \frac{V_N}{V_N + |V_a|}$$

$$\text{or } V_a = V_N \left(1 - \frac{1}{R} \right) \quad \text{Eq. 4.2}$$

$$-400 \{V\} = 2266 \{V\} \left(1 - \frac{1}{0.85} \right)$$

The next step was to select a value for the screen grid hole diameter, d_s , based on what the magnitude of the electric field between the grids was going to be. The electric field, E , is normally limited

to a certain maximum value to prevent excessive electrical arcing between the screen and accel grids. It is not easy to say what this value should be, but historically a value of about 2 kV/mm has been used.

In this case, the total accelerating voltage, V_T , is 2666 V. The discharge voltage, V_d , which is basically the difference between the net accelerating voltage and the screen grid voltage, $V_d = V_N - V_s$, is selected as 25 V. A screen grid hole diameter, d_s , of 2 mm gives a grid spacing, l_g , of 1 mm according to the selected ratio of l_g/d_s , 0.5. This sets the electric field between the grids, $E = (V_T - V_d)/l_g$, to be 2641 V/mm, which is an acceptable electric field. Choosing a screen grid hole diameter leads to values for the rest of the geometrical grid parameters, which are shown in Table 4.2.

Table 4.2 Geometry of the example grid set for a specific impulse of 5000 seconds.

Geometrical Parameter	Value
l_{cc}	2.327 mm
d_s	2 mm
d_a	1.2 mm
l_g	1 mm
t_s	0.4 mm
t_a	0.8 mm

4.2 Grid Design Goals

In order to design a grid set, it is important to first identify what the grid is being designed for. For ion thrusters, the most important design goal might be a long lifetime, or really a large propellant throughput capability. A different goal might be to extract a large amount of current per unit area.

In any case, it is also good to aim for a high grid transparency to ions in order to keep the discharge losses to a minimum. A low divergence angle, or high thrust factor, might be desirable if the beam divergence is important to the intended application of the ion source.

4.3 Geometry Effects

Grid Spacing (l_g). The minimum grid spacing is set according to the maximum allowable electric field between the screen and accel grids. Historically, a maximum electric field of 2 kV/mm has been used as a guideline. The grid spacing largely controls how much current per hole can be extracted.

Aperture Center-To-Center Spacing (l_{cc}). The center-to-center spacing is discussed relative to the screen grid hole diameter, d_s , in terms of the open area fraction, ϕ_s . Increasing the open area fraction increases the maximum amount of current that can be extracted from a certain size discharge chamber by increasing the grid transparency to ions. On the other hand, having a relatively large center-to-center spacing makes the grids stronger and also decreases the transparency to neutral atoms, which minimizes the

amount of propellant loss. The minimum center-to-center spacing can also be limited by sheath interactions among neighboring holes, as discussed in the screen grid thickness section.

Screen Grid Thickness (t_s). The screen grid thickness is chosen to control the plasma sheath that forms upstream of each hole pair and to make the screen grid as lightweight as possible. Decreasing the thickness of the screen grid generally increases the grid transparency to ions because the plasma sheath moves upstream, allowing more ions to enter the hole instead of striking the screen grid. However, if the screen grid is too thin, the plasma sheath can move upstream into the discharge chamber and hole interactions can take place. Having a combination high open area fraction (according to l_{cc} and d_s) and thin screen grid can cause the ion beamlet to become noncircular. Most importantly, this increases the crossover limit, which can cause problems at low perveance fractions where many ion thrusters are typically operated. On the other hand, if the screen grid is too thick the ion transparency is reduced unnecessarily.

Screen Grid Hole Diameter (d_s). The screen grid hole diameter, along with the grid spacing, controls how much current per hole can be extracted. It is scaled based on the grid spacing, which is in turn scaled according to the electric field strength. For a fixed grid spacing, having small holes can be troublesome if hole misalignment is present. Having larger holes relative to the grid spacing can lessen the problems that misalignment can cause.

Accel Grid Thickness (t_a). The accel grid dimensions in general affect the lifetime of the grid set and the ability to prevent electron backstreaming. Increasing the accel grid thickness a) increases the accel grid lifetime because there is more mass in the accel grid, b) helps to prevent electron backstreaming by making the potential along the beamlet centerline more negative, and c) reduces the transparency to neutral propellant molecules. However, increasing the accel grid thickness increases the accel grid weight, reduces the operating range of the grids in terms of the crossover and perveance limits, and magnifies problems with hole misalignment.

Accel Grid Hole Diameter (d_a). Decreasing the accel grid hole diameter helps mainly to prevent electron backstreaming by reducing the potential along the beamlet centerline. Having a small diameter also limits the loss of neutral atoms from the discharge chamber and increases the accel grid lifetime as there is more accel grid mass to be eroded away. A larger hole diameter, however, both decreases the crossover limit and increases the perveance limit, allowing a greater beamlet current operating range.

4.4 Parameter Effects

The following sections will expand the understanding of how each grid set parameter affects the performance of a grid set. The parameters that are discussed are the grid-to-grid spacing (l_g), accel grid hole diameter (d_a), accel grid thickness (t_a), screen grid thickness (t_s), and accel grid voltage (V_a).

4.4.1 Grid Spacing, l_g

As discussed previously, the minimum grid spacing is largely dictated by the maximum allowable electric field between the grids. According to the perveance equation, Eq. 4.3, the maximum attainable current density goes roughly as the square root of the grid spacing.

$$f_p = \frac{j}{P_{Max}} \frac{l_e^2}{V_T^{3/2}} = \frac{\frac{J_b}{\pi d_s^2}}{P_{Max}} \frac{(l_g + t_s)^2 + \frac{d_s^2}{4}}{V_T^{3/2}} \quad \text{Eq. 4.3}$$

The grid spacing is one of the harder parameters to control, and measure, in actual operation of large ion thrusters. This can be a problem because grid spacing has a large effect on perveance and crossover limits. Figure 4.2 shows impingement limit curves for four grid spacing values, the standard grid spacing being 1.0 mm.

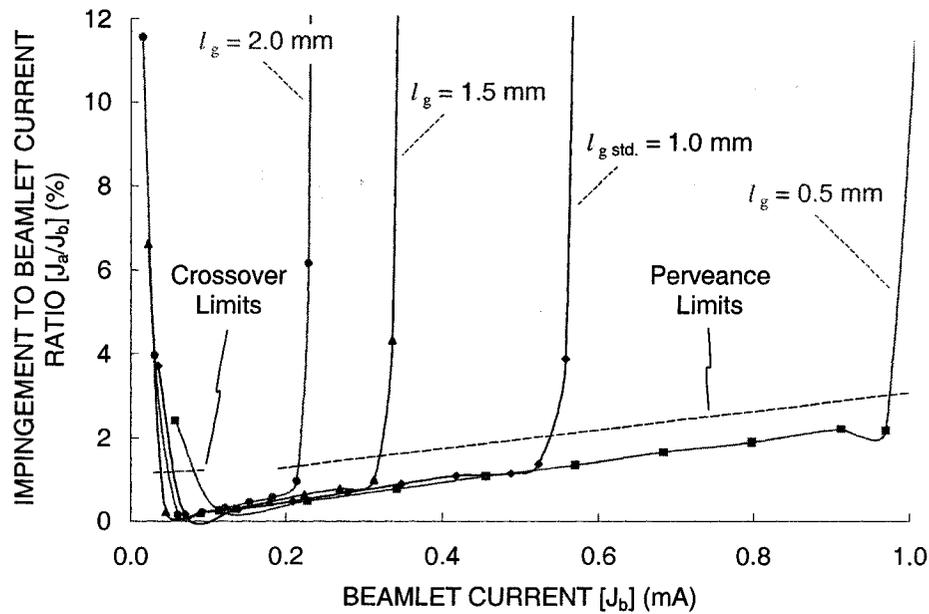


Figure 4.2 Crossover and perveance limits with changing grid spacing, l_g .

Figure 4.3 shows the crossover and perveance limits taken from Figure 4.2 above plotted as a function of effective grid spacing, l_e . instead of grid spacing, l_g . From the perveance fraction equation, the beamlet current, J_b , should vary as l_e^{-2} . This is with the idea that the perveance fraction, f_p , should be constant at the perveance limit for all grid spacings.

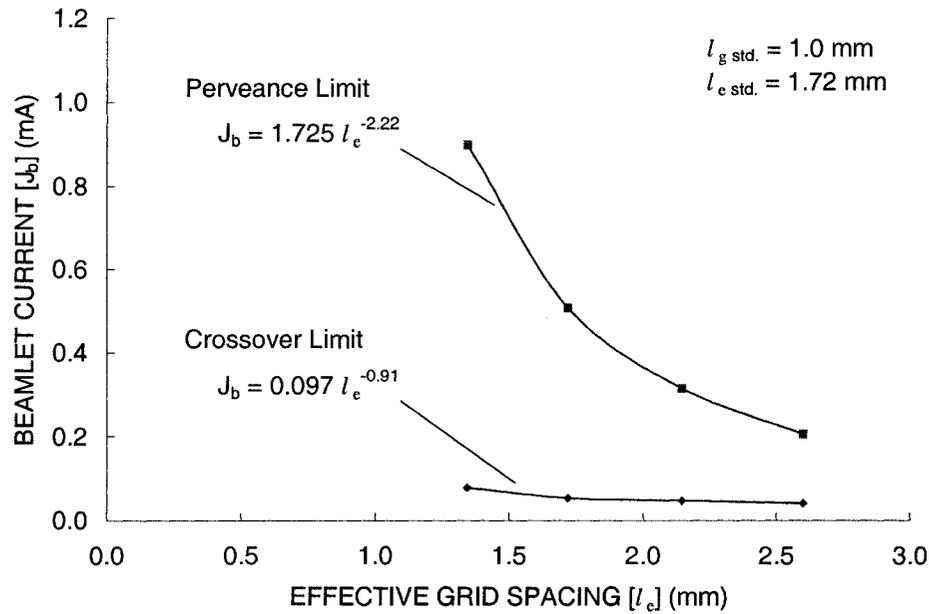


Figure 4.3 Crossover and perveance limits with respect to effective grid spacing, l_e .

The data in Figure 4.3 show that the perveance limit varies as $l_e^{-2.22}$ instead of l_e^{-2} . This is because f_p is not constant at the perveance limit. Figure 4.4 shows the crossover and perveance limits in terms of perveance fraction plotted against effective grid spacing. From this figure, one can see that the perveance limit increases and the crossover limit decreases as grid spacing is decreased. This shows that the operating range increases with decreasing grid spacing.

The perveance equation does not predict what the crossover limit will do with grid spacing. It turns out that the crossover limit beamlet current, $J_{b \text{ crossover}}$, increases with decreasing grid spacing. Many ion thrusters designed for long lifetime are operated at low beamlet currents. If the operating grid spacing is smaller than intended, crossover impingement is more likely to occur.

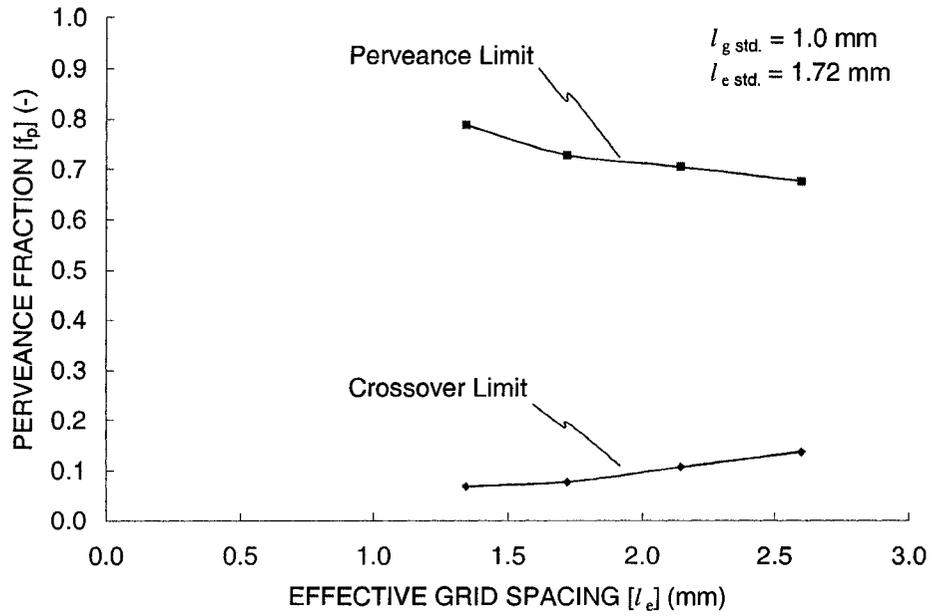


Figure 4.4 Crossover and perveance limits in terms of perveance fraction.

The grid transparency to ions gets better at smaller grid spacings, as shown in Figure 4.5. The physical open area fraction of the grids is 67 percent. Also note that this figure shows perveance fraction on the x-axis. When looking at a constant beamlet current and varying the grid spacing, the increase in ion transparency with decreasing grid spacing is even more apparent.

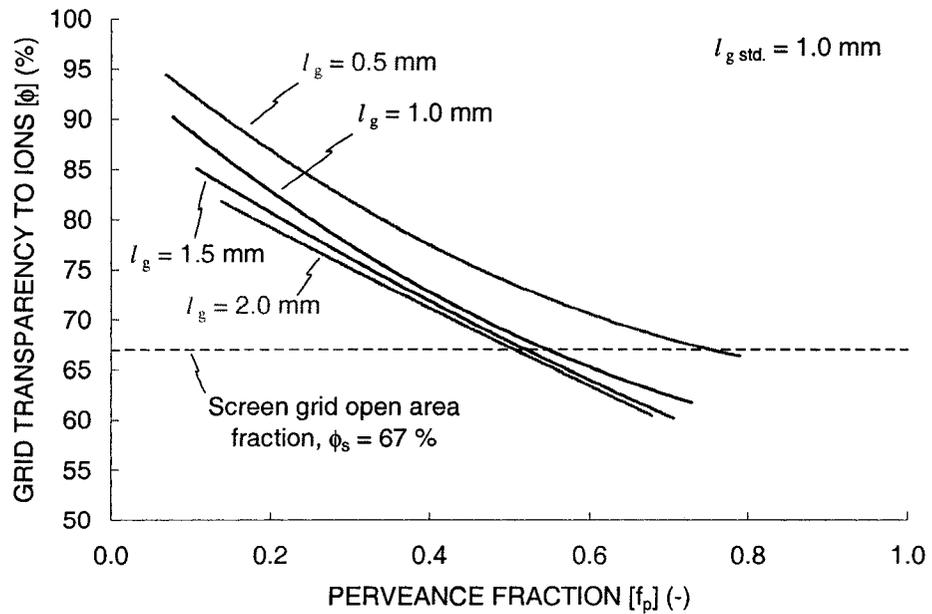


Figure 4.5 Transparency to ions for different grid spacings.

Grid spacing has a very important effect on the accel grid voltage needed to prevent electron backstreaming, as shown in Figure 4.6. This figure was generated using plots of the centerline saddle point potential, $V_{\text{SaddlePoint}}$, as a function of accel grid voltage at different perveance fractions (or beamlet currents).

At a perveance fraction of 0.4, decreasing the grid spacing from 1.0 mm to 0.5 mm increases the required accel grid voltage magnitude by about 57 V, or an increase of 40 %. This can increase the accel grid erosion rate significantly.

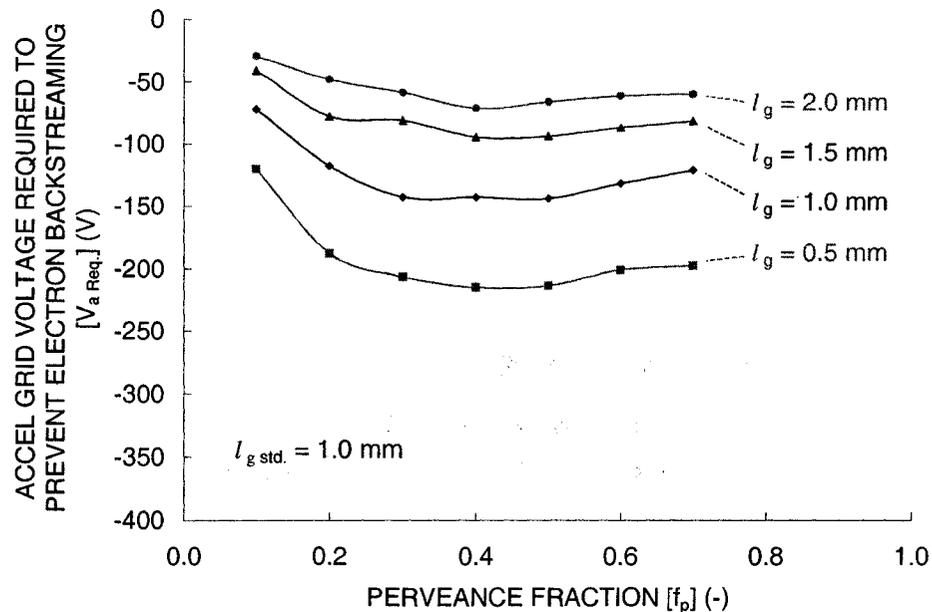


Figure 4.6 Resistance to electron backstreaming with variable grid spacing.

A certain voltage margin against electron backstreaming is often selected for actual use. In this case, at a perveance fraction of 0.4 for example, the required accel grid voltage is -143 V ($l_g = 1.0$ mm). Using a margin of 50 V, the accel grid voltage, V_a , becomes -193 V. Selecting this voltage minimizes the erosion rate while preventing electron backstreaming.

Figure 4.7 shows the predicted propellant throughput capability as a function of beamlet current for four grid spacings. Because the grid spacing is different in each case, the perveance fraction cannot be put on the second x-axis. This chart shows that if the beamlet current is held constant, greater propellant throughput is achieved if a smaller grid spacing is used. However, the data in this chart is all for a constant accel grid voltage of -400 V.

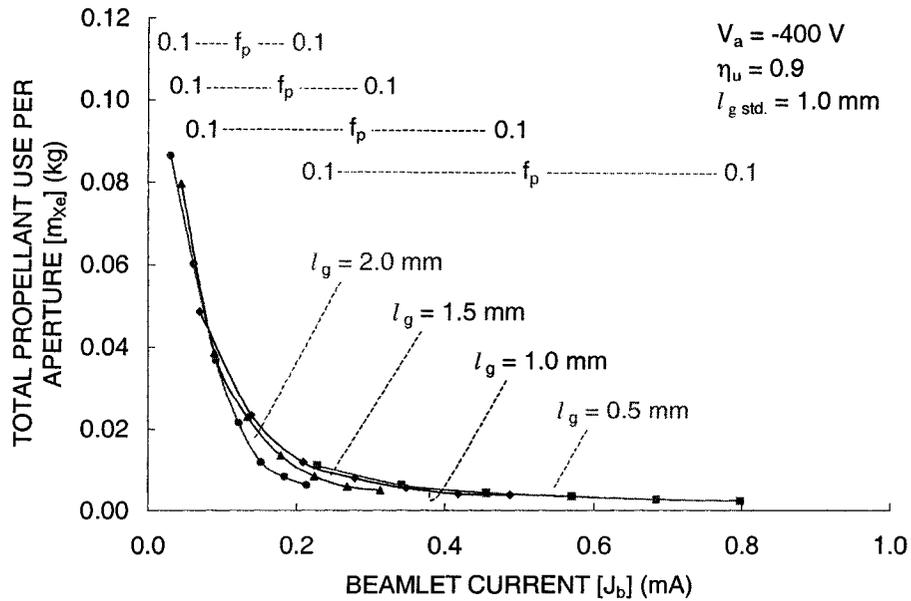


Figure 4.7 Propellant throughput capability as a function of beamlet current at several grid spacings.

The data in Figure 4.7 is plotted again in Figure 4.8 versus perveance fraction instead of beamlet current. This chart shows that if the perveance fraction is held constant, greater throughput is achieved at larger grid spacings. This is mainly because the beamlet current decreases with increasing spacing at a constant perveance fraction.

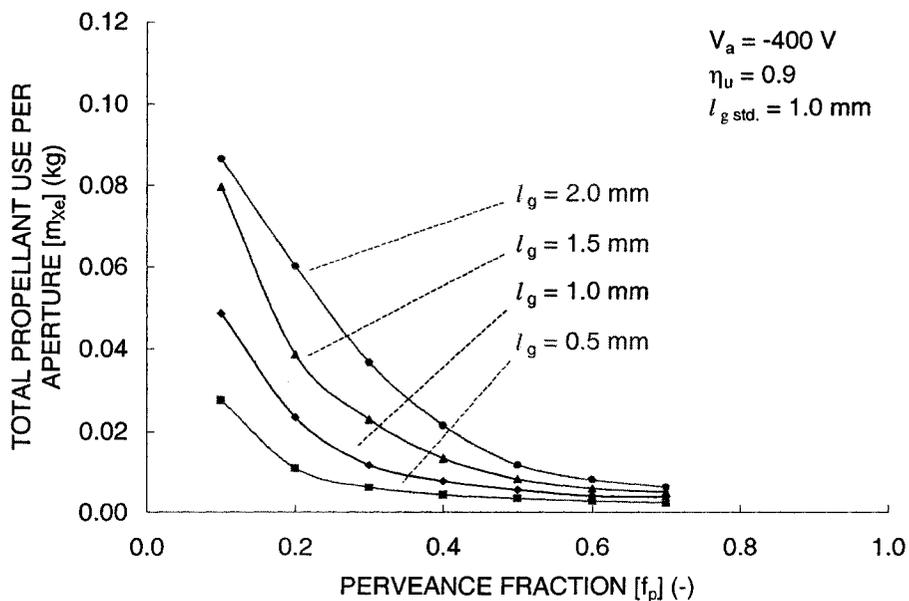


Figure 4.8 Propellant throughput capability as a function of perveance fraction at several grid spacings.

For a certain beamlet current (J_b), the grid spacing (l_g) and the accel grid voltage (V_a) can be varied together in order to maximize the propellant throughput capability. Figure 4.9 shows how this was done at a beamlet current of 0.2 mA. For each grid spacing from 0.5 to 1.75 mm, the accel grid voltage was adjusted to give a 50 V margin against electron backstreaming. The predicted propellant throughput values are based on the initial erosion rates.

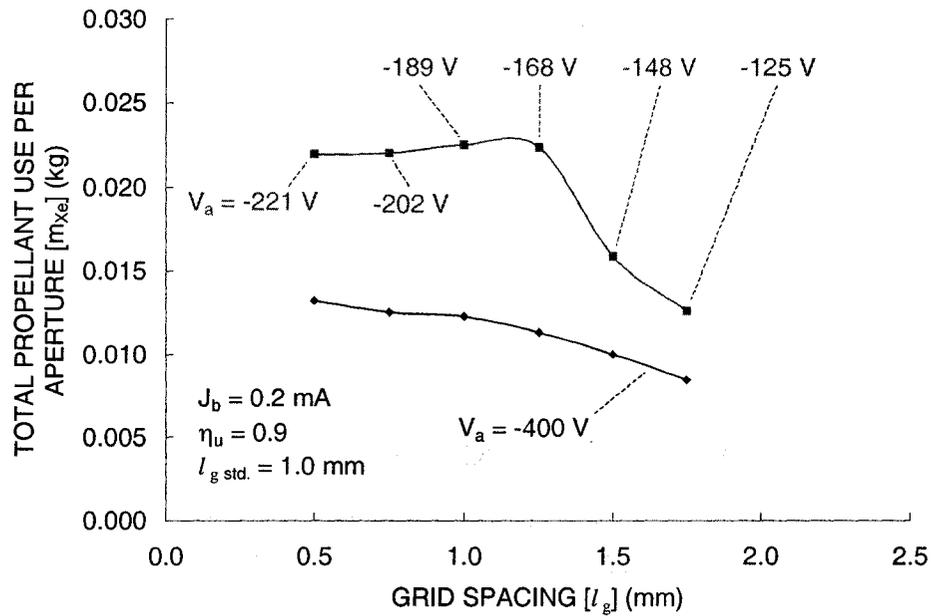


Figure 4.9 Variation in accel grid voltage and grid spacing to obtain a 50 V margin against electron backstreaming.

Following the same procedure for 0.1, 0.3, and 0.4 mA as was done for 0.2 mA in Figure 4.9, the data in Figure 4.10 are produced. At every data point in this figure, the accel grid voltage is set differently such that there is a 50 V margin against electron backstreaming.

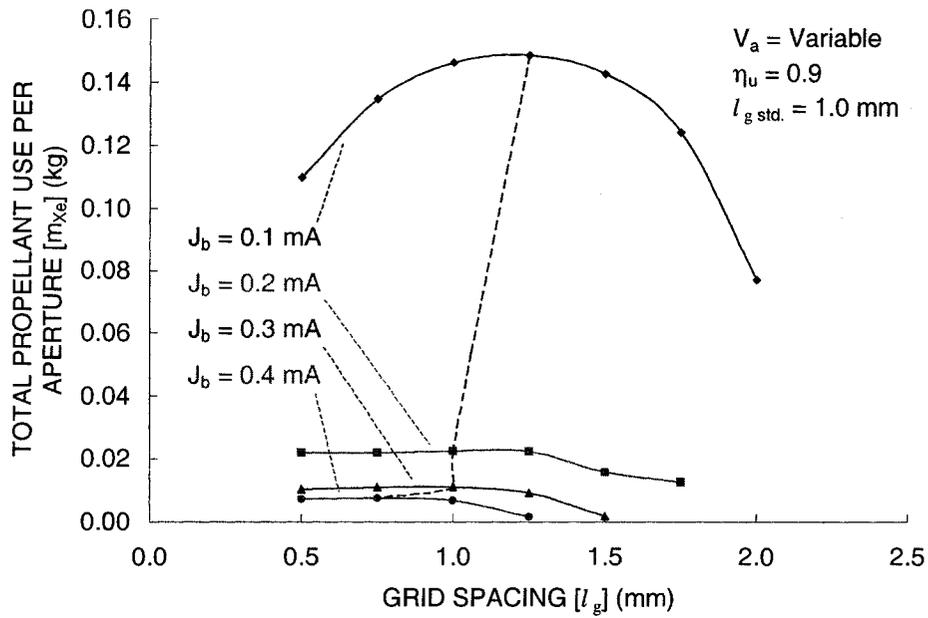


Figure 4.10 Finding the optimum grid spacing and accel grid voltage combination as a function of beamlet current.

In Figure 4.10, the optimum grid spacing occurs at a different grid spacing for each beamlet current. The optimum spacing values are plotted in Figure 4.11. Because the grid spacing changes with beamlet current, the perveance fraction is not a linear function of beamlet current. The optimum perveance fractions are shown along with the perveance fraction values if the grid spacing was held constant at 1.0 mm. Using the variable spacing makes the optimum perveance fraction stay at least a little more constant than when using a constant spacing.

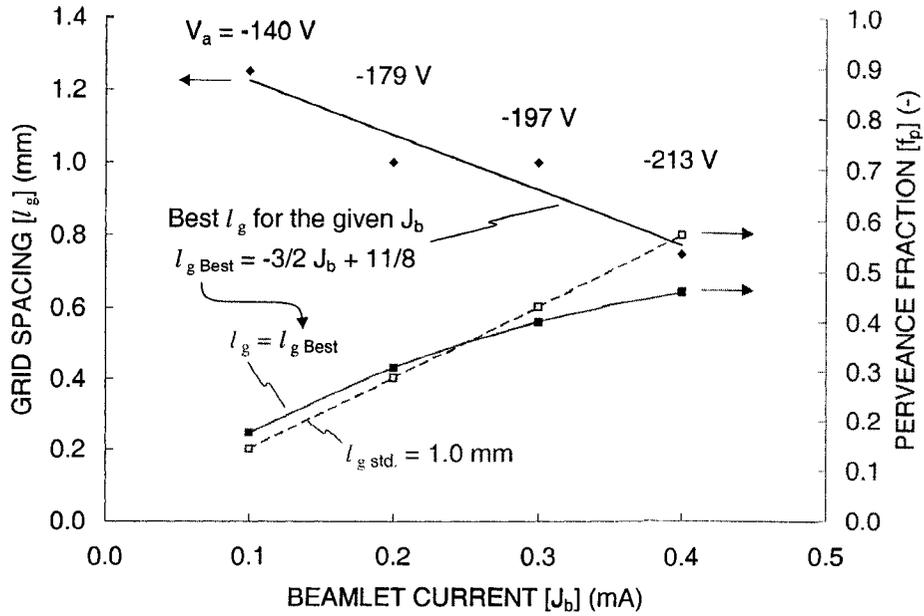


Figure 4.11 Optimum grid spacing and accel grid voltage as a function of beamlet current.

The point of this exercise was to increase the propellant throughput capability by varying the grid spacing and accel grid voltage, as shown in Figure 4.12. Compared to the standard accel grid voltage of -400 V, the propellant throughput capability is doubled or more when using the optimized grid spacing and accel grid voltage combinations.

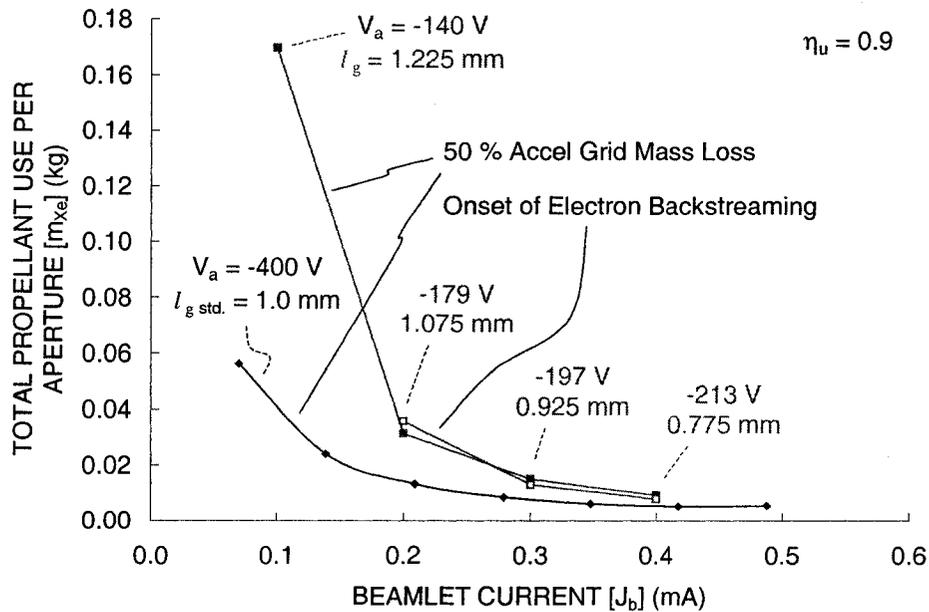


Figure 4.12 Propellant throughput capability using variable spacing and accel grid voltage.

4.4.2 Accel Grid Hole Diameter, d_a

A small accel grid hole diameter serves to help prevent electron backstreaming and to reduce neutral propellant atom loss from the discharge chamber. A large hole diameter, on the other hand, increases the available beamlet current operating range.

Figure 4.13 shows the crossover and perveance limits as a function of accel grid hole diameter. The reference hole diameter is 1.2 mm, 60 % of the 2.0 mm screen grid hole diameter. The crossover limit decreases, and the perveance limit increases, with increasing accel grid hole diameter, both of which are desirable results. The crossover and perveance limits were defined as the points where the impingement to beamlet current ratio, J_a/J_b , went above 1 %.

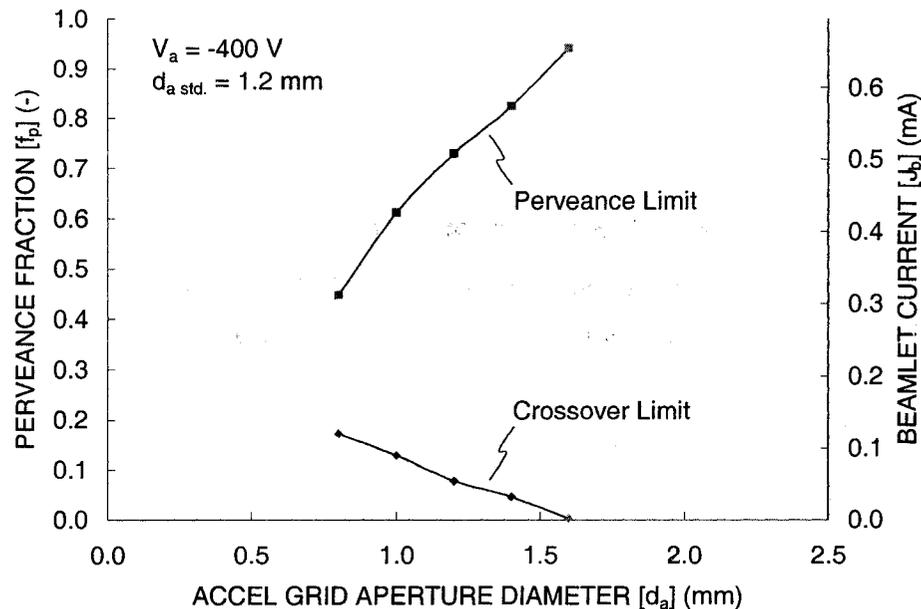


Figure 4.13 Crossover and perveance limits for different accel grid hole diameters.

The main drawback of increasing the accel grid hole diameter is that it allows electron backstreaming to occur more readily. Figure 4.14 shows the centerline saddle point potential between the crossover and perveance limits as a function of perveance fraction, and beamlet current, for the five hole diameters. Electron backstreaming occurs when the saddle point potential nears the beam plasma potential, which in this case is 0 V. At a perveance fraction of 0.2, a 10 percent change in the accel grid hole diameter results in a 16 V, or 6 percent, change in the saddle point potential.

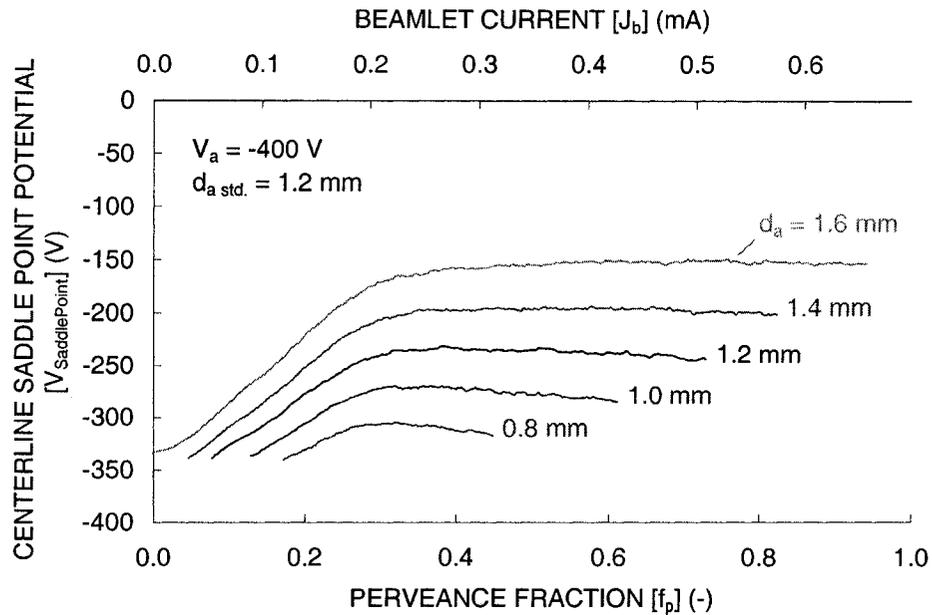


Figure 4.14 Resistance to electron backstreaming as a function of accel grid hole diameter.

Although it is not shown here, the divergence angle, α , of the beamlet gets slightly better (smaller) with increasing accel grid hole size. At a perveance fraction of 0.2, a 10 percent increase in the hole size results in about a 0.16 degree reduction in the 95 % divergence angle. Additionally, no noticeable change in the grid transparency to ions, ϕ , was seen in changing the accel grid hole diameter.

A small study was done to optimize the accel grid hole diameter while varying the accel grid voltage at the same time. Figure 4.15 shows the total mass of propellant per aperture obtained as a function of accel grid hole diameter at four beamlet currents. The accel grid voltage was adjusted at every point in order to obtain a 50 V margin against electron backstreaming.

When the accel grid hole diameter is reduced, the accel voltage does not have to be as negative, which reduces the erosion rate. Also, a smaller diameter increases the amount of mass that is available to be eroded away. However, the accel grid impingement current increases as the diameter is reduced because more charge exchange ions that are created in the hole barrel region, and in the region farther upstream between the grids, strike the accel grid instead of being accelerated downstream. This effect reduces the predicted accel grid lifetime. Another benefit of having a larger diameter hole is that, for a constant propellant utilization efficiency, the neutral density near the grids is reduced as it is easier for unionized propellant to flow through the grids. This reduces the charge exchange ion production rate and in turn reduces the accel grid impingement current and erosion rate.

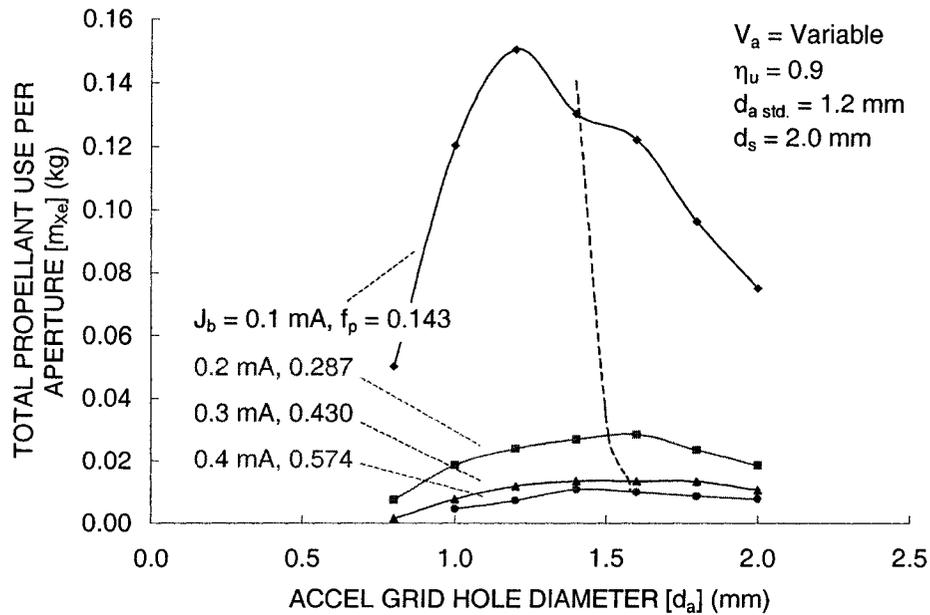


Figure 4.15 Optimizing the accel grid hole diameter for different beamlet currents.

Table 4.3 below lists the accel grid voltage and propellant throughput predictions at the four best points on the four curves in Figure 4.15. The optimum accel grid hole diameters were about 75 percent of the size of the screen grid hole diameter, 2.0 mm.

Table 4.3 Optimum accel grid hole diameter and accel grid voltage points.

f_p (-)	J_b (mA)	d_a (mm)	V_a (V)	m_{Xe} (kg)
0.143	0.1	1.2	-141	0.151
0.287	0.2	1.6	-244	0.028
0.430	0.3	1.4	-215	0.013
0.574	0.4	1.4	-206	0.011

The reason the optimum accel grid hole diameter increases in size with increasing beamlet current is that the beamlet diameter increases with beamlet current. Charge exchange ions created relatively far upstream, between the grids, cause a very high erosion rate when they strike the accel grid. With increasing beamlet diameter, more of these ions strike the accel grid upstream surface and hole barrel regions. To alleviate this problem, the optimum accel grid diameter increases so that more of these high energy charge exchange ions miss the accel grid, even at the expense of requiring a more negative accel grid.

The lifetime predictions in Figure 4.15 above were calculated using the initial accel grid erosion rates. Full lifetime simulations were performed for several accel grid hole diameters at a beamlet current of 0.2 mA to investigate these results. Figure 4.16 compares the propellant use to grid failure predicted using the initial erosion rates and full lifetime simulations. Additionally, using the full lifetime simulations, the propellant use at the onset of electron backstreaming can also be compared. For this case, electron

backstreaming occurs before structural failure at the smallest hole diameter. The best hole diameter for electron backstreaming is about 1.4 mm.

The best accel grid hole diameter in terms of 50 % accel grid mass loss is different when predicted using the initial erosion rates, 1.6 mm, compared to the full lifetime simulations, 1.2 mm. With a small hole diameter, charge exchange ions from the hole barrel region can strike the accel grid, causing the initial erosion rate to be relatively high. After some initial operation time, the accel grid hole diameter can erode away to the point where fewer ions strike the accel grid, reducing the erosion rate and ultimately resulting in a longer lifetime prediction.

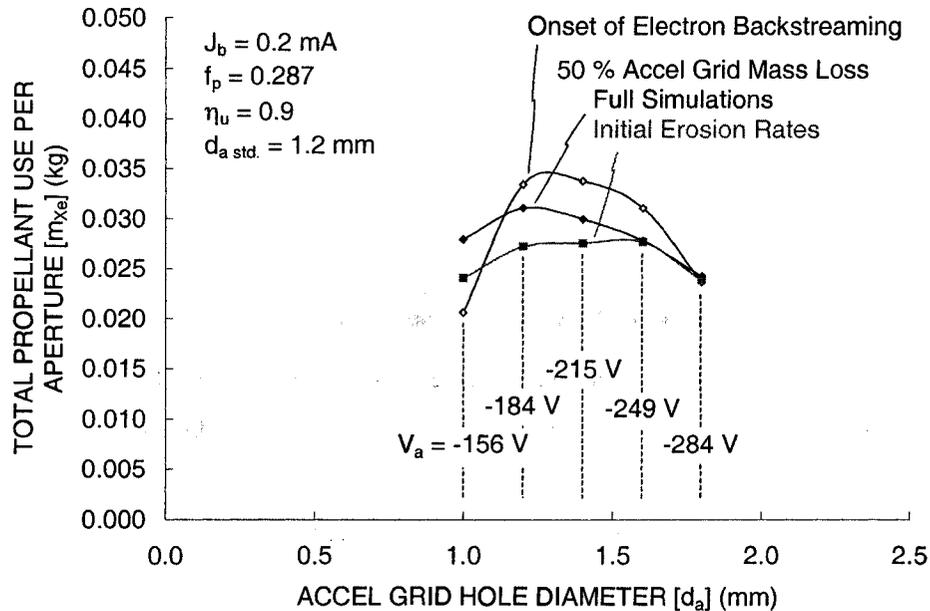


Figure 4.16 The optimum accel grid hole diameter found using full lifetime simulations.

4.4.3 Accel Grid Thickness, t_a

One of the main advantages of increasing the accel grid thickness, t_a , is that it makes the potential along the beamlet centerline more negative, increasing the resistance to electron backstreaming given a constant accel grid voltage. The other main advantage of increasing the accel grid thickness is that there is more mass in the accel grid available to be sputter eroded, which increases the lifetime of the grid. One drawback, however, is that the total mass, and therefore weight, of the accel grid is increased with greater accel grid thickness.

Another possible advantage of increasing the accel grid thickness is that the grid transparency to neutral propellant molecules is reduced, leading to higher neutral densities within the discharge chamber which can help electrical efficiency. However, this also increases the neutral density in the grid region,

which leads to greater charge exchange ion production, greater accel grid impingement current, and greater erosion rates which can offset the gain in grid mass.

The crossover and perveance limits are shown as a function of accel grid thickness, t_a , in Figure 4.17. In this case, the perveance limit does not change drastically with accel grid thickness. This is because direct impingement of ions onto the accel grid often begins on the upstream side of the accel grid surface, and is therefore independent of the grid thickness. The crossover limit, however, is a greater function of accel grid thickness. This is because crossover impingement starts on the downstream side of the accel grid hole barrel, and is therefore directly related to the length of the hole.

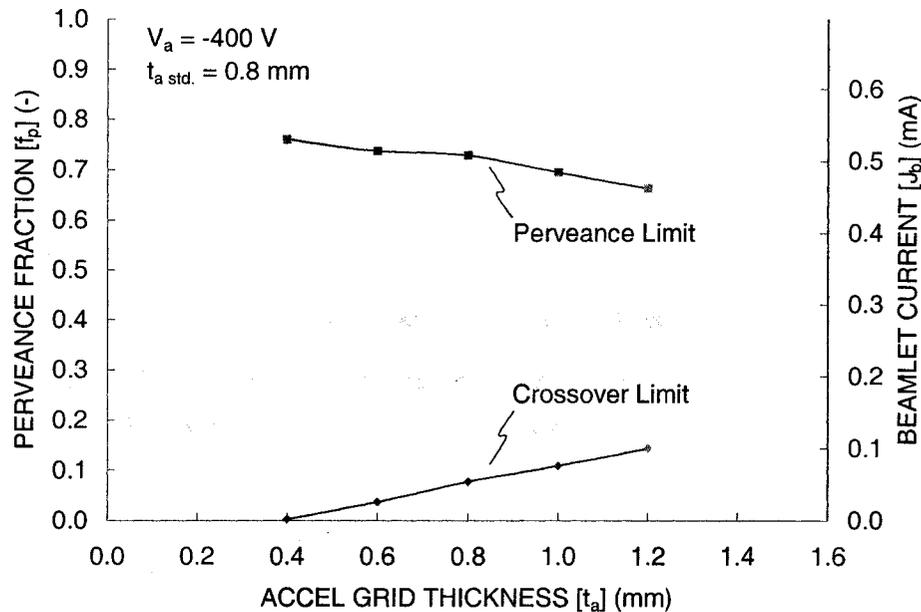


Figure 4.17 Available beamlet current operating range as a function of accel grid thickness.

Figure 4.18 shows two ion beamlets near the crossover and perveance limits. The beamlet shapes stay nearly the same for the two accel grid thickness values. The effect of the accel grid thickness on the perveance limit depends on the grid spacing, l_g . With a relatively close grid spacing, direct impingement would likely start on the upstream side of the accel grid, making the perveance limit independent of the accel grid thickness. With a large spacing, however, direct impingement might begin on the downstream side of the accel grid, and the accel grid thickness would affect the perveance limit. In this case, the grid spacing is between the two extremes, resulting in a modest dependence on the accel grid thickness.

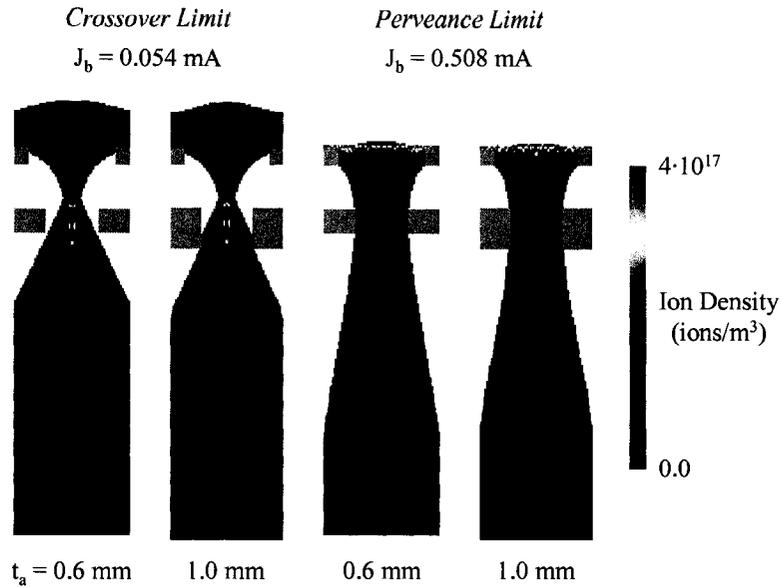


Figure 4.18 Effect of accel grid thickness on crossover and perveance limits.

Figure 4.19 shows how the centerline saddle point potential changes with both perveance fraction and accel grid thickness. As expected, a thicker accel grid leads to a greater resistance to electron backstreaming given a fixed accel grid voltage.

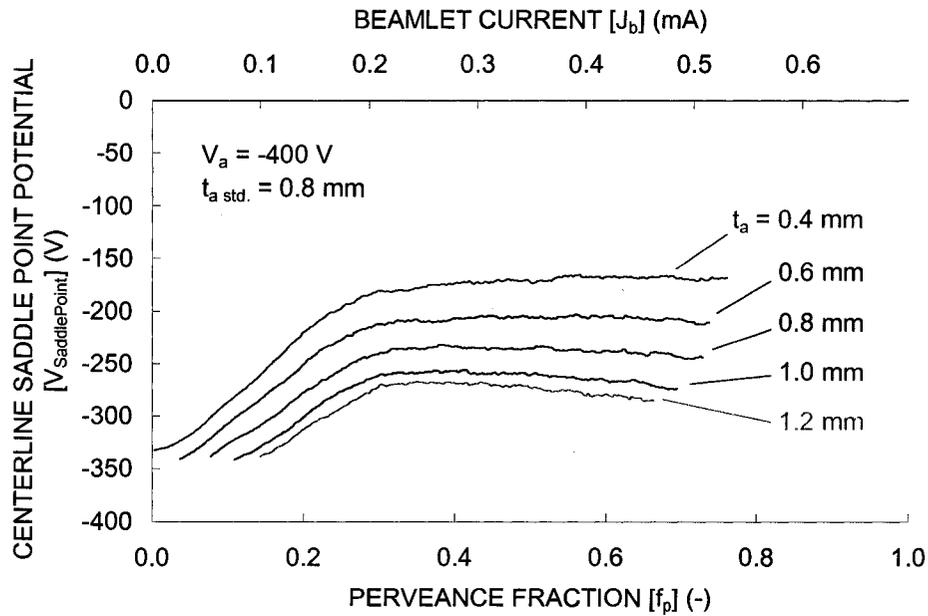


Figure 4.19 Resistance to electron backstreaming with variable accel grid thickness.

As with the accel grid hole diameter, d_a , the divergence angles of the beamlets were slightly better (smaller) with a thin accel grid than with a thick accel grid. Furthermore, no noticeable difference was seen in the grid transparency to ions, ϕ , with changing accel grid thickness.

Figure 4.20 shows how the predicted mass throughput capability changes as a function of accel grid thickness at four beamlet currents. At every point, the accel grid voltage was adjusted to achieve a 50 V margin against electron backstreaming. The propellant throughput capability increases with increasing accel grid thickness because the required accel grid voltage to prevent electron backstreaming decreases, resulting in a lower erosion rate, and because there is more accel grid mass available to be eroded away.

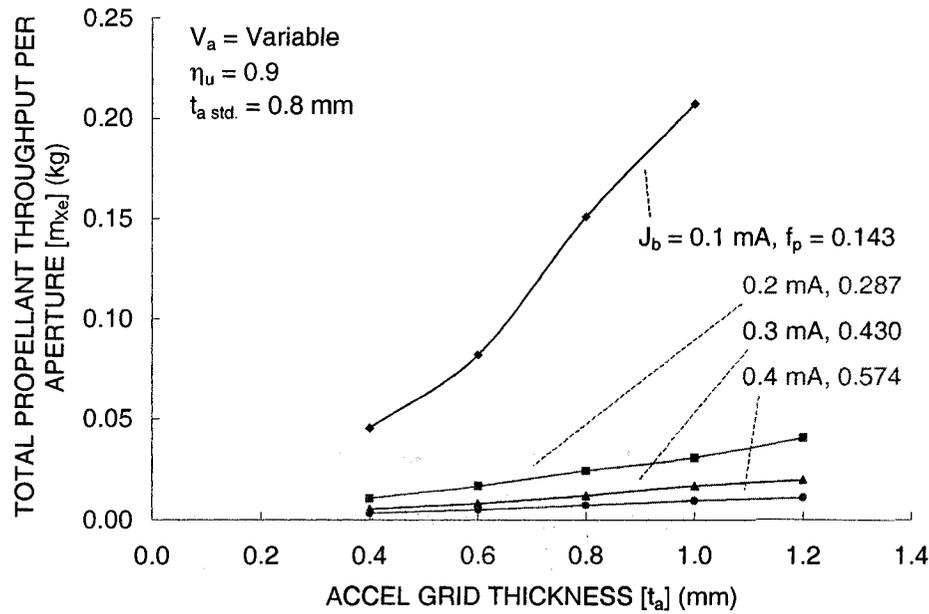


Figure 4.20 Predicted propellant throughput capability with variable accel grid thickness.

4.4.4 Screen Grid Thickness, t_s

The main advantage of reducing the accel grid thickness, t_s , is to increase the grid transparency to ions, ϕ . However, using a screen grid thickness that is too small can result in interactions among the sheaths of adjacent holes, which can cause the beamlets to become noncircular. Another advantage of reducing the accel grid thickness is a reduction of the screen grid mass.

Figure 4.21 shows how the crossover and perveance limits change as a function of screen grid thickness, t_s . Observing the perveance equation, Eq. 4.4, the screen grid thickness affects the effective grid spacing, l_e . The beamlet current, J_b , should vary as l_e^{-2} if the perveance fraction, f_p , stays constant at the perveance limit. Using the data points from 0.2 to 0.8 mm, the beamlet current at the perveance limit actually varies as $l_e^{-1.77}$. The perveance equation does not predict the variation in the crossover limit, where the beamlet current varied as $l_e^{-6.95}$.

$$f_p = \frac{j}{P_{Max}} \frac{l_e^2}{V_T^{3/2}} = \frac{\frac{J_b}{4} \frac{\pi d_s^2 (l_g + t_s)^2 + \frac{d_s^2}{4}}{P_{Max} V_T^{3/2}}} \quad \text{Eq. 4.4}$$

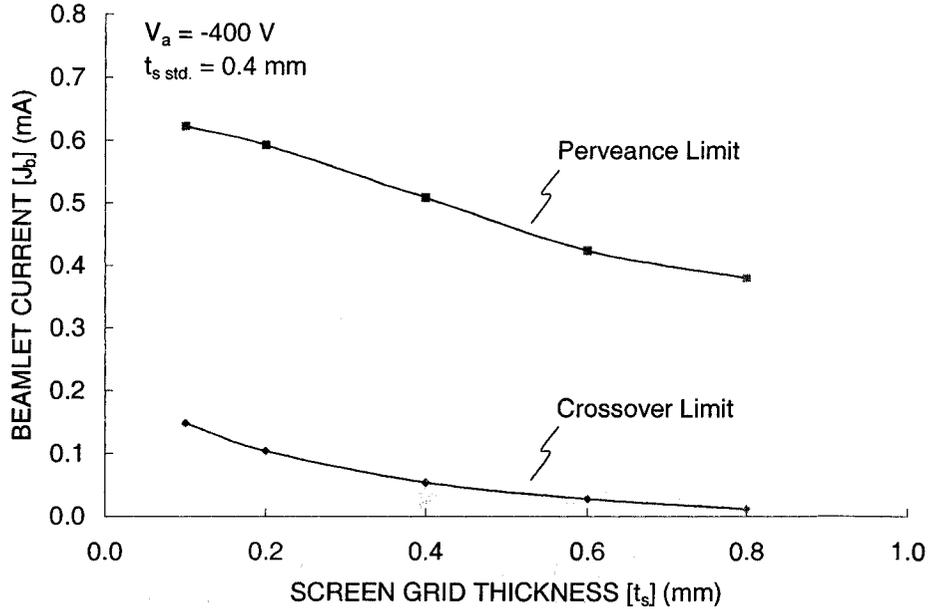


Figure 4.21 Shift in operating range as a function of screen grid thickness.

The perveance fraction, f_p , and the beamlet current, J_b , are not linearly related in terms of the screen grid thickness using the perveance fraction equation so a second x-axis cannot be used. The crossover and perveance limits are given in Figure 4.22 in terms of perveance fraction instead of beamlet current. If the perveance fraction equation perfectly modeled the perveance limit, the perveance fraction at the perveance limit should be a constant. Instead, the perveance fraction increases with increasing screen grid thickness. The perveance fraction at the crossover limit decreases with increasing screen grid thickness.

The effective acceleration length, l_e , is supposed to describe the distance over which the ions are accelerated through the total voltage, V_T . The actual acceleration distance is approximately the distance between the sheath, generally located upstream of the screen grid, and the most negative point along the beamlet centerline, or the location of the centerline saddle point potential. The reason the perveance fraction, f_p , is not constant at the perveance limit is because the sheath position changes when varying the screen grid thickness.

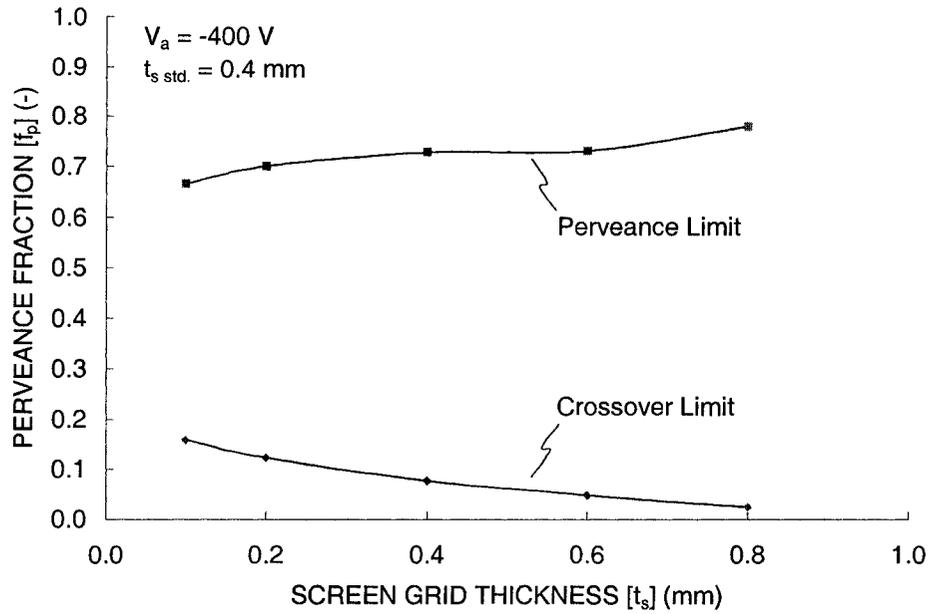


Figure 4.22 Non-constant perveance fraction at the perveance limit with variable screen grid thickness.

The centerline saddle point potential is shown in Figure 4.23 as a function of beamlet current for five screen grid thickness values. These data are presented in terms of beamlet current, J_b , but perveance fraction, f_p , could also have been used. Generally, the resistance to electron backstreaming increases with increasing screen grid thickness.

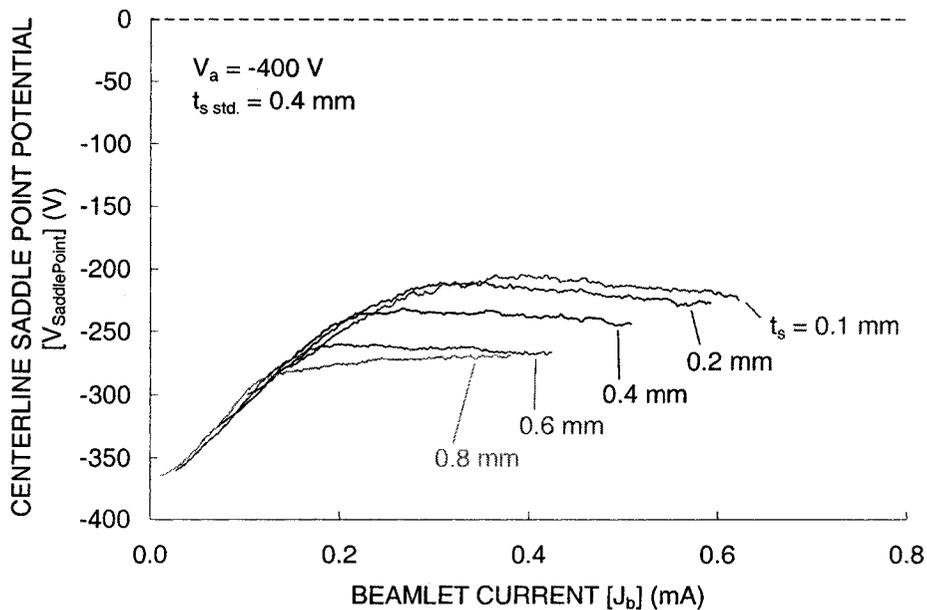


Figure 4.23 Resistance to electron backstreaming as a function of screen grid thickness.

The main reason for reducing the screen grid thickness is to increase the grid transparency to ions, ϕ . Figure 4.24 shows how the transparency to ions increases as screen grid thickness is reduced. The physical open area fraction of the screen grid in this case is 67 %. At a beamlet current of 0.3 mA, for example, the transparency to ions can be increased from 40 to 89 % in changing the screen grid thickness from 0.8 to 0.1 mm.

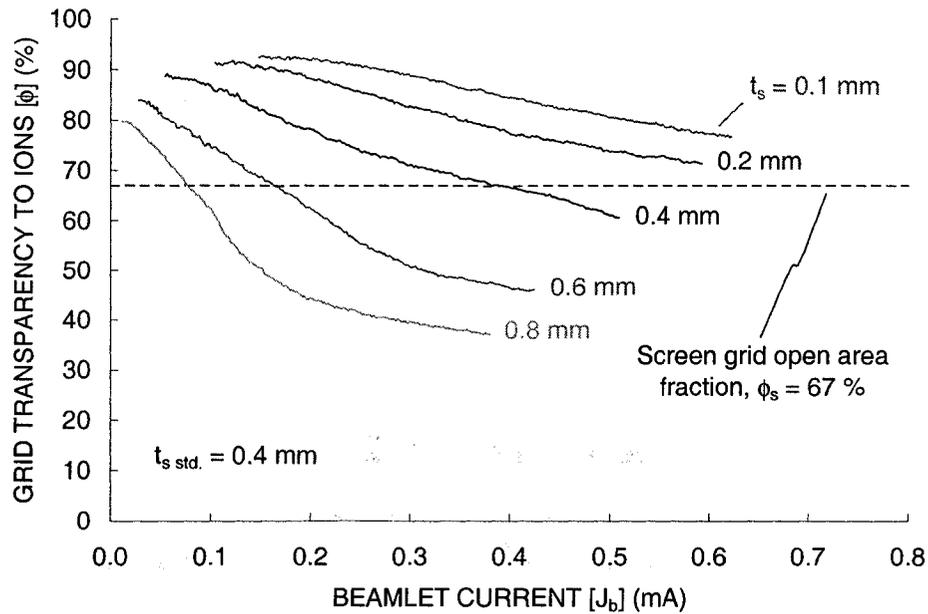


Figure 4.24 Grid transparency increase with decreasing screen grid thickness.

The screen grid thickness, t_s , affects the position of the sheath upstream of the beamlet, which in turn results in the drastic variation in ion transparency. Figure 4.25 shows how the sheath position changes as a function of screen grid thickness at four beamlet currents. Only the accel grid is shown along with the sheath in each case. Looking across a row, where the screen grid thickness is constant, the sheath moves downstream with increasing beamlet current. Looking down a column, where the beamlet current is constant, the sheath again moves downstream with increasing screen grid thickness.

When the sheath is far upstream of the screen grid, as it is at low beamlet currents with thin screen grids, a large percentage of ions entering the sheath are vectored into the beamlet, missing the screen grid. This results in a low screen grid current and therefore a high grid transparency to ions. When the sheath is far downstream, at high beamlet currents with thick screen grids, many ions from the discharge chamber strike the screen grid immediately after passing through the sheath. This results in a high screen grid current and a low ion transparency.

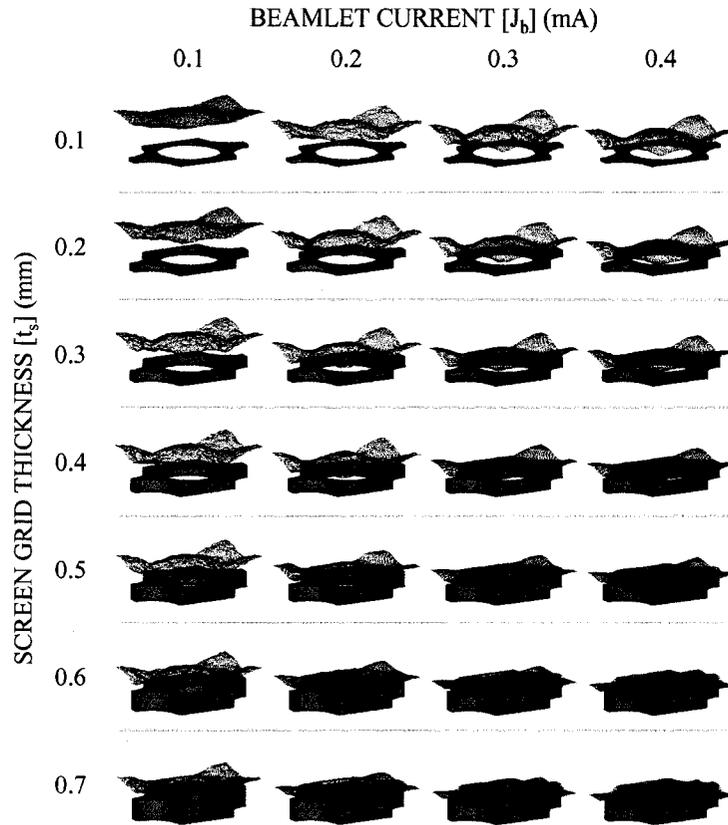


Figure 4.25 Sheath position as a function of screen grid thickness and beamlet current.

Figure 4.26 shows ion beamlet cross sections as a function of screen grid thickness at four beamlet currents. These cross sections are taken through a plane at the downstream edge of the accel grid. At a constant beamlet current of 0.4 mA, for example, the beamlet diameter increases as the screen grid thickness increases.

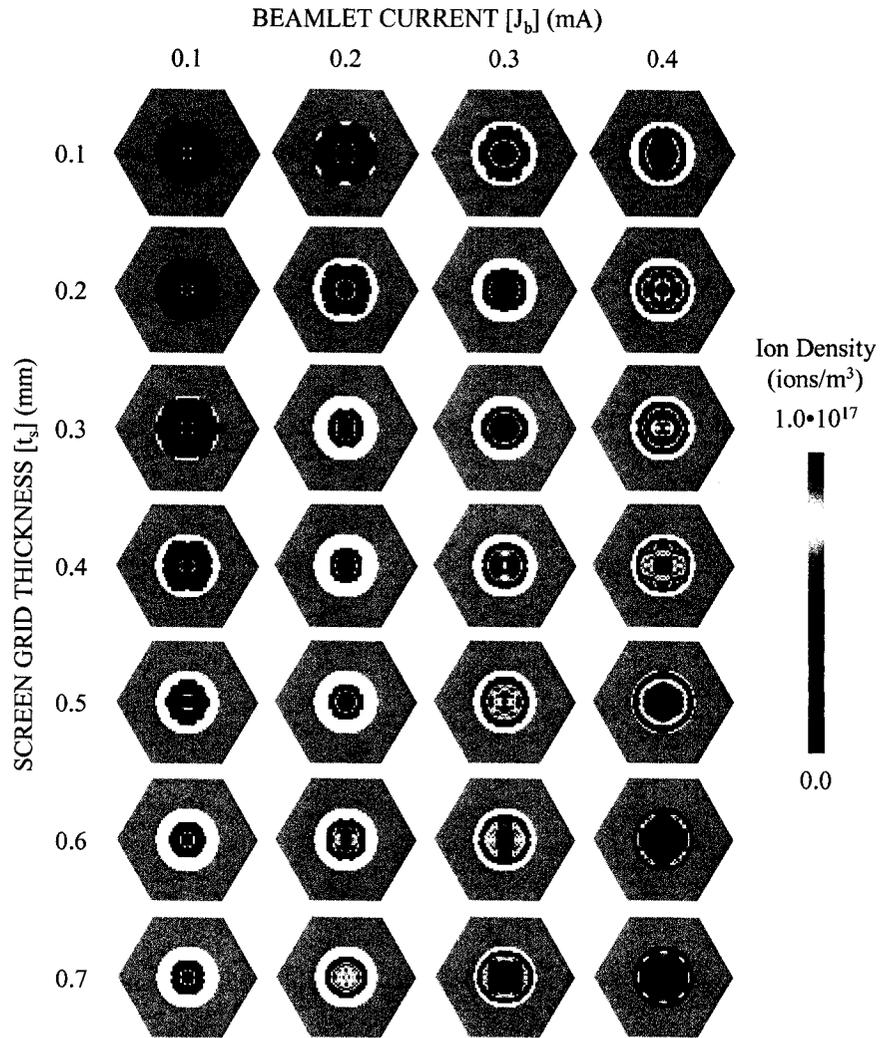


Figure 4.26 Beamlet cross sections, taken through a plane at the downstream edge of the accel grid, as a function of screen grid thickness and beamlet current.

When the sheath is located upstream, interactions between the sheaths of adjacent holes can take place. Figure 4.27 shows beamlet cross sections taken in the low beamlet current, thin screen grid range. The ion density scale in this figure has been reduced to show the lower density structures in the cross sections. As the screen grid thickness and beamlet current are reduced, the sheath interactions become greater and a six sided star pattern emerges in the cross section.

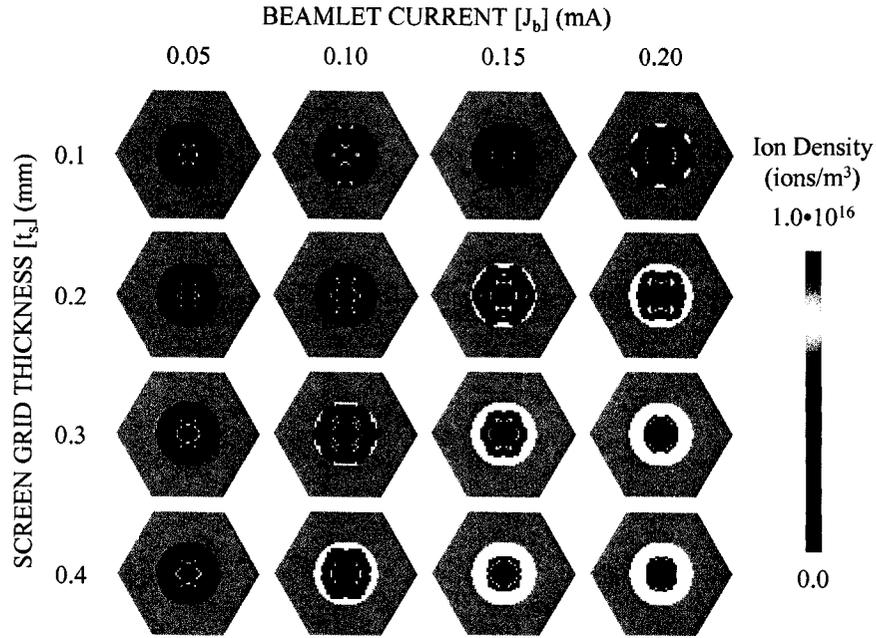


Figure 4.27 Beamlet cross sections at relatively low beamlet currents and relatively thin screen grids.

When the sheath moves upstream, away from the screen grid, the sheath is closest to the screen grid at the intersection points between any three holes in the hexagonal aperture pattern. Ions from the discharge chamber plasma passing through the sheath near these points experience a great change in radial velocity. When these ions reach the accel grid, they create the six points of the star pattern. The star pattern is expected to be the most well defined when the sheath is heavily dished, as it is at a beamlet current of 0.15 mA and a screen grid thickness of 0.1 mm for example. The star pattern, or a noncircular beamlet in general, that comes about when using a relatively thin screen grid increases the crossover limit.

Increasing the screen grid thickness tends to make the beamlet more well focused as the sheath moves closer to the screen grid and becomes flatter. Figure 4.28 shows how the divergence angle, α , of the beamlet changes as a function of beamlet current for five screen grid thickness values. At the lower beamlet currents in this case, the sheath is located upstream of the screen grid and moves a great deal with changing screen grid thickness. A line connecting a constant perveance fraction of 0.5 is drawn in this figure. Past this perveance fraction, the sheaths at all screen grid thickness values are located near the screen grid and no interactions take place among neighboring holes. As a result, the divergence angles of the beamlets become the same as the sheaths become less dependent on the screen grid thickness.

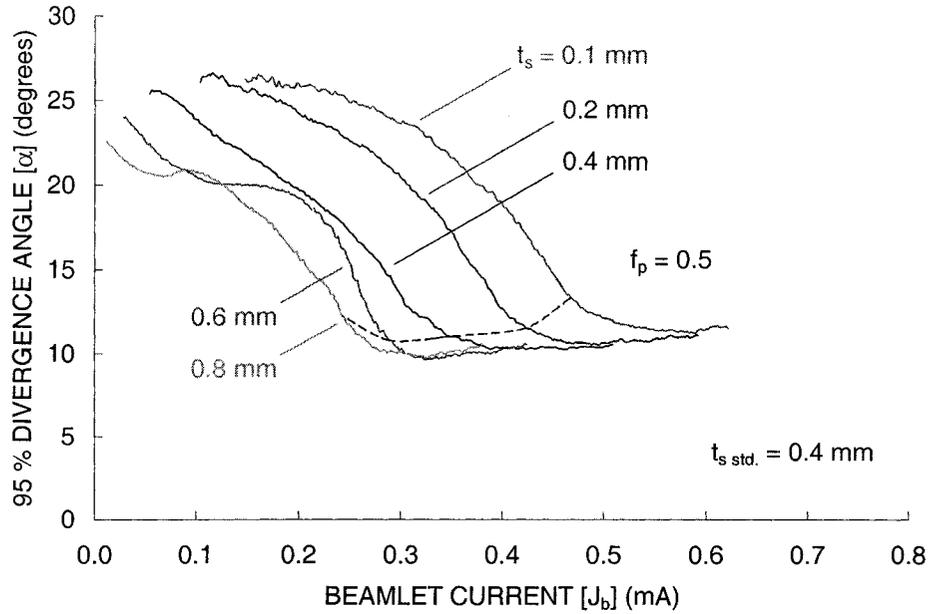


Figure 4.28 Divergence angle reduction with increasing screen grid thickness as a result of the sheath moving closer to the screen grid.

As with the other grid parameters, a small study was done to see if the predicted propellant throughput capability could be optimized at a constant beamlet current, J_b , by allowing the screen grid thickness, t_s , to vary. Figure 4.29 shows how the predicted propellant throughput varies with screen grid thickness at four beamlet currents. The accel grid voltage was varied at each point to achieve a 50 V margin against electron backstreaming. The only beamlet current to show a significant dependence on screen grid thickness was 0.1 mA. In general, the throughput capability increases with increasing screen grid thickness because the required accel grid voltage decreases.

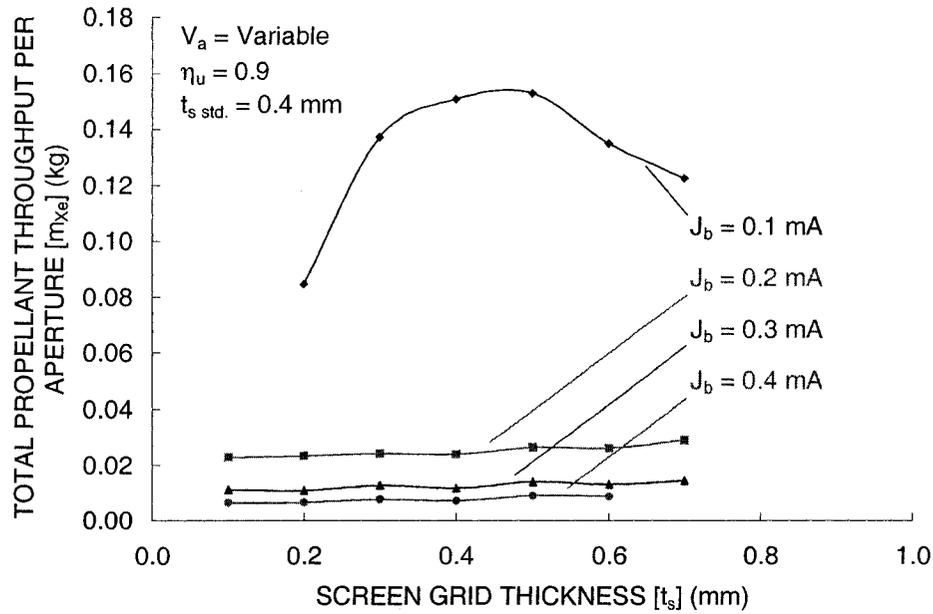


Figure 4.29 Mild variation in the predicted propellant throughput capability with screen grid thickness.

4.4.5 Accel Grid Voltage, V_a

The accel grid voltage is biased negative in order to both extract ions from the discharge chamber and to prevent electron backstreaming. Because many of the ions that erode away the accel grid originate downstream of the grids in a region of near zero potential, the accel grid voltage has a very direct effect on the accel grid erosion rate. The accel grid voltage should always be kept sufficiently negative to prevent electron backstreaming, but not excessively negative such that the accel grid erosion rate is increased needlessly.

The accel grid voltage, V_a , affects the perveance fraction, f_p , equation through the total accelerating voltage, V_T , which is the sum of the net accelerating voltage, V_N , and the accel grid voltage magnitude, $|V_a|$. Figure 4.30 shows the crossover and perveance limits obtained as a function of perveance fraction and beamlet current for five accel grid voltages. Although the perveance fraction to beamlet current ratio is technically different for each accel grid voltage, meaning a second x or y-axis should not strictly be used, the figures that follow use the ratio at the reference accel grid voltage of -400 V. The beamlet current axis in each case is more correct.

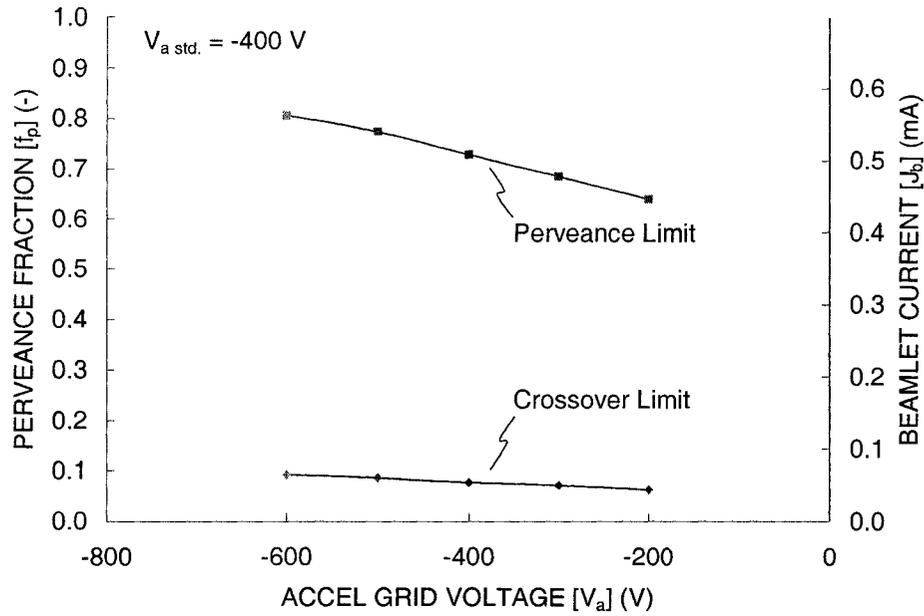


Figure 4.30 Increase in the amount of extracted current with increasing accel grid voltage magnitude.

According to the perveance fraction equation, the beamlet current, J_b , at the perveance limit should vary as $V_T^{3/2}$. If the equation correctly models the situation. Plotting the beamlet current data versus total accelerating voltage, using a net accelerating voltage of 2266 V, shows that the beamlet current at the perveance limit varies as $V_T^{1.56}$. Similarly, the crossover limit beamlet current varies as $V_T^{2.49}$.

The centerline saddle point potential is shown in Figure 4.31 as a function of perveance fraction for five accel grid voltages. At any beamlet current, the saddle point potential varies nearly linearly with accel grid voltage. In this case, at beamlet currents less than about 0.15 mA, a 100 V change in V_a results a 97 V change in $V_{\text{SaddlePoint}}$. At beamlet currents greater than about 0.25 mA, a 100 V change in V_a results in a 88 V change in $V_{\text{SaddlePoint}}$.

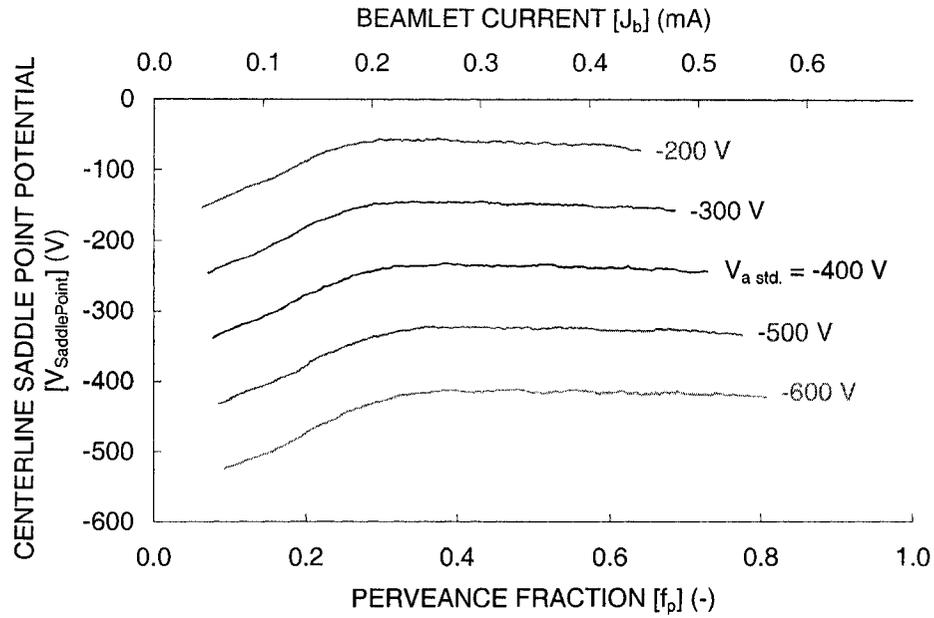


Figure 4.31 Near linear saddle point potential variation with accel grid voltage.

The 95 % divergence angle is shown in Figure 4.32 as a function of perveance fraction. The divergence angle decreases with less negative accel grid voltages. When using a less negative accel grid voltage, the neutralization surface downstream of the grids moves upstream, closer to the accel grid. Downstream of the neutralization surface, the electric fields are weak compared to the momentum that the ions gain in passing through the grids. After the beam ions pass through the neutralization surface, their trajectories stay nearly constant until they pass out of the simulation volume. A neutralization surface located farther upstream is desirable in that the ion beamlet will expand mainly in the region upstream of the neutralization surface.

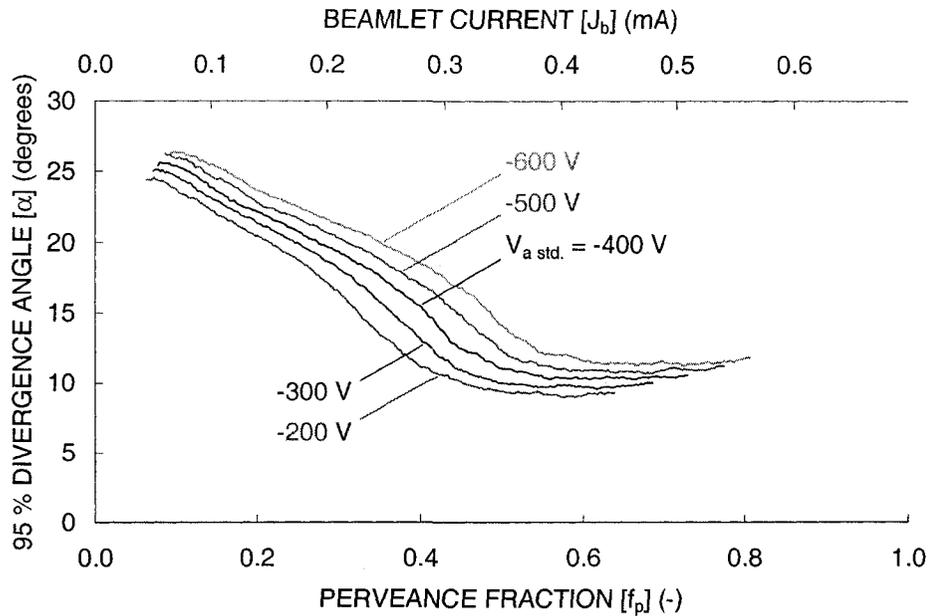


Figure 4.32 Divergence angle improvement with less negative accel grid voltages.

Although it is not shown here, the grid transparency to ions increases with a more negative accel grid voltage. With a very negative accel grid voltage, the plasma density does not need to be as high as it does with a less negative accel grid voltage in order to extract a constant beamlet current. Making the accel grid more negative thus moves the plasma sheath farther upstream, increasing the ion transparency.

If a constant accel grid voltage is used over the entire life of the grids, it is desirable to select an accel grid voltage that will result in structural failure and the onset of electron backstreaming occurring at the same time, thereby maximizing the life of the grids. Figure 4.33 shows the predicted propellant throughput as a function of perveance fraction for grids operating with accel voltages of -200 , -400 , and -600 V. The data in this figure are all from full lifetime simulations.

Using an accel grid voltage of -200 V results in structural failure (50 % accel grid mass loss) and the onset of electron backstreaming occurring at nearly the same propellant throughput values for beamlet currents from about 0.3 to 0.4 mA. Thus, -200 V would be a good choice for the accel grid voltage if the grids were to be operated in that beamlet current range. For beamlet currents lower than 0.3 mA, electron backstreaming occurs after 50 % accel grid mass loss when using an accel grid voltage of -200 V, thus the accel grid could be made more positive to further extend the grid lifetime.

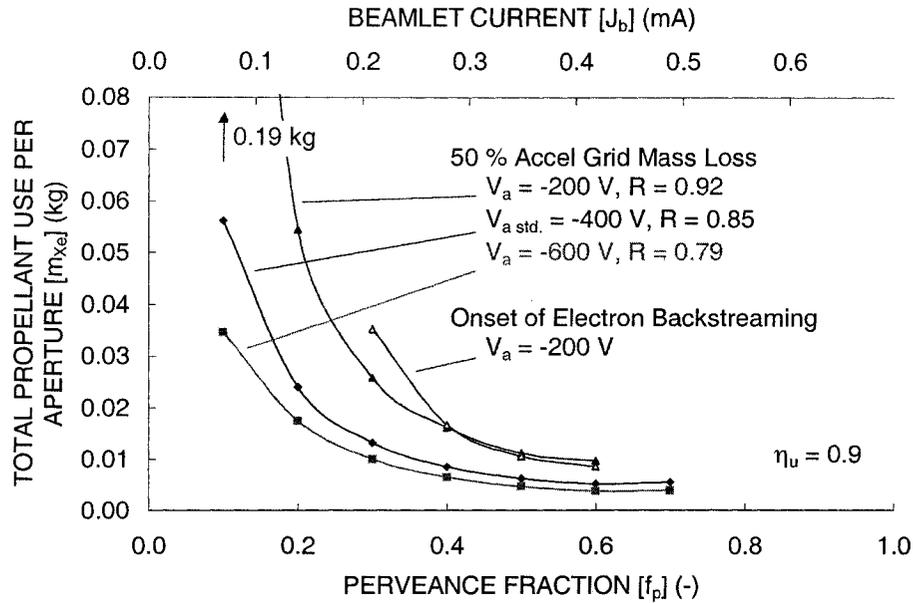


Figure 4.33 Increase in propellant throughput capability with less negative accel grid voltages.

The grid lifetime can be maximized by varying the accel grid voltage during operation such that the accel grid is only kept just negative enough to always prevent electron backstreaming. Keeping the accel grid as positive as possible minimizes the erosion rate. An example of this is shown in Figure 4.34. Here, an initial accel grid voltage of -175 V is used for a beamlet current of 0.3 mA . Every time the centerline saddle point potential, $V_{\text{SaddlePoint}}$, comes within 10 V of the downstream plasma potential, 0 V , the accel grid voltage is made more negative by 25 V .

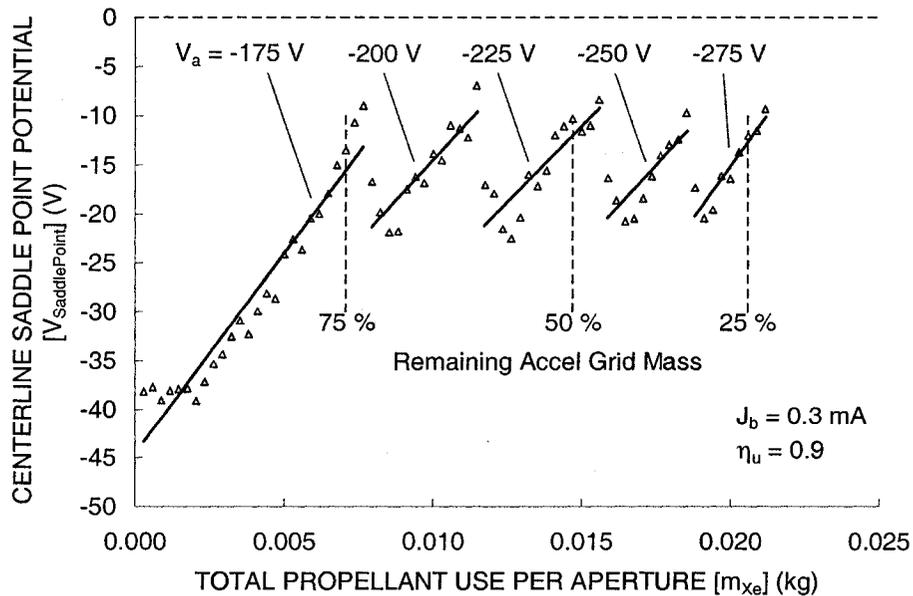


Figure 4.34 Lowest erosion rate achieved with least negative accel grid voltage.

Again using a variable accel grid voltage, Figure 4.35 shows the remaining accel grid mass as a function of propellant throughput for four beamlet currents. The initial accel grid voltage was -175 V in all cases. The alternating thickness lines indicate when an accel grid voltage magnitude increase was made. At the 0.1 mA beamlet current, the accel grid voltage never needed to be made more negative than -175 V to prevent electron backstreaming. For the greater beamlet currents, 50 % accel grid mass loss was seen to occur when the accel grid voltage was -225 V.

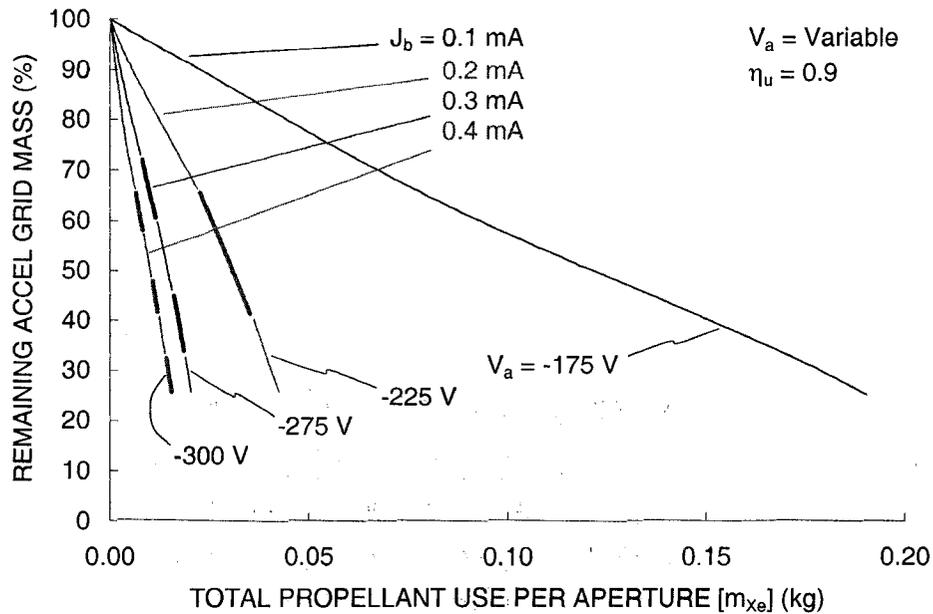


Figure 4.35 Maximizing the grid lifetime by using a variable accel grid voltage.

The goal of using a variable accel grid voltage is to maximize the grid lifetime. Figure 4.36 compares the propellant throughput predictions for the variable accel grid voltage cases with the constant -200 and -400 V cases. Because the initial accel grid voltage of -175 V is close to -200 V, the variable voltage cases are only slightly better.

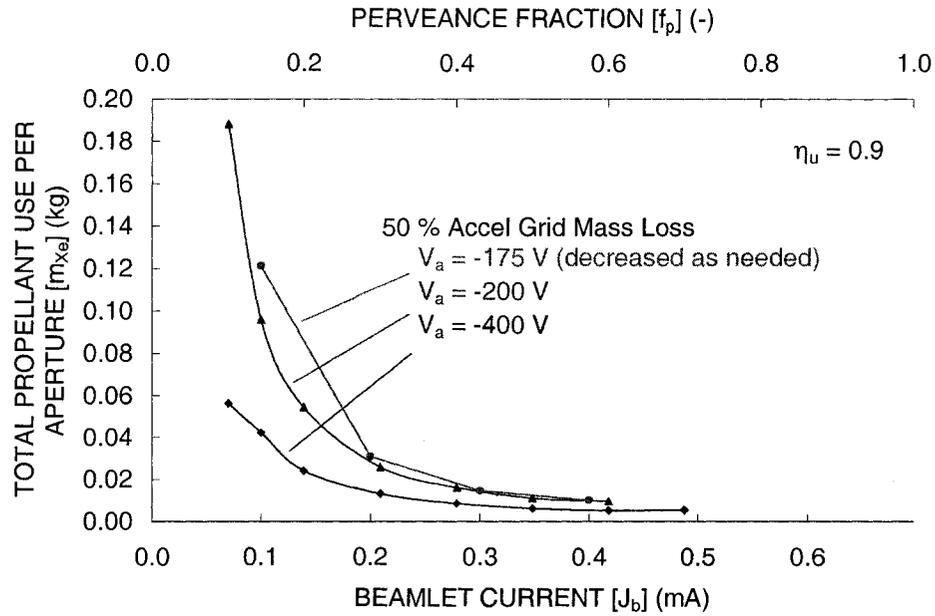


Figure 4.36 Propellant throughput capability using the least negative accel grid voltage.

For illustration, lifetime simulations of grids operating at very low R values were performed, and the results are shown in Figure 4.37. The three cases in this figure are all operating at a perveance fraction of 0.4, and a constant net accelerating voltage, V_N , of 2266 V. The beamlet current ranges from 0.279 mA at an R value of 0.85 to 1.055 mA at an R value of 0.35. The erosion pattern in the 0.35 R value case is markedly different than at the higher R values. The pits are very wide and deep compared to the grooves. Also, the deepest part of the groove does not occur in the center between two apertures. Instead, two deeper depressions are seen just to the sides of the midpoint between holes.

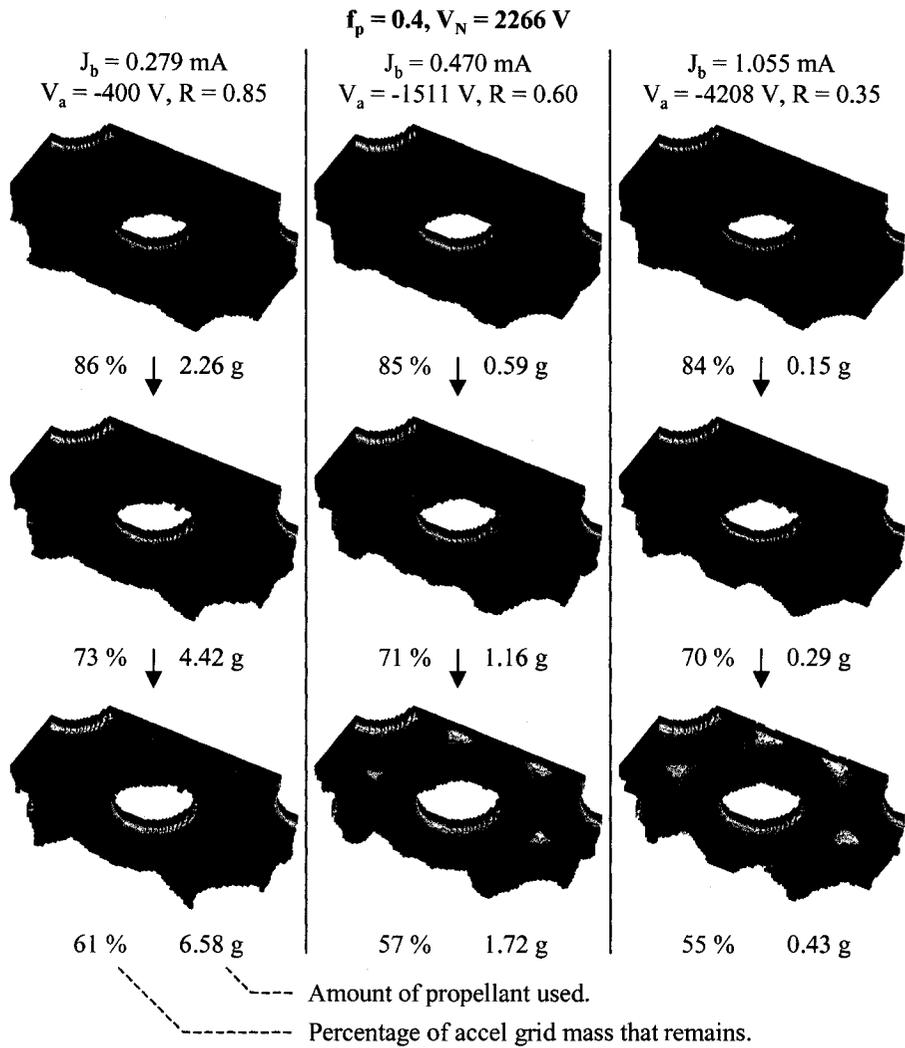


Figure 4.37 Erosion of grids operating at low R values.

5 Special Investigations

This chapter extends the utility of the ffx code through several interesting investigations. Simulation of grid sets with finite numbers of apertures is relevant to the study of sub-scale gridlets and edge apertures. Non-uniform aperture cross sections arise regularly in standard grid manufacturing processes. A study of misaligned aperture pairs helps to shed light on grid performance concerns.

Investigations into the effects of charge exchange ion space charge and initial energy help to direct the appropriate simulation of ion thruster optics. Finally, a method is introduced that utilizes information from individual beamlets to predict overall thruster current density profiles.

5.1 One, Three, and Seven Aperture Grids

For a hexagonal aperture layout, an aperture in the interior of a full grid set is surrounded by 6 apertures operating at nearly the same beamlet current. As a result, symmetric boundary conditions are used in the ffx analysis volume and only a small portion of the grid set needs to be simulated.

It turns out that apertures on the edge of the grid set, which are not symmetrically surrounded by other apertures, behave differently than the interior apertures. This is especially true with respect to the crossover limit.

The ffx code was used to simulate grids with 1 aperture, 3 apertures, 7 apertures, and the standard hexagonal aperture layout. The same code was used to simulate all of these cases, all still using the symmetric boundary conditions. To simulate 1 aperture for example, a quarter hole was placed at the corner of a large simulation square, essentially simulating a large distance to the next aperture. The hexagonal aperture layout is basically the case of an infinite number of adjacent apertures.

Figure 5.1 shows an impingement limit curve extending over the low beamlet current range. The crossover limits are defined here as the points where the impingement to beamlet current ratio, J_a/J_b , goes above 1 %. Charge exchange ion impingement current is not included in the calculation of the impingement current ratio in this figure. The crossover limit is greatest for the 1 aperture grid. As more apertures are simulated, the crossover limit decreases.

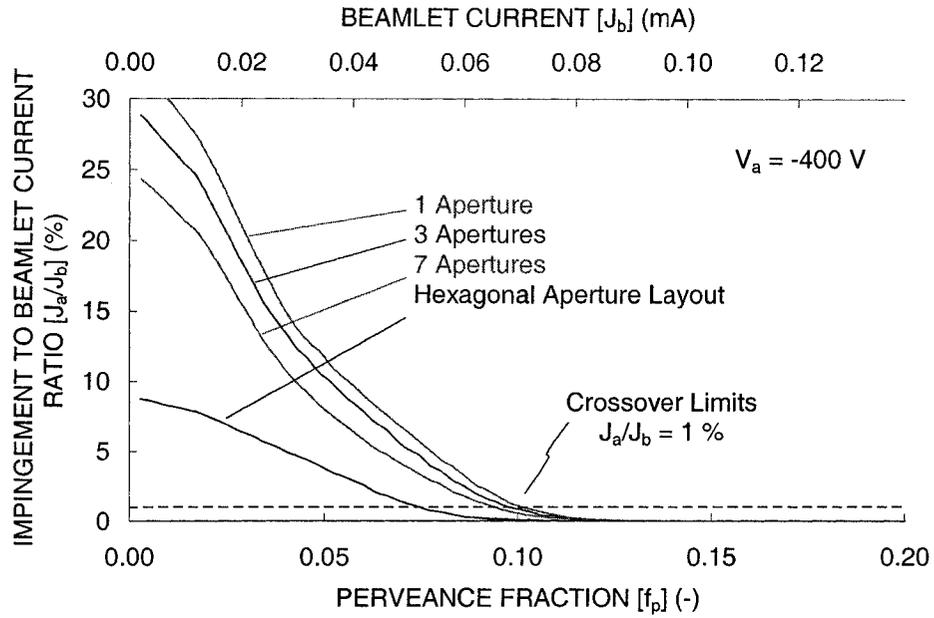


Figure 5.1 Crossover limit change with the number of simulated apertures.

Figure 5.2 shows the crossover limits from the impingement limit curve as a function of the number of simulated apertures. The crossover limit found with a 1 aperture grid is 33 % greater than the crossover limit found with the hexagonal aperture grid. The difference in the crossover limit is, of course, dependent on the exact definition of where the crossover limit occurs.

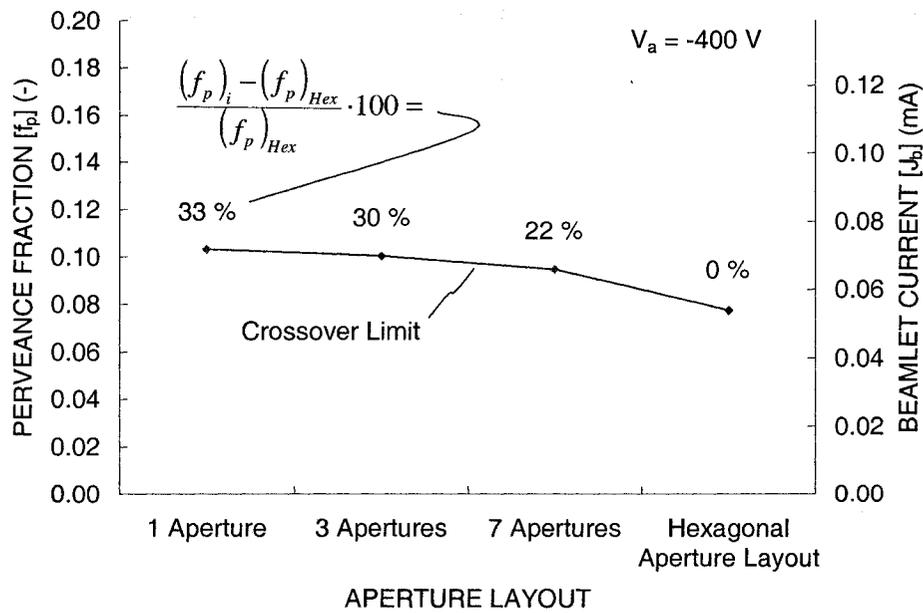


Figure 5.2 Crossover limit decrease with an increasing number of simulated apertures.

The reason the crossover limit changes with the number of simulated apertures is that there are interactions among the sheaths upstream of the apertures. Figure 5.3 shows the sheaths upstream of the four grid cases at a constant beamlet current of 0.035 mA. In the 1, 3, and 7 aperture cases, the sheath is located a fixed distance upstream of the screen grid near the boundary edges, away from any apertures. Between the apertures, the sheaths interact with one another and move farther upstream. In the hexagonal aperture layout case, the sheath is located much farther upstream than in the other cases.

$$f_p = 0.05 (J_b = 0.035 \text{ mA})$$

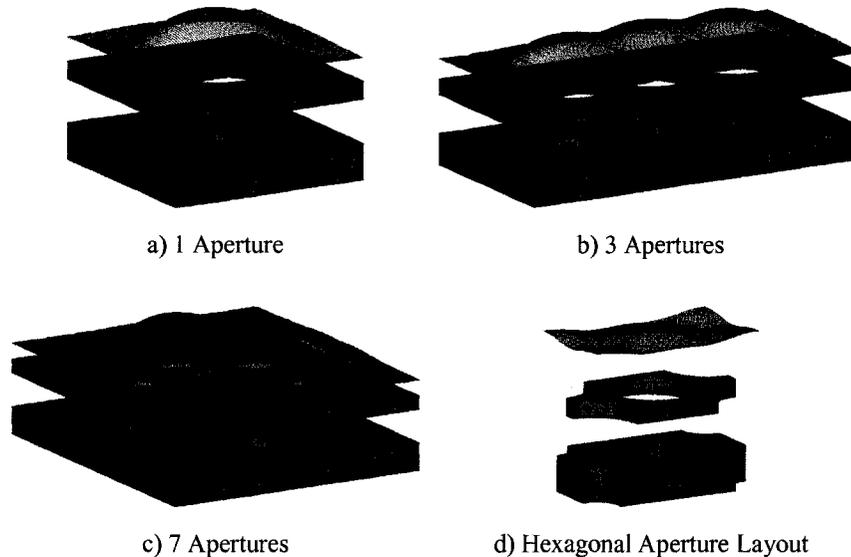


Figure 5.3 Sheath location as a function of the number of apertures.

The greatest crossover limit will be seen in the case where the sheath is the most dishd. Ions entering the sheath at the edge of the aperture are turned inward toward the beamlet centerline. For a highly curved sheath, these ions will receive a large radial velocity, causing impingement onto the accel grid sooner than in a case where the sheath is less curved. Beginning with the 3 aperture case, the sheath is already less curved between the apertures than in the 1 aperture case. Ions entering the sheath between the apertures are therefore less likely to strike the accel grid than ions entering from the sides of the apertures that do not have neighboring apertures. This results in a lower impingement current, and a lower crossover limit.

Figure 5.4 shows ion beamlets that correspond to the four grid cases operating at a beamlet current of 0.035 mA, which is below the crossover limit in all cases. The sheath above the center aperture in the 7 aperture grid is located farther upstream than the six surrounding sheaths.

$$f_p = 0.05 (J_b = 0.035 \text{ mA})$$

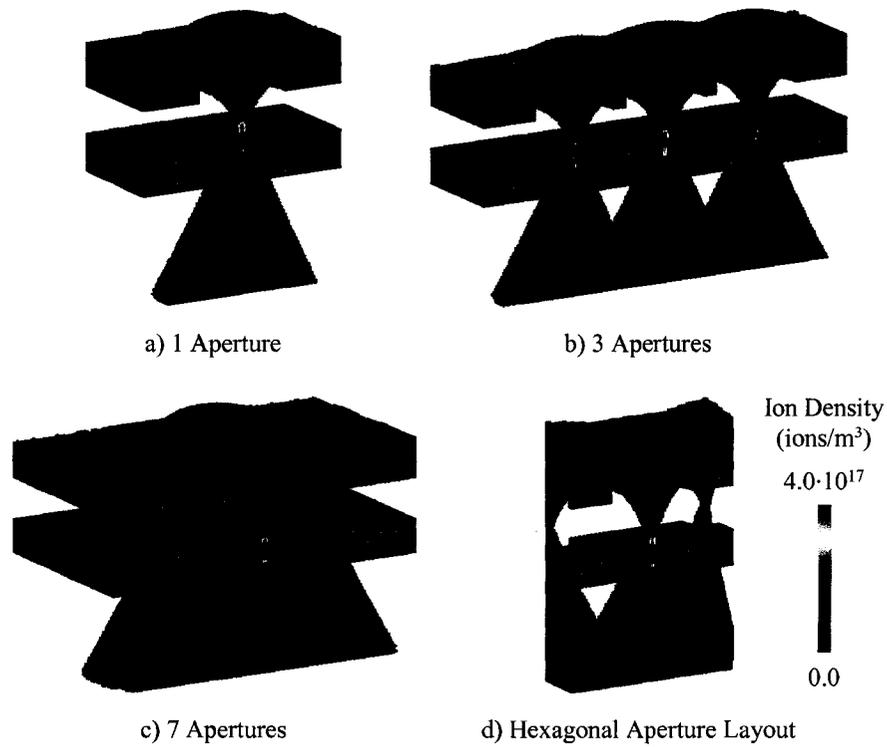


Figure 5.4 Ion beamlets below the crossover limit.

In this case, only the crossover limit is affected by the number of apertures in the grid. Figure 5.5 shows ion beamlets for the four grid cases at a beamlet current of 0.5 mA, which is just below the perveance limit of 0.508 mA. Near the perveance limit, the sheath is located against the screen grid in all cases. As a result, none of the sheaths upstream of the apertures interact, and the perveance limit should be almost exactly the same in each aperture. In cases where the sheath is not against the screen grid at the perveance limit, such as when using a very thin screen grid for example, the number of apertures might affect the perveance limit as a result of sheath interactions.

$$f_p = 0.717 (J_b = 0.500 \text{ mA})$$

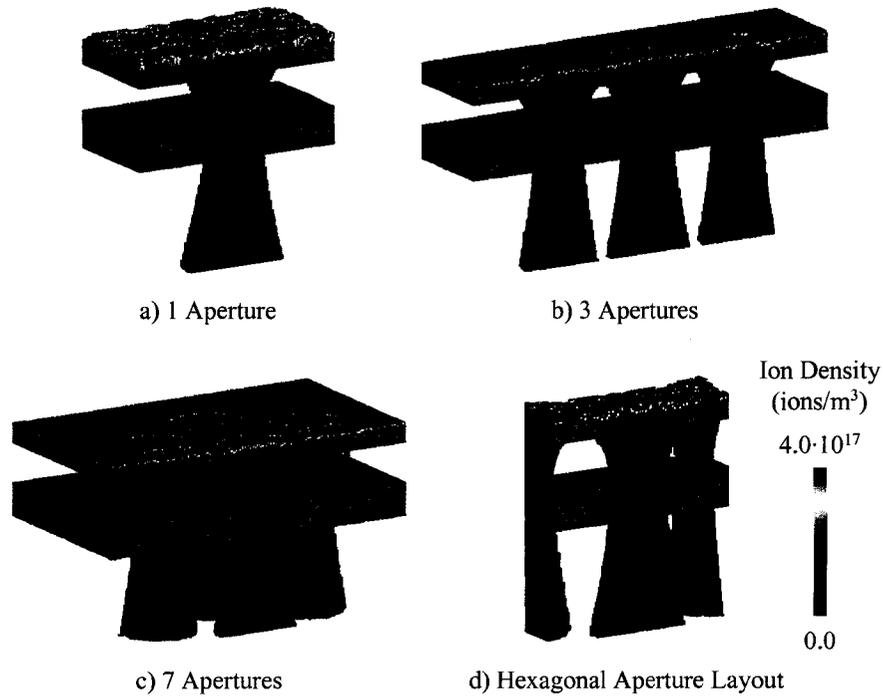


Figure 5.5 Constant perveance limit with the number of simulated apertures.

5.2 Grid Cusps and Linear Grids

Chemical etching grid apertures can result in cusps on the inside walls of the screen and accel grid holes. This is in contrast to drilling the grid apertures, using a drill bit or a laser for example, where the resulting hole walls are flat. Cusp height (h) was varied as a percentage of each grid's thickness (t_s or t_a). Figure 5.6 compares a drilled grid and a cusped grid.

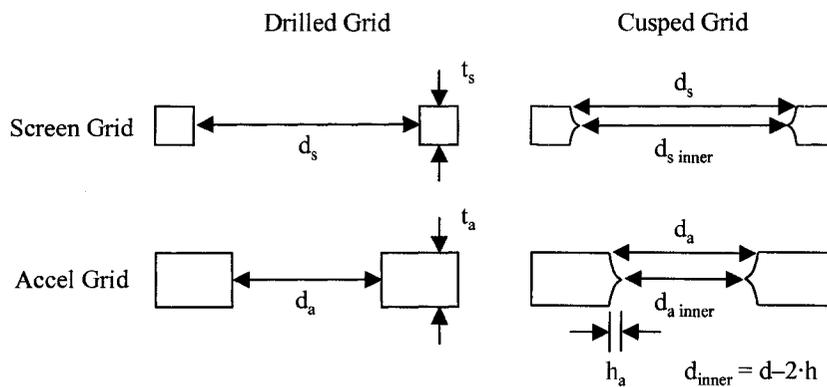


Figure 5.6 An example of a cusped grid.

Figure 5.7 shows how the crossover limit is increased, and the perveance limit decreased, with increasing cusp height. The operating range is reduced essentially because the effective diameter of the accel grid hole is reduced.

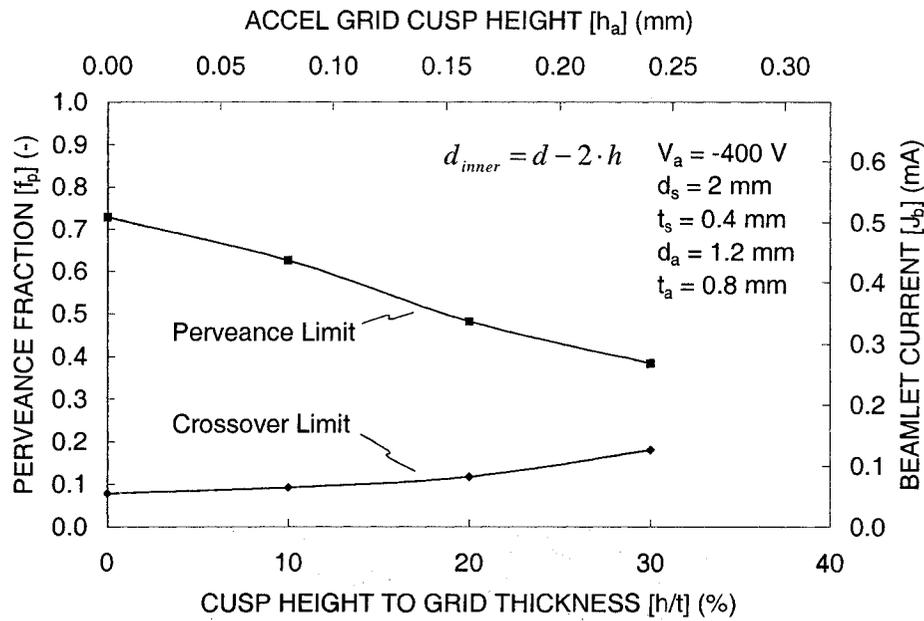


Figure 5.7 Decrease in operating range with increasing cusp height.

Figure 5.8 compares the crossover and perveance limits as a function of accel grid hole diameter for drilled and cusped grids. The crossover limits in the cusped case are slightly lower than in the drilled case because the crossover limit occurs when ions strike the downstream side of the accel grid holes. For the same inner hole diameter, the cusped grids have a greater downstream edge hole diameter, and therefore the beamlet current has to be reduced further compared to the drilled grids to reach the crossover limit.

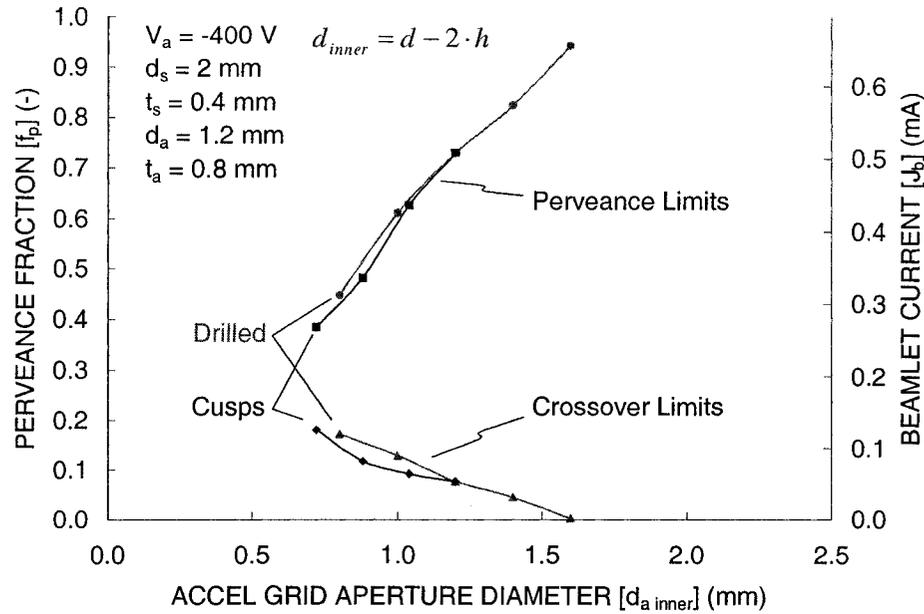


Figure 5.8 Similarity between variations in cusp height and accel grid hole diameter (d_a).

Cases where the hole walls were angled were also studied. These cases are referred to as linear grids here, because there is a linear variation in the hole diameter in moving between the upstream and downstream sides of the grid. The crossover and perveance limit variations with changing hole wall angle are shown in Figure 5.9. Changes in the screen grid wall angle are denoted by the open symbols, while changes in the accel grid wall angle are denoted by the closed symbols. The inner hole diameters were held constant, meaning the smallest part of the screen grid was always 2.0 mm in diameter, and the smallest part of the accel grid was always 1.2 mm in diameter.

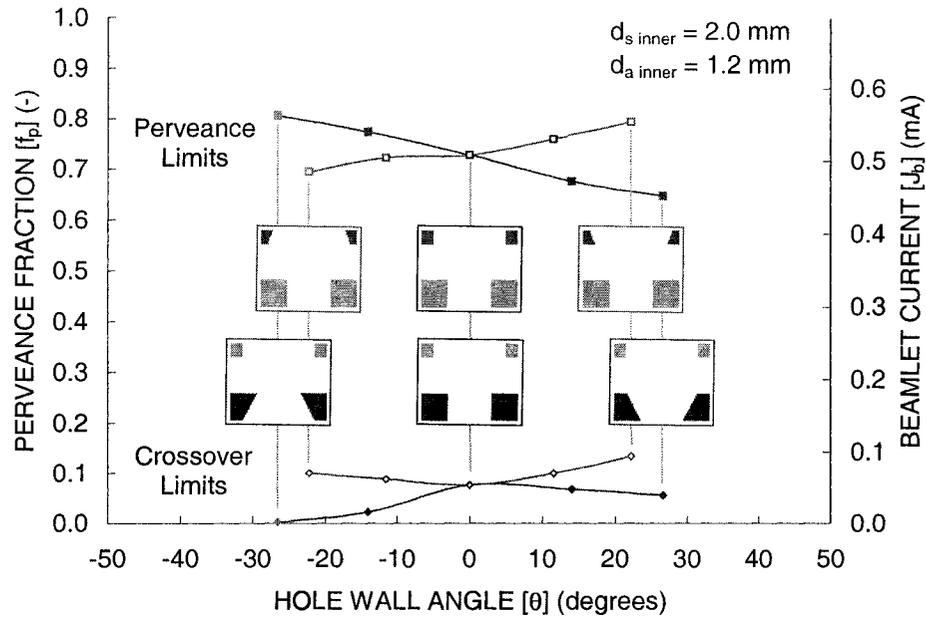


Figure 5.9 Crossover and perveance limits as a function of hole wall angle.

With respect to the accel grid, the drawback of angling the hole wall in either direction is that the resistance to electron backstreaming is reduced as a result of the accel grid being moved farther away from the beamlet centerline, as shown in Figure 5.10. Changing the screen grid wall angle also slightly changes the saddle point potential, but the effect is minor and is therefore not reported.

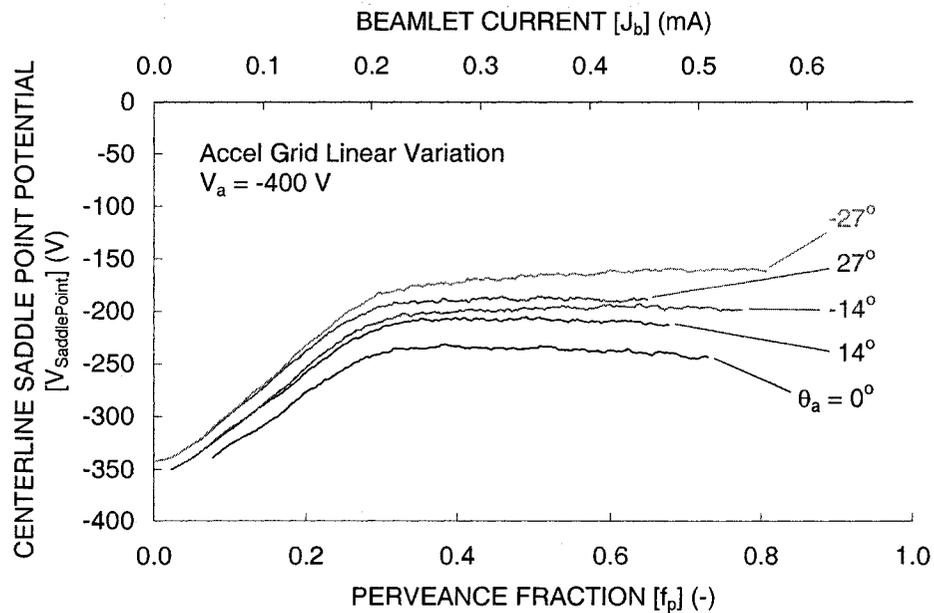


Figure 5.10 Reduction in the resistance to electron backstreaming with angled accel grid hole walls.

When operating an aperture near the perveance limit, high energy charge exchange ions that originate between the grids can erode the upstream side of the accel grid aperture away quickly. This can cause the accel grid to take on the shape of a linear hole that has a positive wall angle. Figure 5.11 shows three beamlets, all operating at a perveance fraction of 0.75, with negative, zero, and positive accel grid wall angles. Erosion at this beamlet current makes the accel grid look like the grid with the positive wall angle, increasing the impingement current and erosion rate.

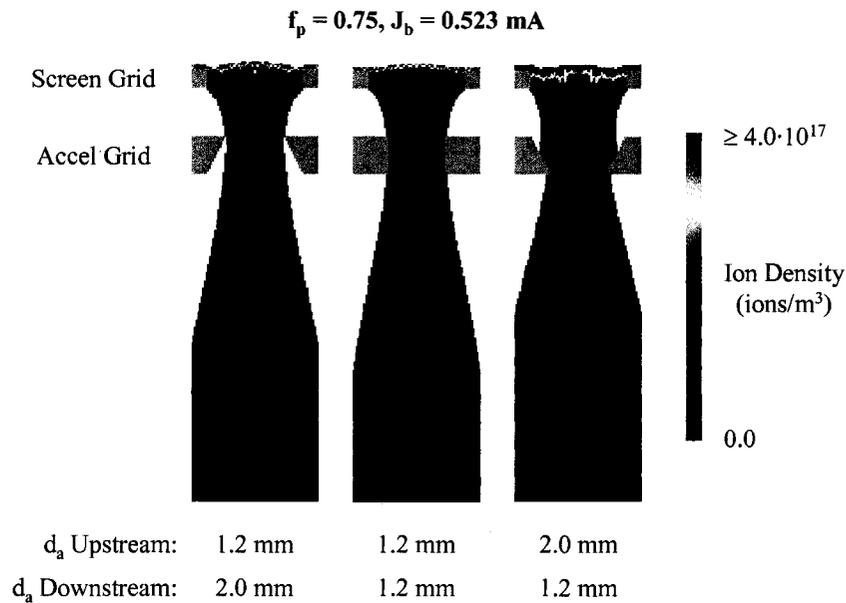


Figure 5.11 Changing impingement current with accel grid wall angle using a constant minimum hole diameter.

Using a positive wall angle for the screen grid and a negative wall angle for the accel grid results in a grid set with a very large operating range (low crossover limit and high perveance limit). It turns out, however, that the ion beamlets for this case have worse divergence angles (α) than for the normal grid case. To minimize divergence angle over the entire operating range, the opposite situation is better, i.e. using a negative wall angle for the screen grid and a positive wall angle for the accel grid. This is shown in Figure 5.12.

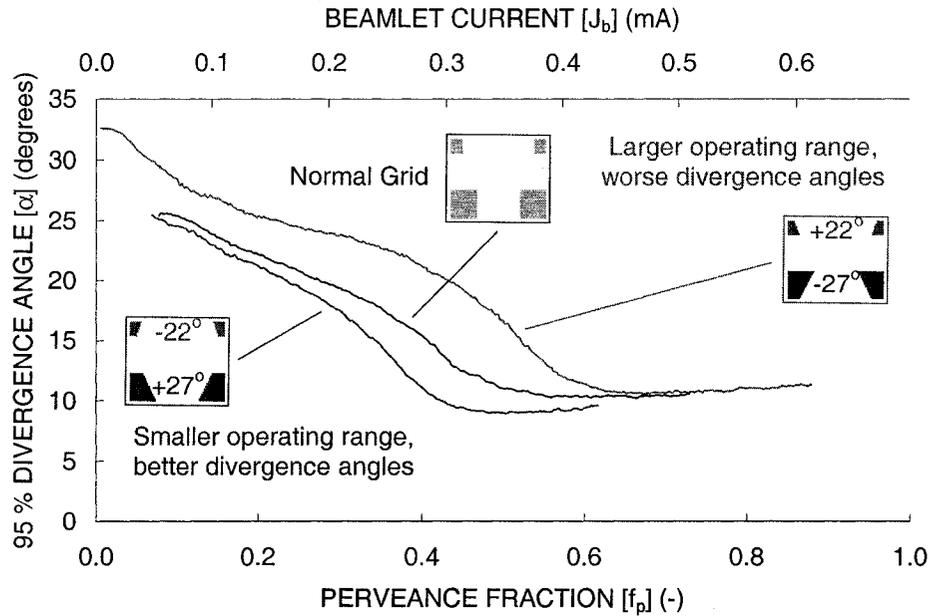


Figure 5.12 Improved divergence angles with slanted aperture walls.

5.3 Offset Apertures

The screen and accel grid holes should be exactly aligned to provide the lowest possible crossover limit and highest possible perveance limit. In reality, it is difficult to align every hole pair considering that there are often thousands of apertures in a grid. It is beneficial then to be able to see how misaligned apertures perform compared to perfectly aligned apertures.

Hole misalignment was simulated using a single aperture. Note that the crossover and perveance limits found using a single hole may be different than those found using a full hexagonal aperture layout pattern. Figure 5.13 shows the crossover and perveance limits as a function of accel grid hole offset. The center of the accel grid hole was offset by a percentage of the screen grid hole diameter. A 10 % offset (δ/d_s) for the 2 mm screen grid hole diameter (d_s) is an offset (δ) of 0.2 mm for example. The crossover limit was seen to increase by 124 % at a 15 % misalignment compared to the aligned crossover limit. Similarly, the perveance limit was reduced by 60 % at a 15 % misalignment compared to the aligned perveance limit.

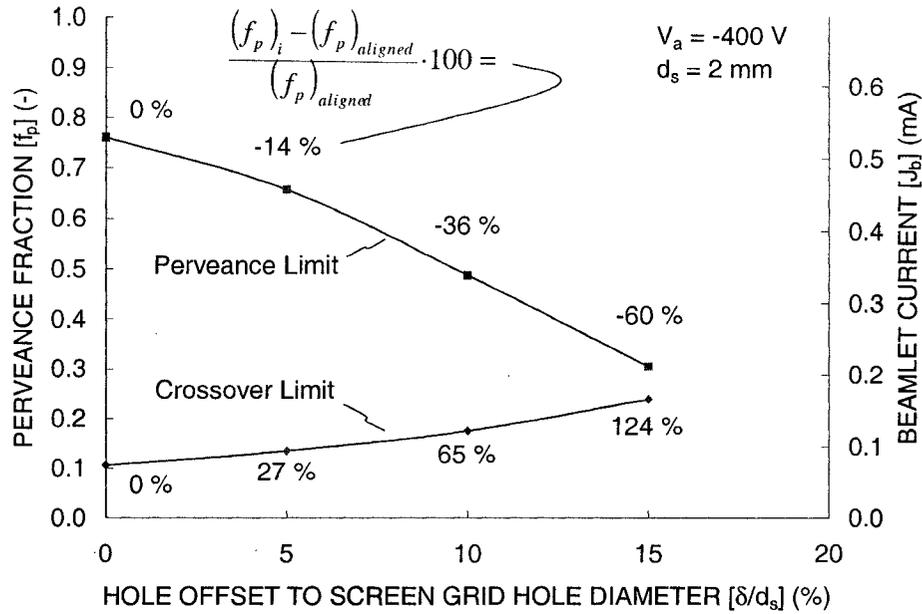


Figure 5.13 Crossover and perveance limit variation with hole misalignment.

Occasionally, the screen and accel grid holes might be intentionally misaligned in order to improve characteristics of the ion beam. To produce a more collimated beam for example, the holes toward the perimeter of the grid can be misaligned in order to steer the ion beamlets toward the beam centerline. Figure 5.14 shows trajectories of ions as they pass through apertures in various states of misalignment. Notice that moving the accel grid in one direction actually has the effect of steering the ion beamlet in the opposite direction.

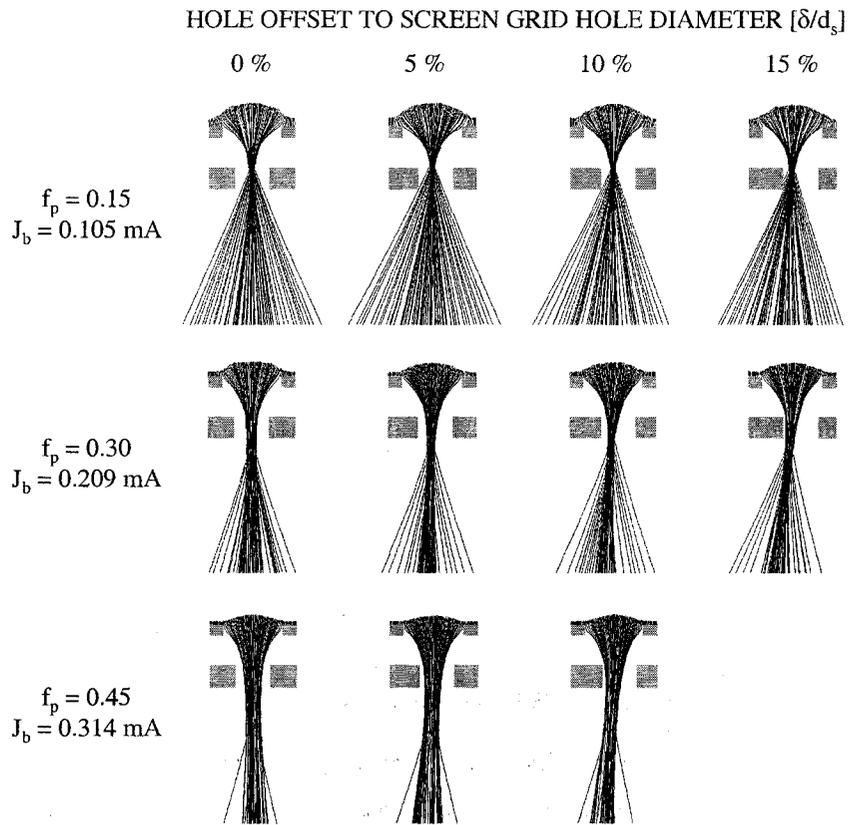


Figure 5.14 Ion trajectories passing through misaligned apertures.

Figure 5.15 shows the steering angle of the beamlet as a function of accel grid hole offset. The steering angle for each case was found by tracking the location of the ion density centroid as a function of axial distance downstream of the accel grid. A 10 % misalignment turns the beamlets by an average of 4.1 degrees. Both the near linear trend and magnitude of deflection angle with hole offset shown in this figure are similar to those reported by Okawa.

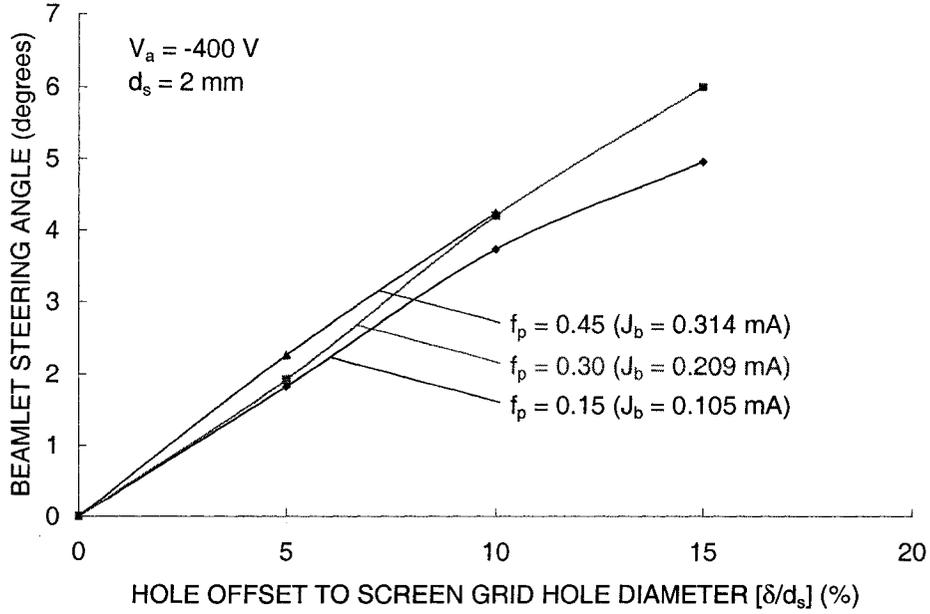


Figure 5.15 Beamlet steering as a result of hole misalignment.

One possible consequence of offsetting the screen and accel grid holes is that the ion beamlet cross section could become noncircular. The circular nature of the cross section was checked three ways. First, the cross sections were simply checked visually. Second, ellipses were fit to the ion density cross sections along the axial distance of the beamlet. The equation of an ellipse is given in Eq. 5.1. At each axial location, the center of the ellipse was placed at the ion density centroid (C_x, C_y), and the ellipsoid parameters (a and b) were adjusted to find the smallest ellipse in terms of area (πab) that encompassed a given percentage of the ions on that plane. The ratio of the semi-major to semi-minor axis lengths was determined to describe how noncircular the beamlet was. Third, the ion density variation passing through the centerline of the beamlet was checked for skewness, as defined in Eq. 5.2, at several axial locations.

Ellipse

$$\frac{(x - C_x)^2}{a^2} + \frac{(y - C_y)^2}{b^2} = 1 \quad \text{Eq. 5.1}$$

Skewness

$$SK = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad \text{Eq. 5.2}$$

n = number of samples

\bar{x} = mean

s = standard deviation

None of the above tests revealed a significant change in the circularity of the ion beamlet with changing accel grid hole offset. It could be that the changes in circularity were not detectable with the

resolution used in these simulations. It could also be that a different grid setup (changing the accel grid hole diameter, voltages, etc.) might yield greater changes in beamlet deflection and possibly allow the observation of a noncircular cross section.

5.4 Charge Exchange Ion Space Charge

Most of the time, charge exchange ion space charge can be neglected. The effect of charge exchange ion space charge on centerline saddle point potential is shown in Figure 5.16 for three propellant utilization values. For the 0.3, 0.5, and 0.7 perveance fractions, the saddle point potential rises by about 1.5 % and 6.3 % on average for the 0.6 and 0.9 propellant utilization efficiencies respectively.

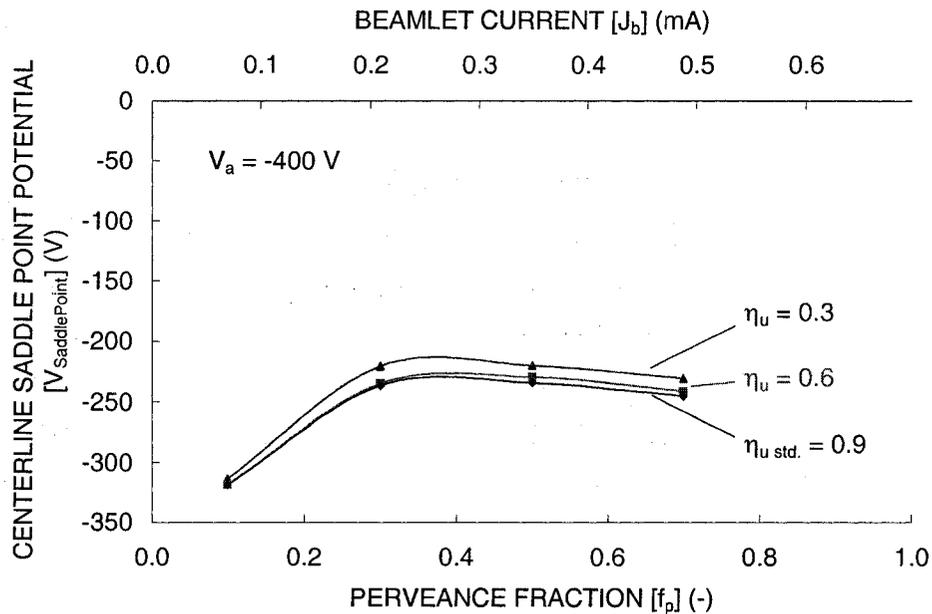


Figure 5.16 Variation in saddle point potential as a function of propellant utilization efficiency when including charge exchange ion space charge.

Figure 5.17 compares the ion density contribution from charge exchange ions to the ion density contribution from beam ions at a perveance fraction of 0.3. Note that the ion density ratio is only shown in locations where beam ions are present. Even at a propellant utilization efficiency of 0.3, the charge exchange ion density is less than 20 % of the beam ion density in the region near the centerline saddle point potential.

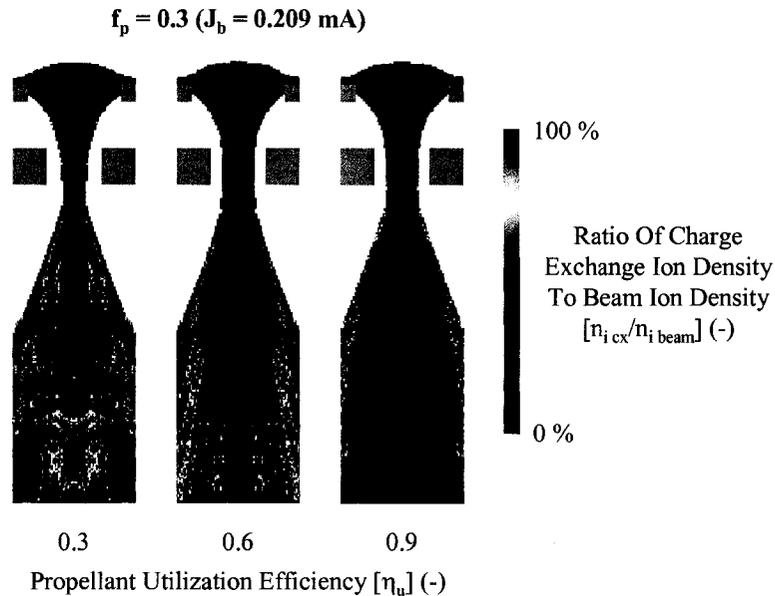


Figure 5.17 Increase in space charge with decreasing propellant utilization efficiency.

5.5 Charge Exchange Ion Initial Energy

When a charge exchange reaction takes place, it is thought that there is no momentum transfer between the ion and neutral molecule. As a result, charge exchange ions have a very low kinetic energy upon creation, only due to the thermal velocity of the neutral molecules before the reaction.

For nearly all of the results presented here, the initial energies of charge exchange ions are obtained by sampling a Maxwell Boltzmann distribution characterized by a temperature of 500 K (0.043 eV).

Changing the initial charge exchange ion energy changes the resulting erosion pattern. Figure 5.18 shows three erosion patterns at a perveance fraction of 0.4 using three charge exchange initial temperatures. The familiar pit and groove erosion pattern is most acute with the lowest charge exchange ion temperature, 0.043 eV. As the temperature is increased, the erosion is more evenly distributed across the accel grid downstream surface. This occurs because the ions have more initial energy to overcome the small electric fields that would otherwise vector the ions into the pits and grooves. Because the expected erosion pattern occurs most visibly at the neutral propellant temperature, 500 K, it is thought that using this temperature is most correct.

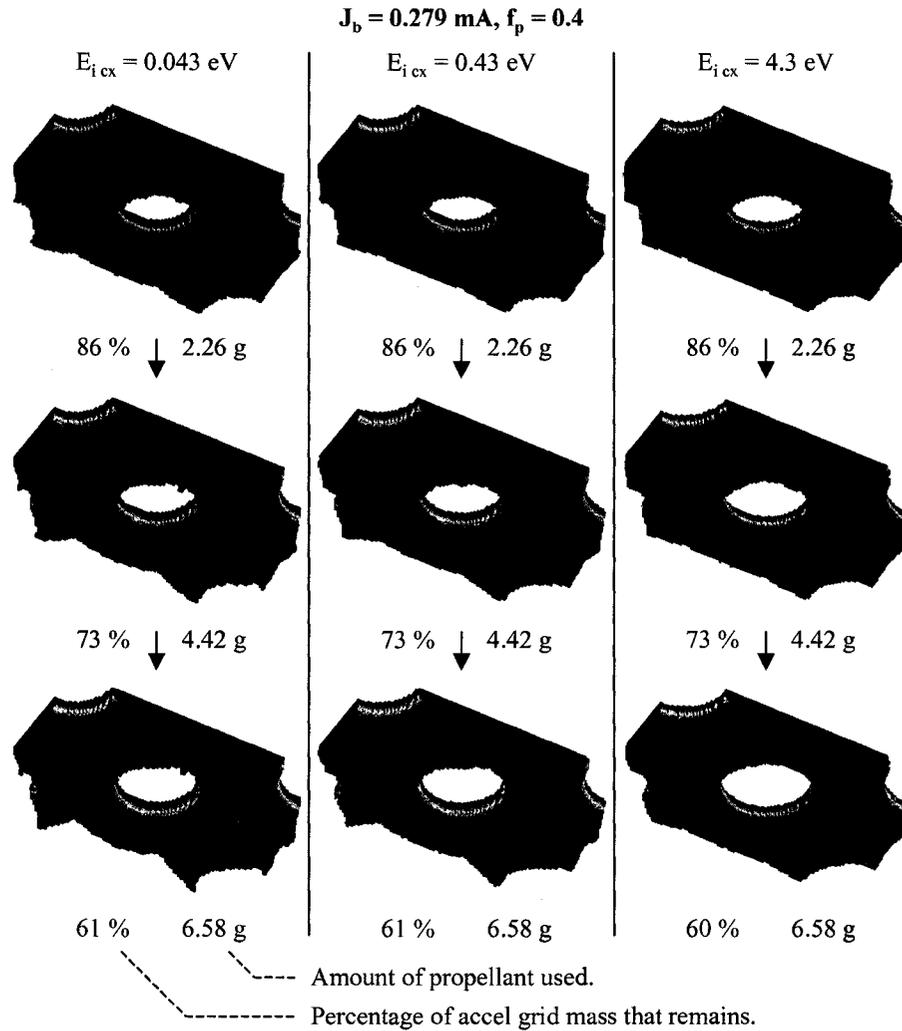


Figure 5.18 Accel grid erosion pattern as a function of charge exchange ion initial energy.

5.6 Fixed Emission Surface

Normally, the sheath between the discharge chamber plasma and ion optics sets up automatically, finding the appropriate location according to the plasma density and grid voltages. As a result, a crossover limit is often found at low plasma density when the sheath moves upstream, and a perveance limit is found at high plasma density when the sheath moves downstream.

If the sheath could be fixed in a single location instead of allowing it to move, the crossover and perveance limits could be improved for instance. One way to accomplish this has been proposed in the Emissive Membrane Ion Thruster (EMIT) concept [Wilbur].

For this study, it will be assumed that a theoretical fixed surface from which ions are generated, basically the sheath, could be created. This surface is of arbitrary shape, and ions are generated with uniform current density only at desired locations.

One way to specify the fixed sheath shape is by using a Bezier curve. Bezier curves are named for Pierre Bezier, who used them for automobile design starting in 1962. A Bezier curve is fully described by a number of control points, P_i . A parametric parameter, u , where $0 \leq u \leq 1$, describes a single location along the curve.

The fixed sheath location will be described in radial, r , and axial, z , coordinates. Two Bezier curves are needed, one for $r(u)$ and another for $z(u)$. Cubic Bezier curves were selected for use in this application, as given in Eq. 5.3. A cubic Bezier curve, where the greatest power of u is 3, uses four control points: P_0, P_1, P_2 , and P_3 . The terms in front of the control points are called blending functions, B_i .

$$r(u) = (1-u)^3 R_0 + 3(1-u)^2 u R_1 + 3(1-u)u^2 R_2 + u^3 R_3$$

$$z(u) = (1-u)^3 Z_0 + 3(1-u)^2 u Z_1 + 3(1-u)u^2 Z_2 + u^3 Z_3$$

Eq. 5.3

An example Bezier curve is shown in Figure 5.19. At the start of the curve, where $u = 0$, the curve is tangent to the line connecting the first and second control points, P_0 and P_1 . The curve is also tangent to the line connecting the last two control points at the end of the curve. In general, the curve only passes through the first and last control points.

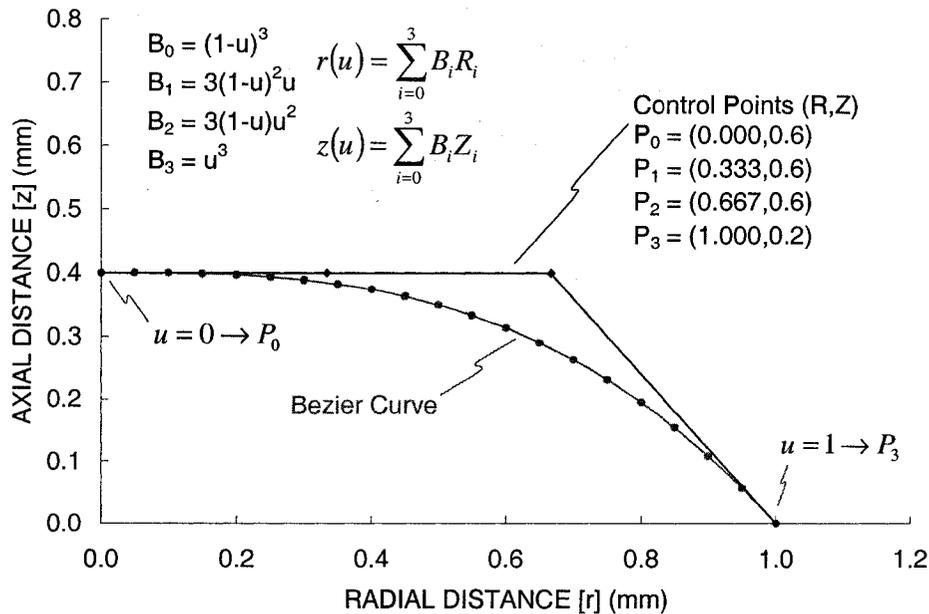


Figure 5.19 An example Bezier curve used as the ion macro particle release surface.

The Bezier curve is rotated around the z -axis to create the fixed sheath, or ion release surface, as shown in Figure 5.20. All points above the fixed sheath are held at a potential equal to the upstream plasma potential. In this case, the fixed sheath was made to extend to the same diameter as the conventional screen grid used previously. Because the screen grid lies entirely upstream of the sheath, in this particular case, it does not affect the simulation and is shown only to reference the ion release surface location. Ion macro

particles were released from this surface within a radius of 0.9 mm from the centerline, less than the radius of the last control point, 1.0 mm.

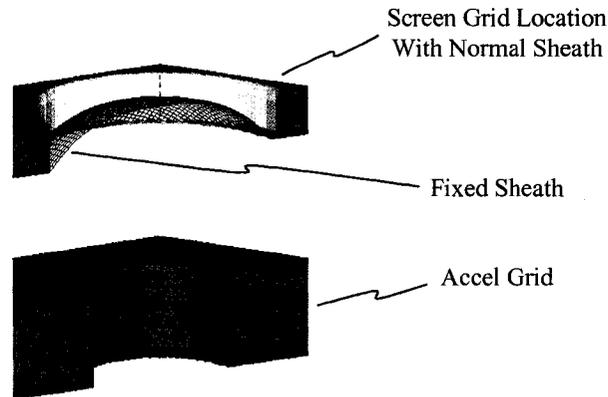


Figure 5.20 The example fixed sheath shown with the screen and accel grids.

Locations from which the ion macro particles are released are selected using a random number generator. Most random number generators provide numbers evenly distributed from 0 to 1 inclusive, i.e. $0 \leq n \leq 1$. Two random numbers are used to obtain each particle's starting location, as summarized in Eq. 5.4. The first number, n_1 , selects u , which in turn sets the r and z particle location. The second number, n_2 , selects θ , the angle to the particle about the z -axis, which sets the x and y particle location using r .

$$\begin{aligned}
 u = n_1 &\rightarrow \begin{aligned} &r(u) \\ &z(u) \end{aligned} \\
 \theta = n_2 \frac{\pi}{2} &\rightarrow \begin{aligned} &x = r \cdot \cos(\theta) \\ &y = r \cdot \sin(\theta) \end{aligned}
 \end{aligned}
 \tag{Eq. 5.4}$$

Particle locations sampled evenly in the u - θ space end up being unevenly distributed when placed into the simulation, as shown in Figure 5.21. The particles with small u values are placed closer together than those with large u values. If all particles are equally weighted, meaning each macro particle represents the same number of ions, the current density near $r = 0$ will be greater than the current densities at greater radii.

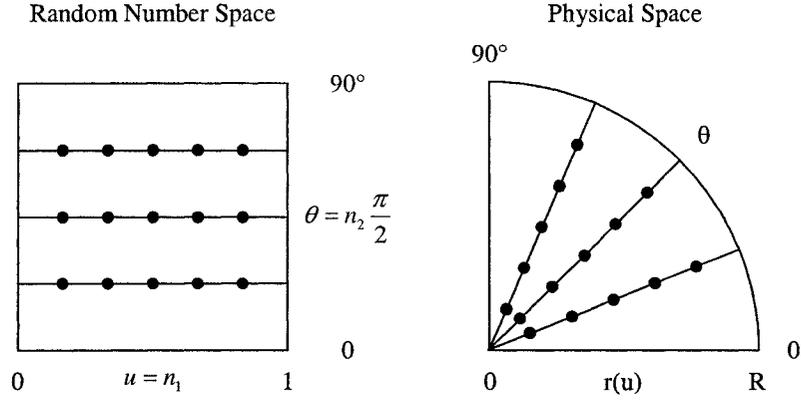


Figure 5.21 Starting particle locations chosen evenly in u and θ may end up unevenly distributed in the physical simulation.

One way to obtain equal current density over the sheath surface is to weight each particle according to its starting location. The surface area, A , of the fixed sheath can easily be found through numerical integration since the equation of the surface is known, as given in Eq. 5.5. The differential surface area, dA , at any value of u is given in Eq. 5.6.

$$A = \int_0^{2\pi} \int_0^1 r(u) \sqrt{\left(\frac{\partial r(u)}{\partial u}\right)^2 + \left(\frac{\partial z(u)}{\partial u}\right)^2} du d\theta \quad \text{Eq. 5.5}$$

$$A = \int_0^{2\pi} \int_0^1 dA(u) du d\theta \quad \text{Eq. 5.6}$$

$$dA(u) = r(u) \sqrt{\left(\frac{\partial r(u)}{\partial u}\right)^2 + \left(\frac{\partial z(u)}{\partial u}\right)^2}$$

The maximum differential area can occur at any u value, not necessarily at $u = 1$, depending on the rate of change of the surface with u , which is affected by the positioning of the control points. The derivative of $r(u)$, which is similar in form to the derivative of $z(u)$, is given in Eq. 5.7. Figure 5.22 shows the differential surface area of the curve specified by the given control points plotted as a function of radial distance. In this particular case, r and u are linearly related because the control points are evenly spaced in the radial direction.

$$\begin{aligned} \frac{\partial r(u)}{\partial u} = & (-3 + 6u - 3u^2)R_0 + (3 - 12u + 9u^2)R_1 \\ & + (6u - 9u^2)R_2 + 3u^2R_3 \end{aligned} \quad \text{Eq. 5.7}$$

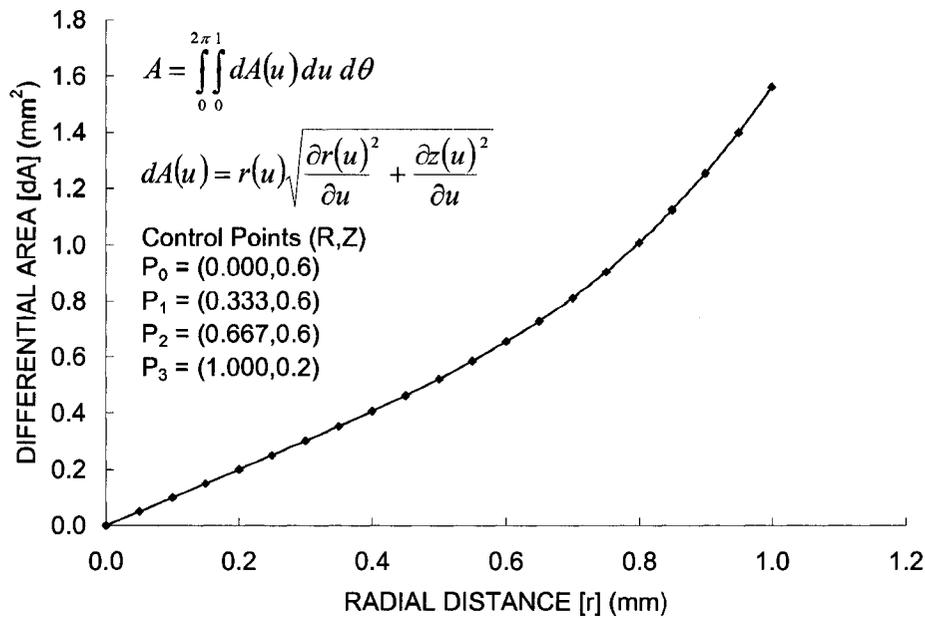


Figure 5.22 The differential surface area variation in the radial direction.

To make the current density constant, the amount of current that an ion macro particle represents is multiplied by the ratio of the differential area where the particle is initially located to the maximum differential area size on the curve. This makes a particle's size proportional to the area where it is located. For example, particle sizes are greatly reduced near the centerline where there is typically a high particle density. Figure 5.23 shows the first 1000 ion macro particles released from the fixed surface with and without weighting applied. The area of each particle is proportional to the amount of current it carries. With weighting applied, the area of particle coverage should be approximately uniform in this view.

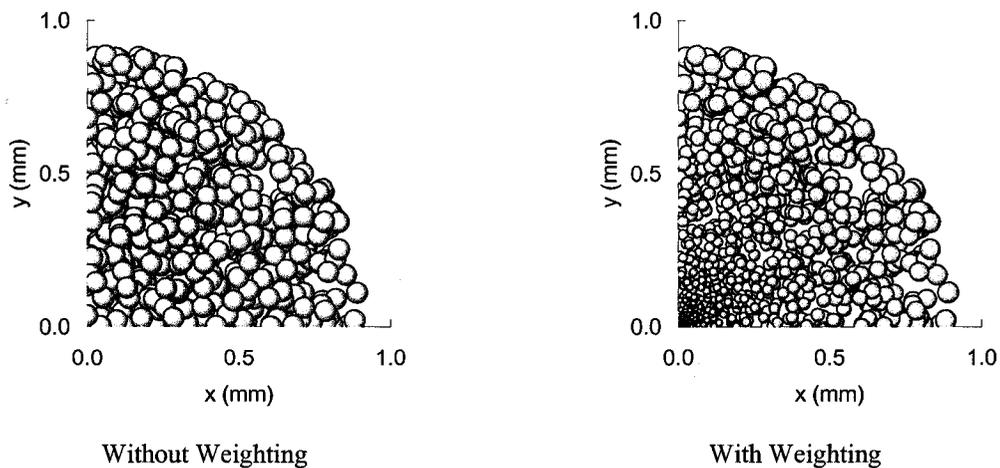


Figure 5.23 Ion macro particle size weighting with differential surface area to obtain constant current density.

Figure 5.24 shows an impingement limit curve comparing the crossover and perveance limits found using the conventional grids and the fixed sheath. This particular choice of the sheath location prevented the crossover limit from occurring, and also extended the perveance limit.

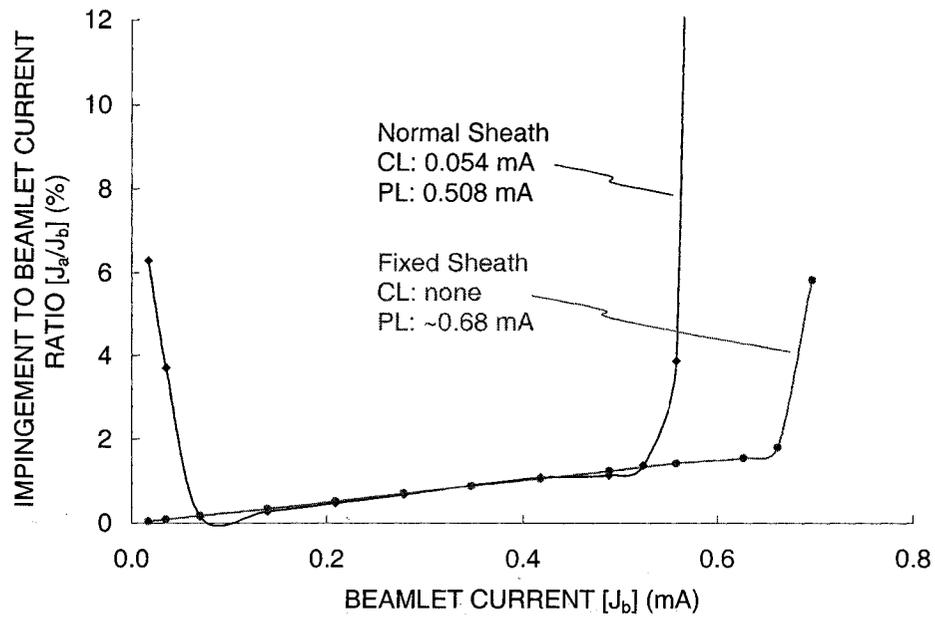


Figure 5.24 Improvement in the crossover and perveance limits using a fixed sheath.

Figure 5.25 compares the beamlet shapes found using the conventional screen-accel grid setup and those found using the fixed sheath surface. The screen grid in the fixed sheath case is shown only to reference the fixed sheath location. With regard to the crossover limit especially, the benefit of keeping the ion release surface far downstream at low beamlet currents is evident.

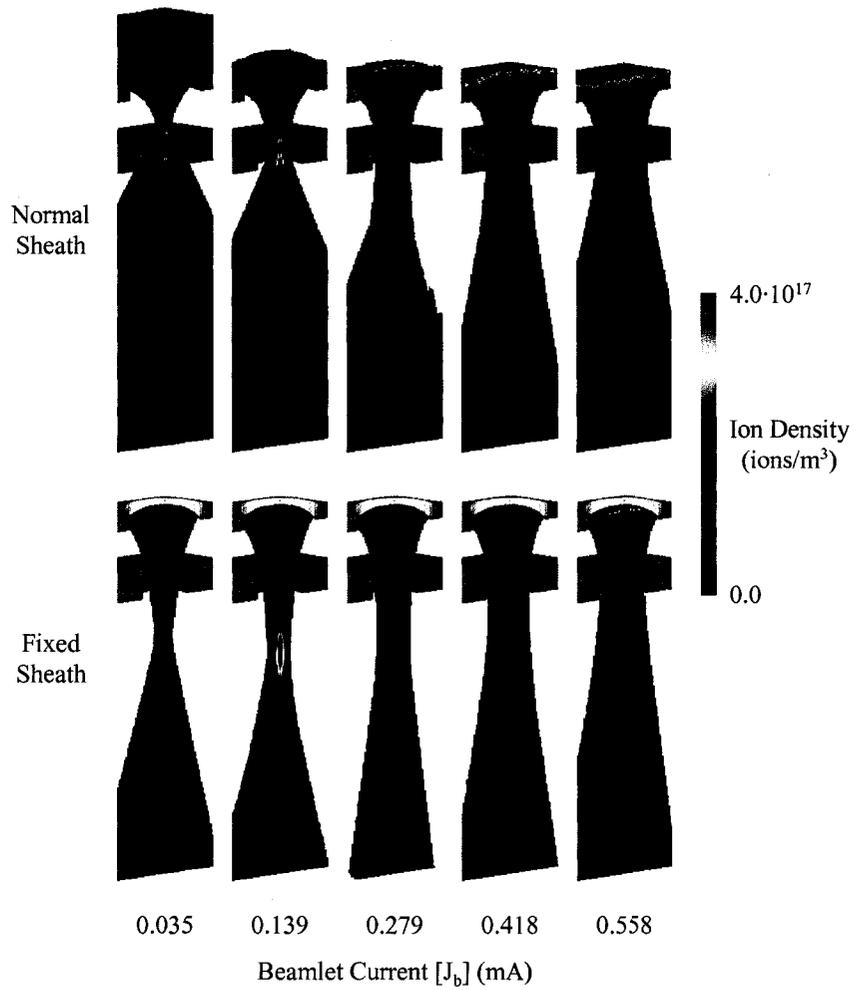


Figure 5.25 Beamlet shapes found using the conventional grids compared to those found using the fixed sheath.

Another possible goal when designing the fixed sheath surface might be to improve the divergence angles of the beamlets for example. In this case, the sheath surface produced beamlets with much lower divergence angles, shown in Figure 5.26.

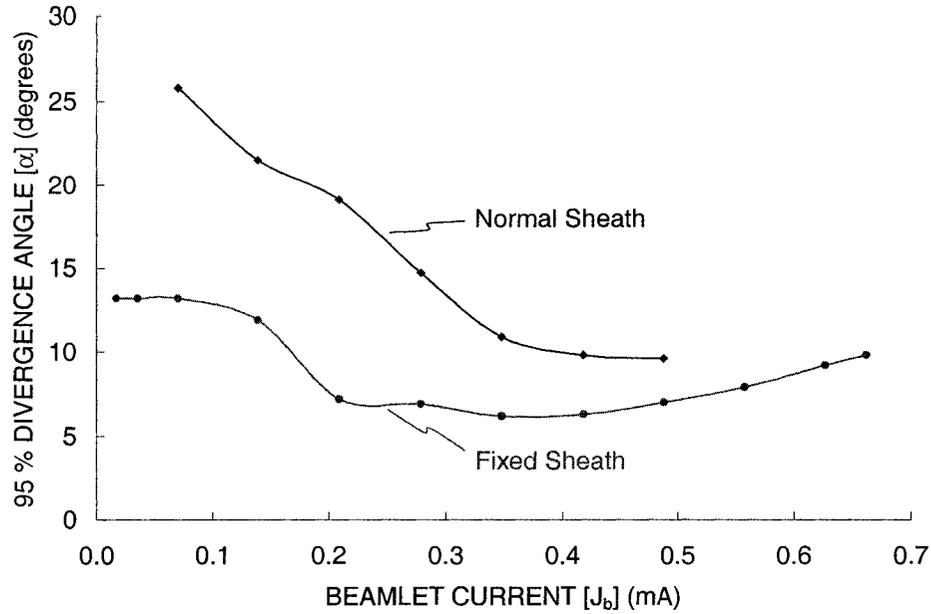


Figure 5.26 Improvement in beamlet divergence as a result of the choice for the fixed sheath.

5.7 Thruster Current Density

It is difficult in practice to measure the beamlet currents of individual apertures over the face of a thruster. Instead, current density profiles are usually taken at various locations downstream of the thruster and the beamlet currents are inferred from those data. Usually, this is done by first extrapolating the current density back to the thruster face and then simply multiplying the current density by the area of an aperture. A program was developed to try to more intelligently determine the beamlet current variation along the face of the thruster given a current density profile measured at some downstream location.

One way to describe a beamlet is to model it with an equation that describes the current density at any location downstream of the aperture. Eq. 5.8, similar in form to that used by Reynolds, gives the current density variation downstream of an aperture as a function of distance, l , and angle, α , to any point, P. In determining the six constant values used in this equation (A_1 , λ_1 , n_1 , A_2 , λ_2 , n_2) for a particular aperture, spherical geometry is assumed such that the integral of the current density, j_b , over a hemisphere downstream of the aperture results in the beamlet current, J_b . This calculation is shown in Eq. 5.9.

$$j_b(l, \alpha) = \frac{A_1}{l^2} e^{-\lambda_1(1-\cos \alpha)^{n_1}} + \frac{A_2}{l^2} e^{-\lambda_2(1-\cos \alpha)^{n_2}} \quad \text{Eq. 5.8}$$

$$\begin{aligned}
J_b &= \int_0^{2\pi} \int_0^{90^\circ} j_b(l, \alpha) l^2 \sin(\alpha) d\alpha d\theta \\
&= 2\pi \int_0^{90^\circ} \left(A_1 e^{-\lambda_1(1-\cos\alpha)^{n_1}} + A_2 e^{-\lambda_2(1-\cos\alpha)^{n_2}} \right) \sin(\alpha) d\alpha
\end{aligned}
\tag{Eq. 5.9}$$

The current density equation models all of the ions as emanating from a single point on the beamlet centerline. A flat probe that is not looking directly at this point will not collect all of the current predicted by the equation. A second angle, β , is used to describe the angle between the probe face and the line that connects the aperture to the probe. If the probe differential area is dA , the effective area of the probe is reduced to $dA \cos \beta$. Figure 5.27 shows the angles α and β . The angle α does not depend on the probe orientation, and α does not equal β in the general case.

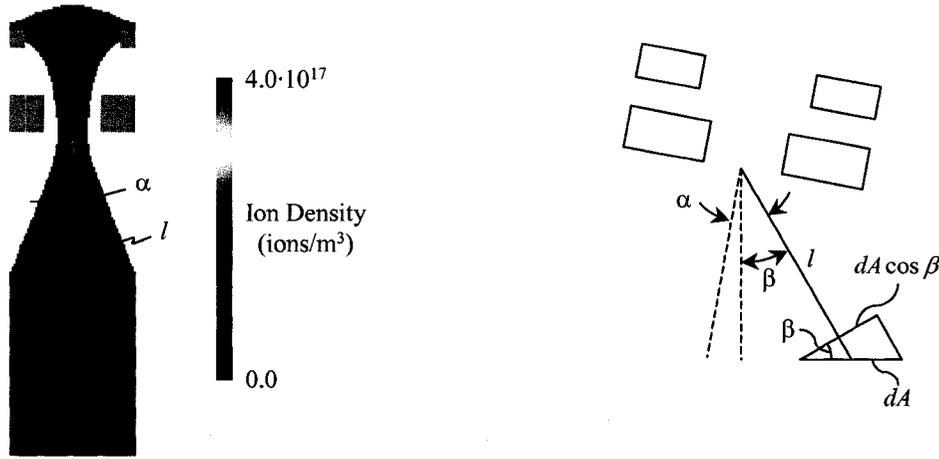


Figure 5.27 Current density at any point P found knowing the distance to the point, l , and the angle to the point from centerline, α . $J_b = 0.209$, $f_p = 0.3$.

The current density that a probe will detect from an individual aperture is given by Eq. 5.10. The total current density measured by the probe at a certain location is a summation over all apertures on the thruster face that contribute significant current, as given by Eq. 5.11.

$$j_b(l, \alpha, \beta) = \left(\frac{A_1}{l^2} e^{-\lambda_1(1-\cos\alpha)^{n_1}} + \frac{A_2}{l^2} e^{-\lambda_2(1-\cos\alpha)^{n_2}} \right) \cos \beta
\tag{Eq. 5.10}$$

$$j_{b \text{ Probe}} = \sum_i^{\text{Total Apertures}} \{j_b(l, \alpha, \beta)\}_i
\tag{Eq. 5.11}$$

There are three possible thruster configurations to consider: a thruster with flat grids, a thruster with grids curved downstream, and a thruster with grids curved upstream. These are shown in Figure 5.28. Spherical geometry is assumed in the two curved grid cases. R_C is the radius of curvature of the grids, R_T is the radius of the area of the thruster face with apertures (measured along the chord instead of the arc

length), and L_0 is the distance from the center of the grids to the plane where the current density measurements are made.

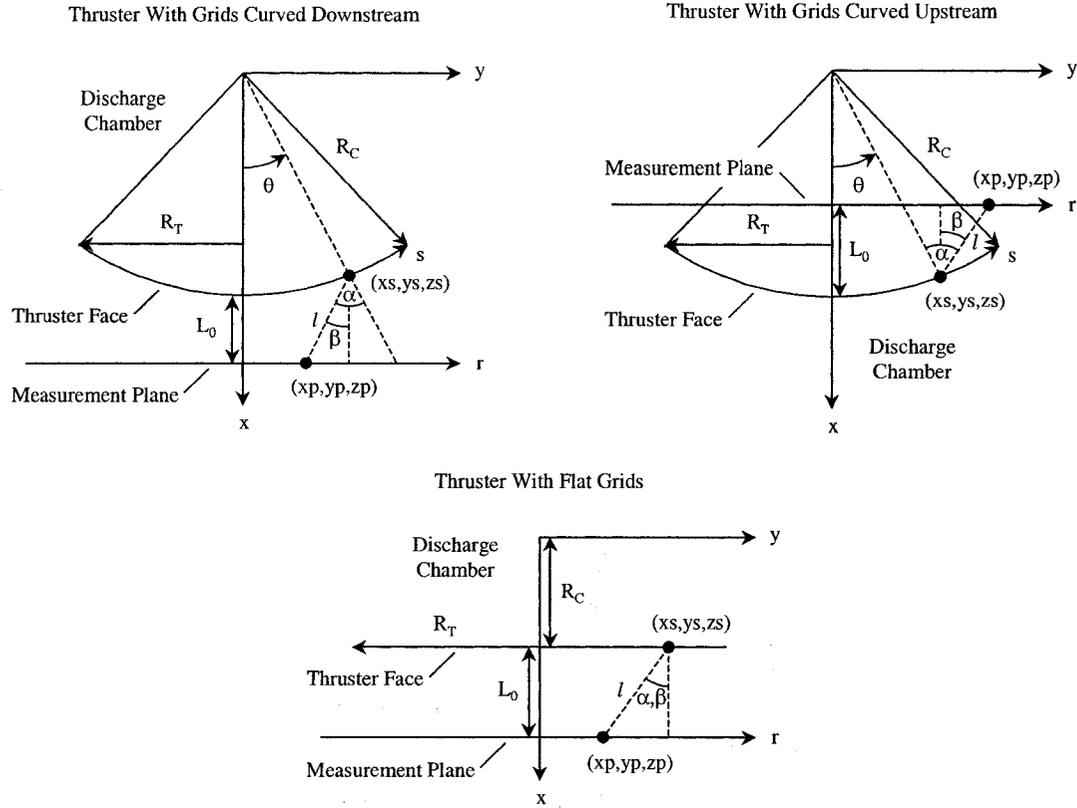


Figure 5.28 Variable definitions for three possible thruster configurations.

If the vector \vec{u} is defined as the vector from the origin to the aperture, and the vector \vec{v} is the vector from the aperture to the probe location, the angle α and the length l are given by Eq. 5.12. Eq. 5.13 gives the angle β .

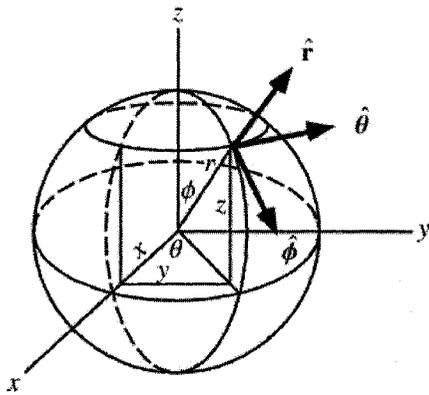
$$\begin{aligned}\vec{u} &= (x_s, y_s, z_s) \\ \vec{v} &= (x_p - x_s, y_p - y_s, z_p - z_s) \\ \alpha &= \cos^{-1} \left(\frac{u_x v_x + u_y v_y + u_z v_z}{|\vec{u}| |\vec{v}|} \right)\end{aligned}\tag{Eq. 5.12}$$

$$l = |\vec{v}|$$

$$\beta = \cos^{-1} \left(\frac{|x_p - x_s|}{l} \right)\tag{Eq. 5.13}$$

Figure 5.29 shows the spherical geometry used for the cases where the grids are curved upstream or downstream. The center of the grid set was taken to be along the x -axis, where $\theta = 0$ and $\phi = 90^\circ$. The hole layout was assumed to be hexagonal, where the rows of holes run along lines of constant ϕ . In other

words, while ϕ is held constant, θ is varied by an amount $\Delta\theta$, such that the arc length change is equal to the hole center-to-center spacing, l_{cc} . The rows of apertures are separated by $\Delta\phi$, and odd rows are offset by $(\Delta\theta)/2$. The location of the center of an aperture on the grid surface (x_s, y_s, z_s) can then be easily found using the transformations from spherical to rectangular coordinates listed in the figure.



$$\Delta\theta = \frac{\Delta s}{R_C} = \frac{l_{cc}}{R_C}$$

$$\Delta\phi = \frac{\sqrt{3} l_{cc}}{2 R_C}$$

$$x_s = R_C \cos \theta \sin \phi$$

$$y_s = R_C \sin \theta \sin \phi$$

$$z_s = R_C \cos \phi$$

Figure 5.29 Spherical geometry for the cases where the thruster grids are curved downstream or upstream.

For the case where the grids are curved upstream, the measurement plane is placed upstream of the thruster face such that the value L_0 is negative. Also, the ion beamlets are taken to be directed toward the origin. Slightly different equations are used for the flat grid case. The grid radius of curvature (R_C), for example, is simply taken as the distance from the origin to the thruster face, and can be set to an arbitrary value.

A stand alone program was written to curve fit the six unknown constants in the current density equation to the ion trajectory distribution of a beamlet. This program uses the Levenberg-Marquardt method, which is the standard in nonlinear least-squares curve fitting [Press]. An example result of the curve fitting program is shown in Figure 5.30 for a beamlet current of 0.209 mA.

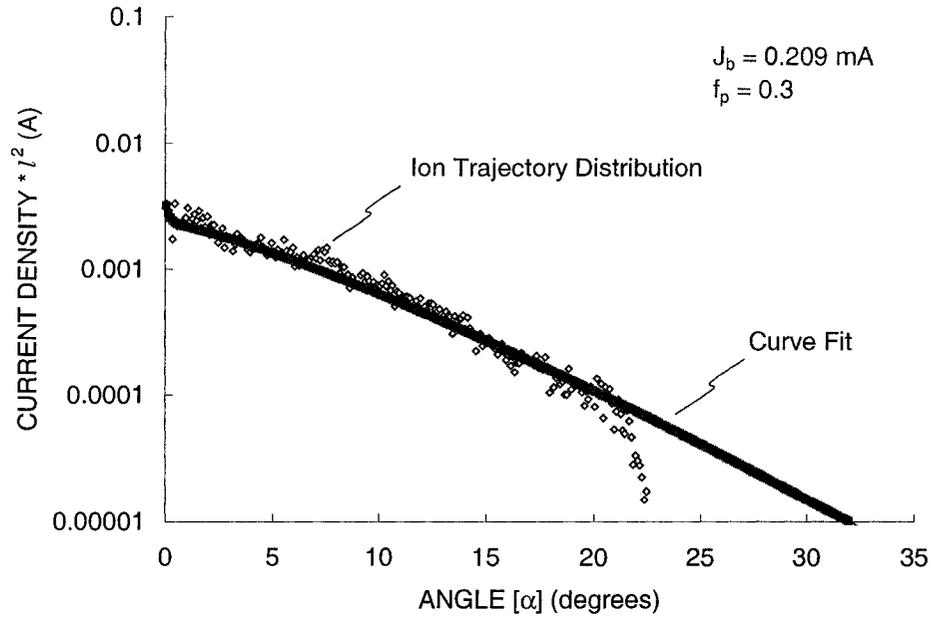


Figure 5.30 The current density equation fit to the ion trajectory distribution of a beamlet.

The curve fitting program was used to find constants for several beamlet currents, as given in Table 5.1. The graphical representations of these curves are given in Figure 5.31.

Table 5.1 Values for the constants used in the current density equation.

J_b (mA)	f_p (-)	A_1 (A)	λ_1	n_1	A_2 (A)	λ_2	n_2
0.035	0.05	0.000150	23	0.94	0.000099	500	0.68
0.070	0.10	0.000197	57	1.38	0.000173	500	1.34
0.139	0.20	0.001624	26	0.72	0.007803	500	0.50
0.209	0.30	0.002320	18	0.62	0.001153	500	0.50
0.279	0.40	0.005284	127	1.01	0.004618	500	0.50
0.348	0.50	0.012676	500	0.50	0.012676	500	0.50
0.418	0.60	0.012681	126	0.91	0.006608	500	0.50
0.488	0.70	0.009398	500	1.32	0.004402	500	0.50
0.523	0.75	0.009080	500	1.35	0.004578	500	0.50

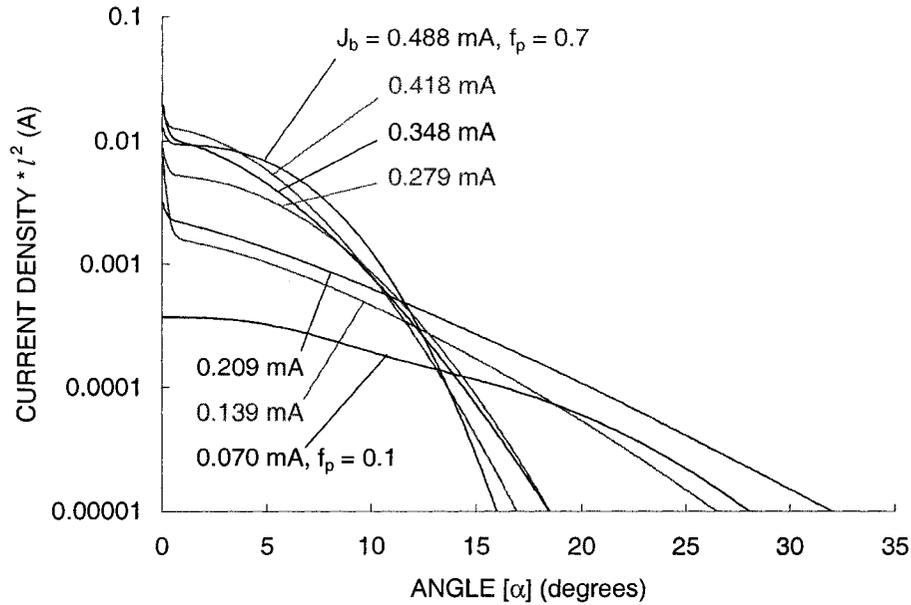


Figure 5.31 Current density equation curve fits as a function of beamlet current.

For these examples, the thruster radius, R_T , was taken to be 10 cm, which results in 6697 apertures in a flat grid. First, all apertures on the thruster were assigned an equal beamlet current, J_b , of 0.209 mA ($f_p = 0.3$). Figure 5.32 shows current density profiles (large, solid symbols) that would be measured at different locations downstream of the thruster. Near the grids, where L_0 is 5 cm, the beam profile extends only slightly beyond the edge of the thruster. Moving downstream, the profile broadens. Note that no beam attenuation as a result of ion scattering or charge exchange is accounted for.

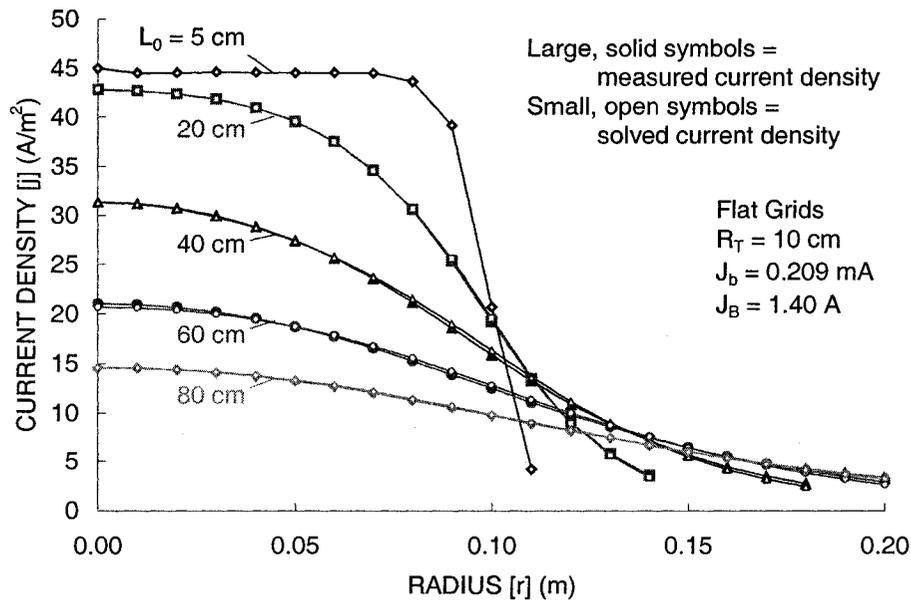


Figure 5.32 Current density profiles downstream of a flat grid set.

Rather than already knowing the beamlet current profile, the goal of the current density program is to determine the unknown beamlet currents based on known current density measurements. The code accepts as input a) a measured current density profile at some downstream location, b) a set of constants for the beamlet profiles, c) an initial guess for the beamlet current variation along the thruster face, and d) the thruster geometry. The code goes through an iterative process to reverse calculate the beamlet current profile. First, the code uses the beamlet current profile to predict the current density at the measurement plane. Then, the code compares the predicted current density profile with the measured current density profile and updates the beamlet current profile to make a better prediction.

The current density profiles obtained previously using a constant beamlet current, 0.209 mA, profile are now taken to be “measured” current density profiles to see if the code can correctly determine the original constant beamlet current profile. Referring back to Figure 5.32, the small open symbol curves are the current density profiles solved for by the current density program. At all downstream distances, the current density profiles match well.

Figure 5.33 shows the beamlet current profiles that correspond to the current density measurements of the previous figure. Ideally, all of the beamlet current profiles should be constant, at 0.209 mA. Close to the grid, where L_0 is 5 cm, the beamlet current profile is correct. With the current density measurements further downstream, the beamlet current profiles diverge from being constant even though the current density predictions match the measured current density values well.

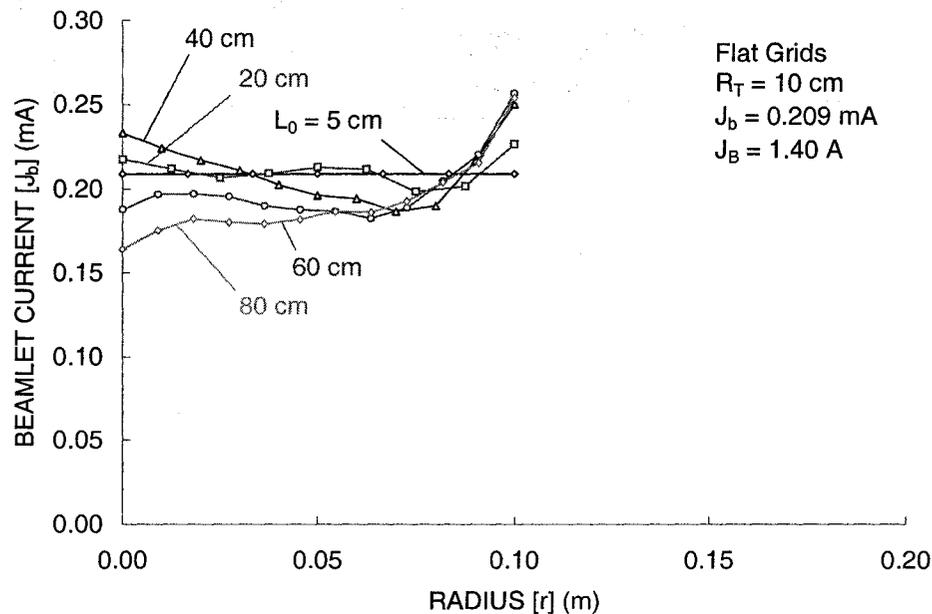


Figure 5.33 Beamlet current profiles corresponding to current density measurements.

Although it is not shown here for brevity, two things can be done to make the beamlet current profiles match correctly. First, better initial guesses for the beamlet current profile can be made. Choosing the initial beamlet current profile to be a constant 0.209 mA results in correct beamlet current profiles.

Second, in a real thruster the beamlet current profile usually decreases with increasing radius. Forcing the beamlet currents to decrease in moving toward the edge of the thruster, in this case, brings the beamlet current profile into perfect agreement.

Finally, current density profiles at three downstream locations are shown for a linearly varying beamlet current profile across the thruster face. This situation is much more realistic than the constant beamlet current cases shown previously.

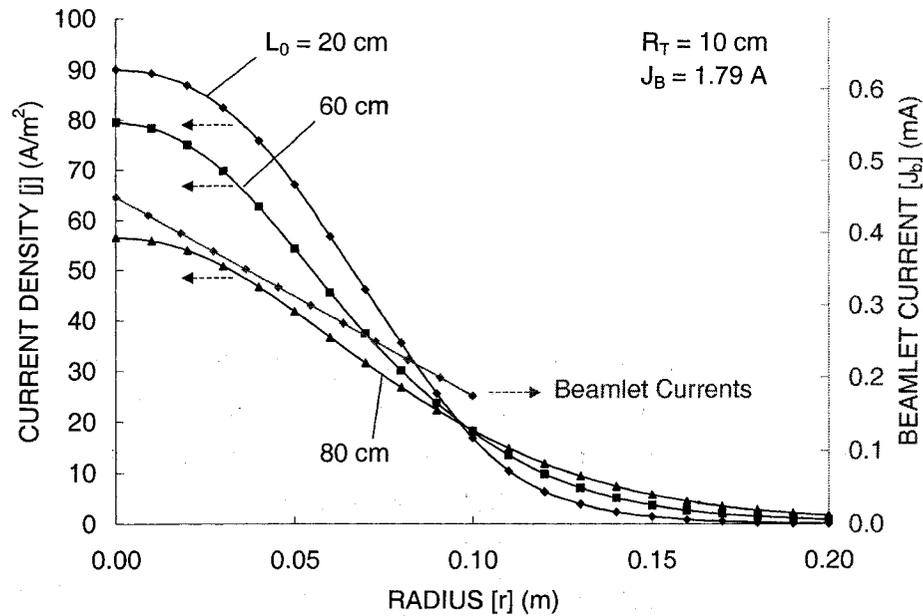


Figure 5.34 Current density profiles with a variable beamlet current profile.

6 Grid Design

The grid geometry parameters discussed previously have complex interactions, which is why designing a grid set is not a simple, straightforward process. An individual parameter, such as the grid spacing for instance, can be varied to find its affect on things like the ion transparency, beamlet divergence, and operating range pretty easily given a fixed set of values for the rest of the dimensions. Trying to vary several parameters at the same time to find optimum operating points is more difficult.

One tool that can be used to search for solutions to problems with many variables is an evolutionary algorithm. Evolutionary algorithms work by applying evolutionary concepts to problems that have been translated into the biological realm. Information concerning evolutionary algorithms was compiled from the publications by Buseti, Obitko, Pohlheim, Srinivas, and Whitley.

6.1 Evolutionary Algorithms

Evolutionary algorithms, in general, attempt to solve problems using the ideas of evolution. Two of the main sub-categories of evolutionary algorithms are genetic algorithms and evolution strategies. Two additional sub-categories related to computer programming and machine learning are genetic programming and evolutionary programming.

Figure 6.1 shows the general outline of what an evolutionary algorithm does. The first step is to select an evolutionary algorithm appropriate to the problem. The next step is to define a set of possible solutions to the problem, called chromosomes or genomes. Then, those possible solutions are evaluated to determine their effectiveness at solving the problem. In doing this, each solution is assigned a fitness value. Then the ideas of selection, crossover, and mutation are used to advance the set of possible solutions to the next generation. After many generations, the set of possible solutions will hopefully converge to a single, best, solution.

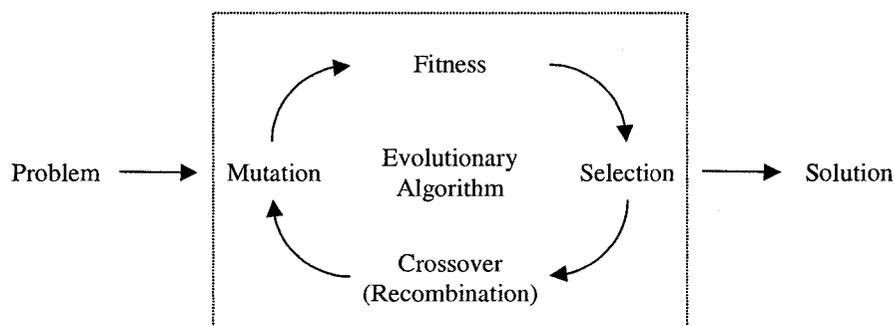


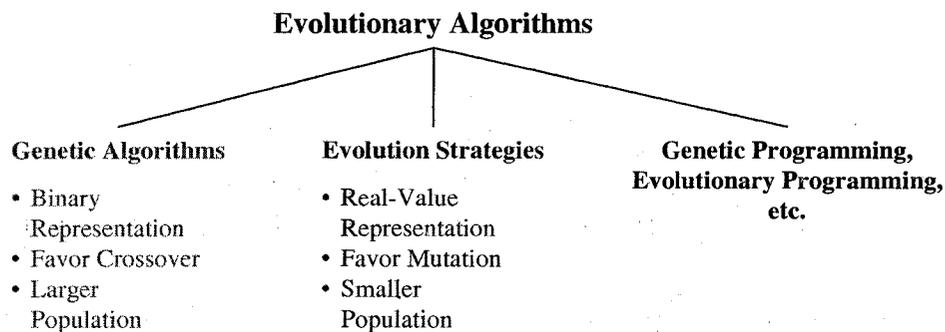
Figure 6.1 Outline of an evolutionary algorithm.

Genetic Algorithms & Evolution Strategies

Genetic algorithms are perhaps the most widely used evolutionary algorithms. Genetic algorithms have been primarily developed in the United States, starting with the work of John Holland in the 1960's.

Evolution strategies have been developed alongside, but fairly independently of, genetic algorithms. Specifically, evolution strategies were first developed through the work of Ingo Rechenberg and Hans-Paul Schwefel in Germany.

Both genetic algorithms and evolution strategies share the same principles of evolutionary algorithms. Genetic algorithms generally have a wider area of application than evolution strategies. Genetic algorithms have been used for search, design, machine learning and optimization problems, whereas evolution strategies are mainly used for optimization problems. Genetic algorithms tend to favor binary encoding of the solutions rather than real-value encoding, crossover rather than mutation of the chromosomes, and tend to have larger solution populations than evolution strategies.



Benefits of Genetic Algorithm and Evolution Strategies

Genetic algorithms and evolution strategies are known as “weak” methods because they try not to make many assumptions about the problem in question. Genetic algorithms and evolution strategies are methods that can be used to search for a global solution to a problem without the use of gradient information.

- Evolutionary algorithms search through the solution space using many parallel solutions rather than a single solution.
- Evolutionary algorithms do not require any derivative type information about the problem.
- Evolutionary algorithms use probabilities to influence the search direction. In this way, paths that look bad at a certain step, when in reality might lead to good solutions, can remain open when they otherwise might have been closed by a more absolute search method.

Example Problem

An example problem is useful for illustrating the principles of evolutionary algorithms. In this example, one might wish to optimize the ion thruster grid geometry in order to achieve the longest possible grid lifetime for a certain operating voltage and current range. In this case, the parameters that might be

varied are the grid hole center-to-center spacing (l_{cc}), screen grid thickness (t_s), screen grid hole diameter (d_s), grid-to-grid spacing (l_g), accel grid thickness (t_a), accel grid hole diameter (d_a), and accel grid voltage (V_a).

Chromosomes

By the evolutionary analogy, each possible solution to the problem in question is called a chromosome, or genome. Each chromosome somehow describes a value for all of the parameters that vary in the problem. Each value is by analogy called a gene, and is encoded onto the chromosome in a certain location.

In this example, each chromosome will give a value for the hole center-to-center spacing (l_{cc}), screen grid thickness (t_s), etc.

Encoding

There are many ways in which chromosomes are typically constructed. The method of encoding values onto the chromosome is often influenced by the problem itself. Binary and real-value encoding are common to genetic algorithms and evolution strategies. Permutation encoding is used for more specific problems, such as the traveling salesman problem. Tree-encoding, not discussed here, is used in genetic and evolutionary programming.

Binary Encoding

The most traditional way of placing solution values onto the chromosome is through binary value encoding. In binary encoding, each parameter is represented by a binary bit string of a certain length. This is the traditional way that genetic algorithms encode values.

In the example problem, the entire chromosome would be 56 bits long if each of the seven parameters were assigned 8 bits of resolution. An example chromosome is shown in Figure 6.2.

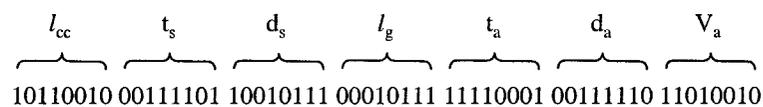


Figure 6.2 Chromosome containing genes that code for parameter values.

Real-Value Encoding

A second common way of encoding values is called real-value encoding. In this encoding scheme, each chromosome is simply the collection of the actual parameter values. This encoding method is traditionally favored in evolution strategies.

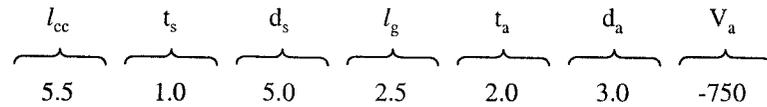


Figure 6.3 An example of real-value chromosome encoding.

Permutation Encoding

Permutation encoding has applications in certain types of problems. Specifically, the problem commonly known as the traveling salesman problem and other task ordering problems use permutation encoding. The traveling salesman problem is the problem of finding the shortest path among many different places that the salesperson must visit. Each chromosome, for example, might have a list of numbers that describe the order of cities to be visited.

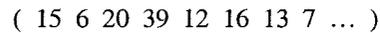


Figure 6.4 An example of permutation encoding for a traveling salesman type problem.

Population

In genetics, there are many chromosomes that exist at the same time, and the complete set of chromosomes in a generation makes up the gene pool. Essentially, the population of chromosomes in a generation are all the possible solutions to the problem. Each problem is different, but typically there might be around 30 chromosomes in the population in each generation. On the first generation, all of the values in all chromosomes are assigned random values as a starting point.

Fitness Values

There must be a way to evaluate how well each chromosome meets the solution. This is called finding the “fitness value” of the chromosome. The fitness value will be used to determine the probability that the chromosome will reproduce to create offspring in the next generation of the algorithm.

In the example problem, if the goal of the problem is to find a grid set with the maximum lifetime, the fitness value of the chromosome might be the predicted operation time until the accel grid loses half of its original mass. The ffx code can be used to test each chromosome at several operating conditions and assign each chromosome in the population a fitness value. For example, if the population is comprised of ten chromosomes, the ffx code will evaluate each chromosome to determine the predicted accel grid lifetime and give each a fitness value. The fitness values for this example are listed in Table 6.1.

Table 6.1 Example fitness values for ten chromosomes labeled 0 through 9.

Chromosome	Fitness f_i	Normalized Fitness f_i/f_A
0	5	1.0
1	0	0.0
2	3	0.6
3	12	2.4
4	6	1.2
5	2	0.4
6	4	0.8
7	7	1.4
8	1	0.2
9	10	2.0
Sum of f_i s:	50	
Average (f_A):	5	

An example fitness value function for the example problem is given in Eq. 6.1. Here, the fitness of chromosome i is given by the grid transparency to ions, ϕ_i , multiplied by one minus the divergence angle, α , multiplied by the predicted grid lifetime, L .

$$f_i = \phi_i \cdot \left(1 - \frac{\alpha}{90^\circ}\right) \cdot L \quad \text{Eq. 6.1}$$

There are many possible choices for the fitness function. For instance, one could desire to minimize the beam divergence angle only, or one could increase the effect of the predicted lifetime by taking it to a power. The fitness function will change with the desired problem solution.

Creating a New Population

Once the fitness values of each chromosome in the population are determined, the next generation of chromosomes can be formed. In this process, the new set of chromosomes will ideally be better at solving the problem than the previous generation of chromosomes. The chromosomes in the current generation are called the “parent” chromosomes, and the chromosomes formed from those chromosomes and placed in the next generation are called the “child” chromosomes.

The process of forming the chromosomes in the next generation involves a) the “selection” of parent chromosomes for reproduction, b) combining the parent chromosomes into child chromosomes through “crossover”, and c) random “mutation” of the child chromosomes. There are many different ways in which to perform these steps in evolutionary algorithms.

Selection

Generally, as is the case in the algorithm developed here, two parent chromosomes are selected for reproduction and two child chromosomes are produced from those chromosomes. The parent chromosomes

are selected according to their fitness values. Here, chromosomes with greater fitness values meet the solution better than chromosomes with smaller fitness values. The chromosomes with greater fitness values are therefore more likely and better candidates to produce child chromosomes, hence the “survival of the fittest” component of the evolution process.

In the example problem, chromosome 3 would have the best chance at being selected to be a parent chromosome because it has the greatest fitness value. Accordingly, chromosome 1 is the least likely to be selected to be a parent.

There are many techniques used to select parent chromosomes for reproduction. These include roulette wheel selection, stochastic sampling, and tournament selection.

Roulette Wheel Selection

In roulette wheel selection, chromosomes are selected for reproduction based on the magnitude of their fitness values. The fitness values can be thought of as being laid out around a wheel, with the portions of the wheel allocated to the chromosomes based on the sizes of the fitness values, as shown in Figure 6.5. To select a parent chromosome, a random number is generated from zero to the sum of the fitness values, and the random number will then point to a parent chromosome. Chromosome 0, for example, has a 5/50 chance (10%) of being selected for reproduction. A second random number is generated to choose the second chromosome for reproduction.

Chromosome	Fitness f_i	Normalized Fitness f_i/f_A	Selection Probability %
0	5	1.0	10
1	0	0.0	0
2	3	0.6	6
3	12	2.4	24
4	6	1.2	12
5	2	0.4	4
6	4	0.8	8
7	7	1.4	14
8	1	0.2	2
9	10	2.0	20
Sum of f_i s:	50		
Average (f_A):	5		

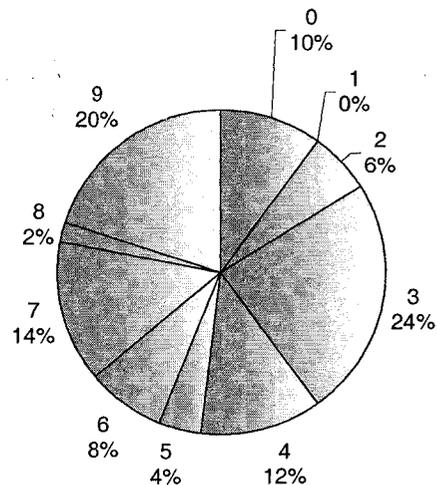


Figure 6.5 Roulette wheel selection.

Stochastic Sampling

Stochastic sampling is often used instead of roulette wheel selection because it enhances the likelihood that the chromosomes selected for reproduction will agree with the distribution of fitness values.

Often in the selection process, an intermediate chromosome population is created between the current generation and the next generation. Chromosomes are selected from the parent generation and placed in the intermediate population. Then, only chromosomes in the intermediate population can be selected for reproduction to place offspring in the next generation.

Chromosomes that have normalized fitness values, f_i/f_A , greater than 1.0 are above average chromosomes, while those that have normalized fitness values less than 1.0 are below average chromosomes. One way to create the intermediate population is to use “remainder stochastic sampling.” In using this method, the normalized fitness values are broken into an integer part and a remainder part. The integer part instructs that many copies of the chromosome to be placed directly into the intermediate population, and the remainder becomes the probability that an additional copy of that chromosome will be placed in the intermediate population.

For instance, if a chromosome has a f_i/f_A value of 1.2, the integer part is 1.0 and the remainder is 0.2. For this chromosome, 1 copy is placed directly into the intermediate population, and the chromosome then has an additional 20% chance of placing another copy of itself into the intermediate population. On average, this will place the correct number of chromosomes into the intermediate population. To achieve the same end result, roulette wheel selection can be used with the sum of the remainders, after the integer copies have been placed, to fill in the rest of the intermediate population.

Something called “stochastic universal sampling” can be used to achieve remainder stochastic sampling reliably. In this method, the f_i/f_A values can be laid out end to end along a line, or around a wheel. The sum of all the f_i/f_A values is equal to the number of chromosomes. An equally spaced line, with segments of length 1 and a total length of one minus the number of chromosomes, is placed alongside the line of f_i/f_A values. This is a “pointer” line, where each of the pointers indicates a chromosome which will be placed into the intermediate population. The entire line is offset by a random number, k , that is generated from zero to one. This method still places the integer number of chromosomes into the intermediate population because the line segments should never skip a chromosome with a length greater than 1 along the line. Figure 6.6 shows an example of stochastic universal sampling. In this particular case, the choice of k dictated that three copies of chromosome 3 would be placed in the intermediate population for example.

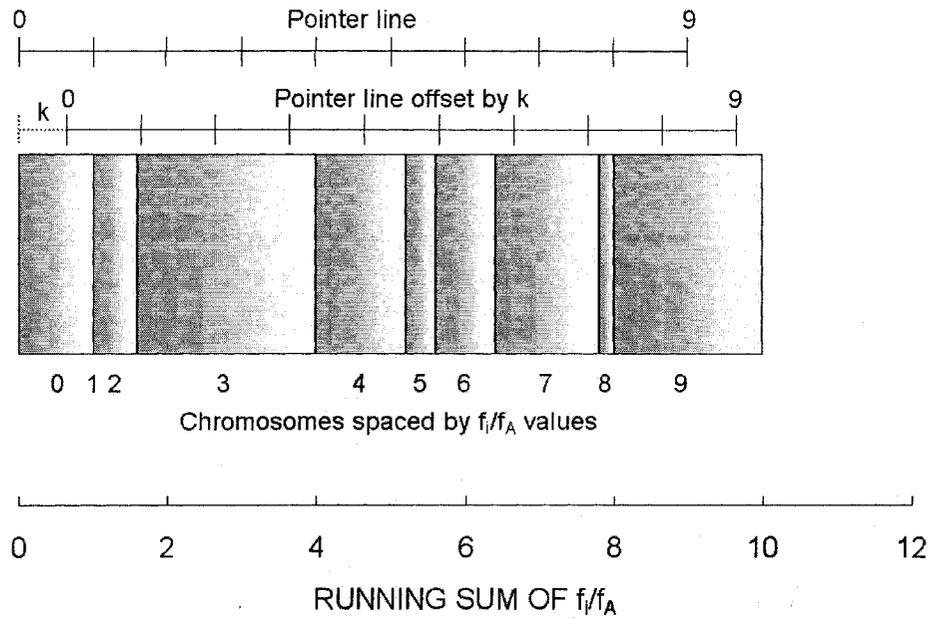


Figure 6.6 Stochastic universal sampling achieves remainder stochastic sampling.

Once the intermediate population has been created, pairs of chromosomes are selected at random for reproduction. Steps can be taken to randomize the intermediate population if desired. It is also common to prevent (incest) breeding among the exact same, or very similar, chromosomes within the intermediate population.

Tournament Selection

Another commonly used selection method is tournament selection. In this method, small groups of chromosomes are randomly selected to participate in a tournament. The winner of this small tournament, the chromosome with the greatest fitness value, will be placed in the intermediate population of chromosomes. For instance, if the tournament size is 2, two chromosomes are selected and the one with the greater fitness value gets placed into the intermediate population. Using larger tournament sizes has the effect of further weeding out the chromosomes with lower fitness values because they are less likely to win the larger tournaments.

Crossover

Once two parent chromosomes are selected for reproduction, crossover takes place to start the process of producing child chromosomes. The term “crossover” tends to be associated with binary value representations, and the term “recombination” tends to be associated with real-value representations, but they refer to the same process. There are many different ways that crossover is implemented. Common methods of performing crossover for both binary and real-value representations are one-point crossover, two-point crossover, uniform crossover and discrete crossover. Most of these methods will be described

with binary encoded chromosome examples, but the extension to real-value encoded chromosomes should be clear. Intermediate recombination is described as applied to real-value encoding.

One-Point Crossover

One-point crossover involves selecting a point at random along the length of the chromosome and then switching the sections either before or after the selected point. Figure 6.7 shows an example of one-point crossover. Here, the crossover point was selected after bit four. For real-value encoding, one can simply imagine real number values in each of the chromosome positions instead of the single binary bits.

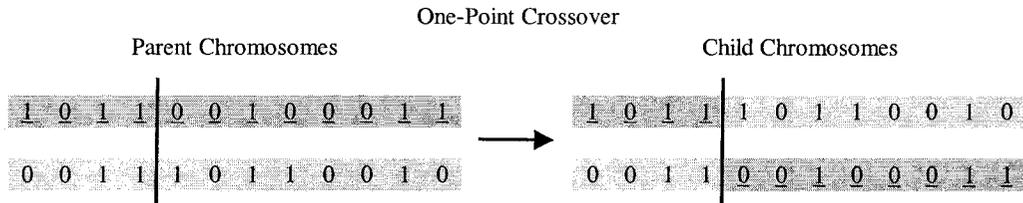


Figure 6.7 One-Point crossover applied to a binary encoded chromosome.

Two-Point Crossover

Two-point crossover is very similar to one-point crossover, the only difference being that two locations along the chromosome are chosen instead of one. Figure 6.8 shows an example of two-point crossover where the split points have been selected after bit three and after bit nine. The section between the two split points are simply switched to create the new child chromosomes.

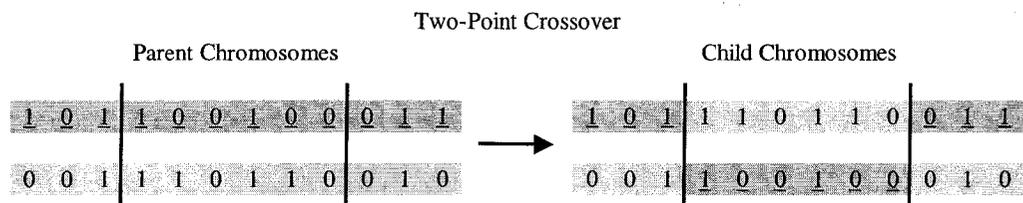


Figure 6.8 Two-point crossover applied to a binary encoded chromosome.

Uniform Crossover

Uniform crossover introduces greater randomness than either one-point or two-point crossover. With uniform crossover, a random bit mask is generated for each pair of chromosomes. The bit mask indicates which bit is taken from which parent. In essence, this is an extension of one-point and two-point crossover, where many crossover points are selected. Figure 6.9 shows an example of uniform crossover. Here, it can be seen that roughly half of each parent chromosome is given to each child chromosome.

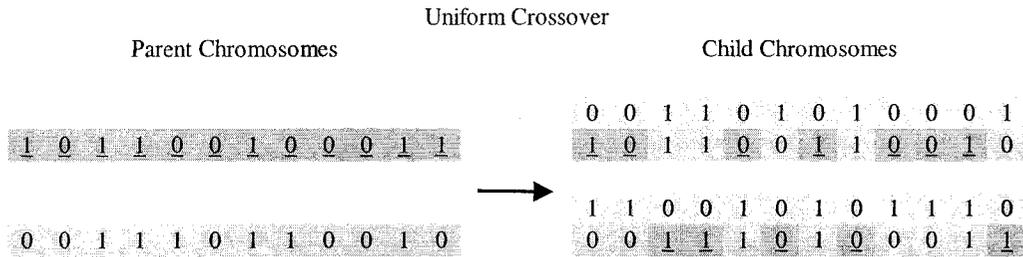


Figure 6.9 Uniform crossover applied to a binary encoded chromosome.

When two chromosomes are very similar, one-point and two-point crossover can skip entire sections of the different parts of the chromosomes. Therefore, uniform crossover (and discrete crossover) is normally better at producing new offspring than one-point or two-point crossover.

Discrete Crossover

Discrete crossover is very similar to uniform crossover. When using discrete crossover, a bit is randomly selected from one of the two parent chromosomes to fill each position in the child chromosomes. This is in slight contrast to uniform crossover, where a single bit mask is used to create the two child chromosomes. Discrete crossover essentially uses two randomly generated bit masks to form the child chromosomes. For instance, in the example shown in Figure 6.10, both child chromosomes took the bit value from the first parent chromosome in the first bit position. In the second bit position, both child chromosomes took the bit value from the second parent chromosome. It is noted that it really doesn't matter which parent chromosome is chosen in the second bit position, since both bits are zero. This is the case at several locations along the chromosome. In this particular case, it turns out that the child chromosomes only differ by a single bit.

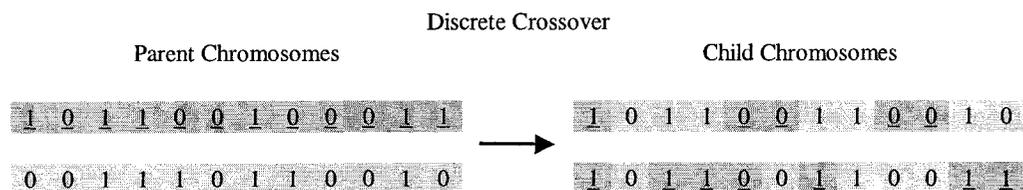


Figure 6.10 Discrete crossover applied to a binary encoded chromosome.

Intermediate Recombination

Intermediate recombination can be used with real-value encoding. The parameters of the child chromosomes are assigned through interpolation among the two values of the parent chromosomes using Eq. 6.2. A new random value of a is chosen for every parameter in the chromosome using the random number r provided by a random number generator. Using $d = 0.25$ statistically results in the domain of the children being as large as the domain of the parents.

$$v_c = v_{p1}a + v_{p2}(1-a)$$

where:

$$a = r + d(2r-1)$$

Eq. 6.2

$$d = 0.25, r = [0,1], \text{ and } a = [-d, 1+d]$$

A variation, called line recombination, results when the same value of a is used for all variables in creating a particular child chromosome. Standard intermediate recombination results when using $d = 0$.

Crossover Comments

One thing to notice in the crossover process is that chromosomes are split at random locations. These locations may fall in the middle of the genes, which code for each physical parameter. Figure 6.11 shows how the crossover points can change some parameter values and leave others the same. In this example, the values of the hole center-to-center spacing, l_{cc} , and screen grid thickness, t_s , are affected by the crossover operation, while the gene that codes for the screen grid diameter, d_s , is unchanged in the child chromosomes. In this particular case, both child chromosomes have different values for the hole center-to-center spacing and screen grid thickness than either of the parent chromosomes. This example serves to illustrate that the child chromosomes can share some of the same traits as their parents, but they can also have new traits altogether.

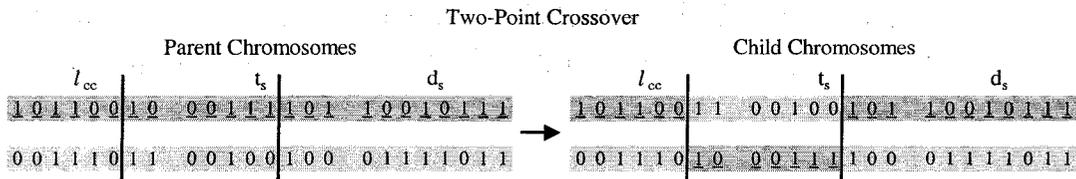


Figure 6.11 Example of two-point crossover indicating how genes can change.

Crossover in genetic algorithms generally occurs at a rate of 70 to 90%. In other words, when two parents are selected to be placed in the new generation, crossover will occur about 80% of the time. If crossover does not occur, child 1 will be an exact copy of parent 1 and child 2 will be an exact copy of parent 2. In that case, the parents survive as children in the next generation.

Mutation

The next process is the mutation of the child chromosomes. During the process of reproduction in genetics, there can be small errors that change DNA in a purely random way. There are many ways to model mutation. Often the mutation method is dictated by the problem in question.

Mutation Using Binary Encoding

One way to model mutation with binary encoding is to simply switch bits from 0 to 1 or from 1 to 0 at a small rate. Mutation usually occurs infrequently, and in many algorithms the rate at which mutation occurs is on the order of a few percent or less. One source suggests that a mutation rate of $1/b$ is a reasonable mutation rate, where b is the number of bits in the chromosome string. Using this rate has the effect of mutating, on average, a single bit in each child chromosome. Figure 6.12 shows an example of mutation using the two-point crossover example, where two bits were inverted in the first child chromosome, and one bit was inverted in the second child chromosome.

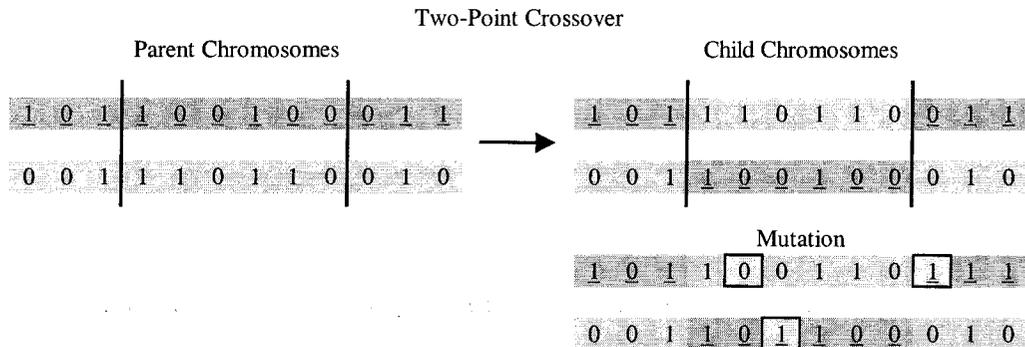


Figure 6.12 Mutation of the child chromosomes. A few bits at random points in each child chromosome are flipped at a small rate.

Mutation Using Real-Value Encoding

Mutation using real-value encoding typically involves adding (or subtracting) a small number to selected values in each child chromosome. Once again, the rate of mutation is suggested to be $1/v$, where v is the number of parameters that are in the chromosome. This has the effect of changing, on average, one value in each child chromosome. The amount that a value is changed is normally a small percentage of the range over which the particular value is allowed to vary. Additionally, it is desirable to make the probability of a small change greater than the probability of a large change.

Figure 6.13 shows one way to achieve these effects. The variable r is the possible amount by which a value could change in either direction, typically chosen as 10% of the total value domain or less. The variable a , which in turn depends on u and k , sets the actual value magnitude of change, where a varies from one to zero. From the plot, for a random choice of u , a is likely to be small, but it does have some probability of being closer to one. The variable s chooses which direction to move the value. Note that the smallest possible variable mutation is dictated by the choice of k , which determines the size of 2^{-k} . The choice of these variables can change during the algorithm to help convergence.

$$Value^{Mutated} = Value + s \cdot r \cdot a$$

$s = -1$ or $+1$, chosen at random

$r = R \cdot domain =$ mutation range,

$R \cong 10\%$ or smaller

$$a = 2^{-u \cdot k}$$

$u =$ uniform random number : $0 \leq u \leq 1$

$k =$ mutation precision, $k \cong 4 \dots 20$

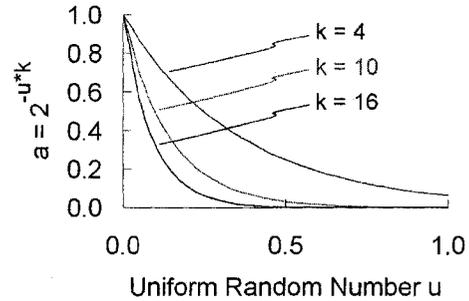


Figure 6.13 Mutation of real-value encoded values.

Adaptive Crossover and Mutation

Srinivas and Patnaik proposed one example of an adaptive crossover and mutation scheme intended to improve algorithm efficiency. A similar scheme is presented here. The idea in general is to use less crossover and only slight mutation with the better chromosomes in the population, while using heavy crossover and mutation with the poorer chromosomes. This tends to preserve and fine-tune the better chromosomes while still searching widely for better solutions using the less-fit chromosomes.

On any generation, the average population fitness value is f_{avg} , and the fitness value of the best chromosome is f_{max} . After selecting two chromosomes, the probability of crossover is dictated by the better of the two chromosomes (with fitness value f') using Eq. 6.3. A linear variation in probability is used if the chromosome fitness value is greater than the average fitness value, and a constant probability ($p_{c_{avg}}$) otherwise. $p_{c_{min}}$ and $p_{c_{max}}$ are the probabilities of the best and average chromosomes undergoing crossover, respectively. Mutation is similar, Eq. 6.4, except that the probability of mutation depends on each individual fitness value (f) rather than the better fitness value (f').

$$p_c = \begin{cases} p_{c_{min}} - (p_{c_{min}} - p_{c_{max}}) \frac{f_{max} - f'}{f_{max} - f_{avg}} & f \geq f_{avg} \\ p_{c_{avg}} & f < f_{avg} \end{cases} \quad \text{Eq. 6.3}$$

$$p_m = \begin{cases} p_{m_{min}} - (p_{m_{min}} - p_{m_{max}}) \frac{f_{max} - f}{f_{max} - f_{avg}} & f \geq f_{avg} \\ p_{m_{avg}} & f < f_{avg} \end{cases} \quad \text{Eq. 6.4}$$

Example adaptive crossover and mutation probabilities, found to work well, are shown in Figure 6.14 for a case where there are six variables per chromosome and real-value encoding is used. In this case, a selective pressure of 2.0 leads to average and maximum fitness values of 1.0 and 2.0, respectively. Most of the chromosomes selected for reproduction will come from the upper half of the population. If the best

chromosome is picked for reproduction, crossover will occur half of the time and one and one half variables will be mutated on average.

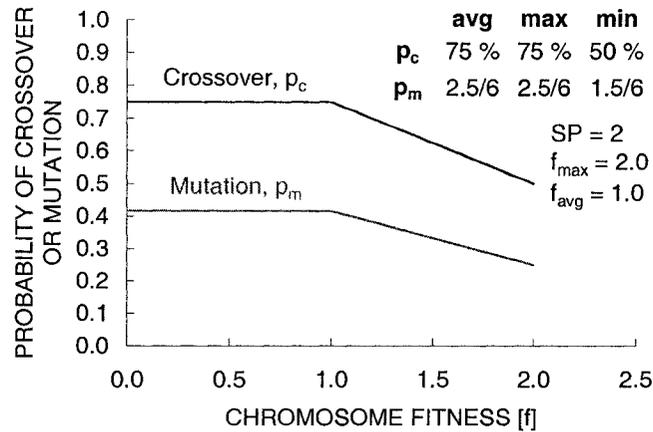


Figure 6.14 An adaptive crossover and mutation scheme.

Elitism

One other common practice is to use “elitism”. Elitism is when the best chromosome, or chromosomes, from the current population is simply copied directly into the next generation generally without modification. Using elitism ensures that the best chromosome from the current generation makes it into the next generation. The chromosome that was directly copied could be the best chromosome in the next generation as well as the current one. Ensuring that the best chromosome is in the next population makes it a candidate to be a parent in that generation.

Algorithm Execution

Once a generation of chromosomes is filled through a combination of elitism and crossover & mutation, the chromosomes are evaluated using the ffx code to determine their fitness values, and the process repeats until some pre-determined stopping criterion has been reached. Typically, good solutions to problems are not found until something on the order of 100 generations have passed. Convergence of an individual gene can be concluded when about 95% of the population shares the same value of that gene.

One problem that can arise is referred to as genetic drift. This is the effect that certain genes in the chromosome can converge to a single value, which may not be the best value. Through crossover alone, a gene that is the same in two parent chromosomes cannot change in going to the child chromosomes. In that case, mutation is the only way that the gene can change. However, if the mutation rate is set too high, the algorithm will start behaving more like a random search rather than searching with a purpose using the fitness values.

Scaling of the Fitness Values

In order to help the algorithm run smoothly, scaling is often applied to the chromosome fitness values to ensure that there is always a range of good, okay, and bad parent chromosome candidates. This relates to the idea of “selective pressure,” which simply means that there is always pressure to move toward a better solution.

If the spread of the fitness values is small, the algorithm will have trouble converging to one solution because all of the solutions look more or less the same. On the other hand, if the spread is too big, individual, super-fit, solutions can heavily dominate the algorithm and convergence to a certain solution can occur too quickly.

There are several ways to control the selective pressure among fitness values. One of the most common ways is rank selection. Another is called windowing.

Rank Selection

To ensure the spread of fitness values among the population is not too big nor too small, rank selection is often used. In rank selection, new fitness values are assigned that are based on each chromosome’s rank within the population. In this manner, the magnitudes of the fitness values only determine the chromosome rank order, which serves to reduce the influence of any super-fit chromosomes while at the same time making sure that there are a range of individuals in the population.

Linear ranking is performed through the use of an equation such as the one given in Eq. 6.5. Here, SP is the selective pressure, which varies from 1.0 to 2.0. Selective pressure is the probability of the best chromosome being selected for reproduction compared to the average probability of all chromosomes being selected for reproduction. A selective pressure of 1.0 results in all of the fitness values being 1.0, while a selective pressure of 2.0 results in the fitness values going from 0.0 to 2.0 linearly. In general, the fitness values will go from $2.0-SP$ to SP . The loss of diversity in Eq. 6.6 is the proportion of chromosomes that will not be selected for reproduction. With a selective pressure of 2.0, 75% of the chromosomes will be selected for reproduction, which means 25% of the population will not have a chance to reproduce.

$$Fitness_{LinearRank} = 2 - SP + 2(SP - 1) \frac{Position - 1}{N - 1}$$

$$1.0 \leq SP \leq 2.0, N = \text{Total Number of Chromosomes} \quad \text{Eq. 6.5}$$

Worst Individual, $Position = 1$

Best Individual, $Position = N$

$$Loss\ of\ Diversity_{LinearRank} = \frac{SP - 1}{4} \quad \text{Eq. 6.6}$$

With a selective pressure of 2.0, about 75% of the parent chromosomes will be chosen from the upper half of the ranked chromosomes, i.e., those with normalized fitness values (f_i/f_A) greater than 1.0. As shown in Figure 6.15, the first 5 chromosomes, out of 10, take up about 78% of the wheel. Because the number of chromosomes is finite, the exact area ratio might not always be 75%.

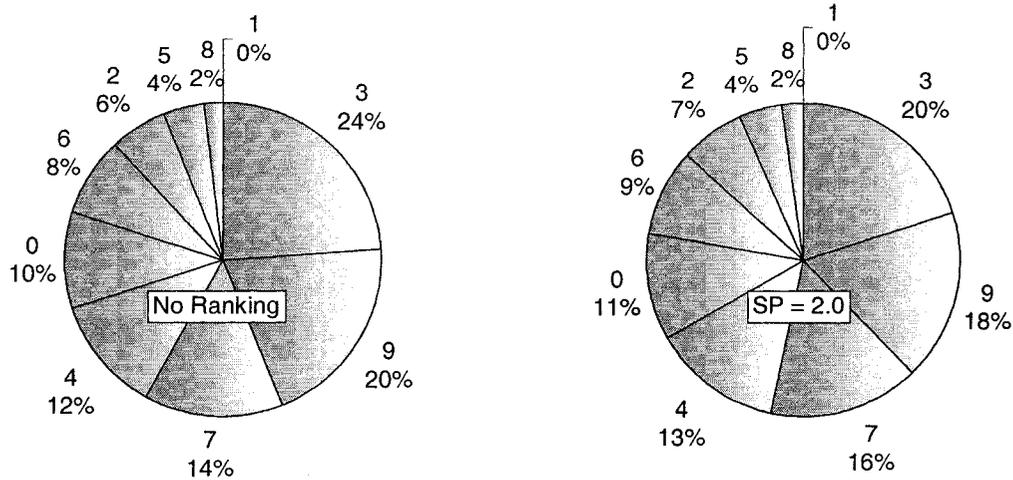


Figure 6.15 Linear ranking with a selective pressure of 2.0.

Windowing

Another method of controlling the selective pressure among chromosomes is called windowing. In this scheme, the (absolute) fitness values of all chromosomes are reduced by the fitness value of the worst chromosome in the population. This serves to increase the selective pressure among chromosomes. For example, if the fitness values all lie between 0.8 and 1.2 for example, applying the windowing technique would then make the fitness values go between 0.0 and 0.4. The relative fitness value of the best chromosome to the worst chromosome will be greater after windowing is applied. Also, the subtracted fitness value could be taken as an average value over several generations to make the process steadier.

Another related method is truncation selection. In this method, a certain percentage of the worst chromosomes, according to their rank in the population, are not allowed to reproduce.

Further Selection Ideas

Another feature of some algorithms is steady-state selection. Here, the idea is to have a larger portion of the previous generation survive to the next generation unchanged. The worst chromosomes in the generation are replaced with the offspring from the better chromosomes.

Parallel Algorithm Execution

One aspect of evolutionary algorithms is that they work on a population of individuals instead of just one individual. This immediately raises concerns about the speed with which the chromosomes can be evaluated. The chromosomes must be evaluated (relatively) quickly so that many generations can be simulated in order to reach an acceptable solution.

The genetic algorithm used with the ffx code has been designed to run in parallel with respect to chromosome evaluation. One “main” program handles all aspects of the algorithm execution. The main program takes care of creating each chromosome generation, either randomly on the first generation or

through ranking, elitism, selection, crossover, and mutation on subsequent generations, and then divides up the chromosomes for evaluation among all the programs working on the algorithm. The main program along with each of the “auxiliary” programs evaluates a share of the chromosomes calling the ffx code when needed.

For instance, 8 programs may be typically set up to evaluate a population size of 25 chromosomes. One chromosome will generally not need to be simulation as it is a result of the elitism step of the algorithm. The main program divides up the 25 chromosomes equally among the main and auxiliary programs for fitness value evaluation. Each of the 8 programs will essentially simulate 3 chromosomes, which should ideally take about the same amount of time. The main program waits for the data to be returned from all auxiliary programs before it moves on to creating the next generation of chromosomes.

Depending on the algorithm goal, the evaluation of a single chromosome can take between 10 and 20 minutes. Using a single program to evaluate all 24 chromosomes would take about 6 hours per generation. Once again, it depends on the goal, but a good solution to the problem might not be found until 50 to 100 generations of the algorithm have passed. If eight programs are used, the evaluation of 24 chromosomes can be accomplished within an hour, which becomes much more reasonable.

Considering that the development of an ion source might take on the order of years to complete, spending a few days of, ideally uninterrupted and unaided, computer time to design a set of grids does not seem like an unreasonable thing to do.

6.2 Previous Genetic Algorithms

The use of a genetic algorithm for grid design was first proposed by Nakayama in a 2001 paper presented at the 37th Joint Propulsion Conference. In that work, Nakayama developed a Thruster Performance Evaluation Code (TPEC) which was used to optimize a discharge chamber design. The results of the genetic algorithm were in good agreement with experimental data. Eight genes were used in each chromosome that coded for the propellant flow rate, discharge voltage, discharge current, and five additional parameters that describe the primary electron paths through the discharge chamber. Nakayama optimized the discharge chamber design using a fitness value that was the propellant utilization efficiency multiplied by the thrust divided by the total power consumption. The fitness value thus aimed for a high propellant utilization and high thrust-to-power ratio. The algorithm was performed with 200 chromosomes and run for 300 generations.

Nakayama also presented a second use of a genetic algorithm in 2001 at the 27th International Electric Propulsion Conference. In that work, he used the algorithm with the igx code which he developed for three-dimensional, high speed, evaluation of ion thruster optics. The fitness value used was one that optimized the thrust factor multiplied by a beam current ratio taken to a power to expand differences between the chromosomes. Nakayama also checked to make sure electron backstreaming was avoided, otherwise the fitness value was set to zero.

$$f_v = (ft \times cr)^{100} = \left(\sum \left(\frac{v_r}{\sqrt{v_z^2 + v_r^2}} \right) \times \frac{(I_b - I_{CX})}{I_b} \right)^{100} \quad \text{Eq. 6.7}$$

Five genes were used per chromosome which coded for the screen grid thickness (t_s), grid gap (l_g), accel grid applied potential (V_a), accel grid hole diameter (d_a), and the accel grid thickness (t_a). The screen grid applied potential (V_s) and the screen grid hole diameter (d_s) were held constant. The algorithm was run at several target beamlet currents, which ended up giving different optimized parameter values at each beamlet current. In those cases, the optimization was considered complete on or before the 80th generation where 100 chromosomes were used per generation.

6.3 Algorithm For Grid Design

6.3.1 Algorithm Guidelines

A set of guidelines for an appropriate evolutionary algorithm was determined through many trial applications. These guidelines are described in Table 6.2. Many of the features of both genetic algorithms and evolution strategies have been programmed into the ffx code. No clear distinction has been made between the two algorithm types to say that one is going to be better than the other at solving problems related to grid design.

Table 6.2 Guidelines for the evolutionary algorithm.

Population	25	Eight programs running in parallel evaluated 3 chromosomes per program per generation.
Encoding	Real Value	Real value encoding was used more often than binary value encoding.
Scaling	Rank Selection	Linear rank selection was used with a selective pressure of 2.0.
Selection	Roulette Wheel	
Elitism	Yes	The best chromosome was copied directly to the next generation.
Crossover	Intermediate Recombination	Intermediate recombination was used with real value encoding and uniform crossover was used with binary value encoding. Adaptive crossover was used, with rates between 50 and 75 %. Both child chromosomes are placed into the next generation.
Mutation	$R \approx 0.5, k = 8$	A larger than typical value of R was used to maintain population diversity. Adaptive mutation was used, with rates between 1.5 and 2.5 values per chromosome.
Randomness		In keeping with the ideas of adaptive crossover and mutation, less than average chromosomes that underwent crossover were completely randomized prior to evaluation to increase population diversity.

6.3.2 Design Goal

One goal of grid design might be to design a set of grids to provide the maximum change in momentum per unit grid area. The question is, given a minimal set of input conditions, what should the grid set look like? The purpose of using an evolutionary algorithm design technique is to specify as few input values as possible, letting the algorithm determine the rest of the unknown variables.

6.3.3 Input Values

The main set of input values are:

1. Net accelerating voltage, V_N . The net voltage can be determined first from a desired specific impulse, I_{sp} .
2. Current density, j . The grids are designed based on a current density instead of the thruster area, A_T , or the total thruster beam current, J_B , because current density is independent of thruster size. Later, given a thruster area or beam current (to obtain a certain amount of thrust for example), the remaining unknown variable is determined using $J_B = j \cdot A_T$.
3. Propellant, xenon. Xenon will be used, as it is the most common ion thruster propellant. The results for xenon will likely apply to other propellants.
4. Grid material, molybdenum. Molybdenum is the most common grid material. The results for molybdenum should apply to grids made of carbon or titanium for example.
5. Discharge voltage, $V_d = 25 V$. The discharge voltage is usually nearly constant for all thrusters. Here, V_d is set to 25 V. This in turn fixes the screen grid voltage, $V_s = V_N - V_d$.

6. Double to single current ratio. $\frac{J_b^{++}}{J_b^+}$. Here, the double to single current ratio will be considered

constant for any possible grid set that the algorithm might come up with. For simplicity, $\frac{J_b^{++}}{J_b^+}$ is set to zero.

7. Propellant utilization efficiency, $\eta_u = 0.9$. Here, it will be assumed that any possible grid design can be operated with a (discharge) propellant utilization efficiency of 90 %.
8. Screen grid physical transparency, $\phi_s = 67\%$. A traditional value for the screen grid open area fraction is 67 %.

The two primary input values are the net accelerating voltage, V_N , and current density, j . The net voltage is specified according to a certain mission, and the current density can be used to scale a thruster to obtain a certain level of thrust.

6.3.4 Variables

Each chromosome in the evolutionary algorithm will contain a set of values for the unknown variables. The unknown variables are:

1. Screen grid thickness, t_s .
2. Screen grid hole diameter, d_s .
3. Grid spacing, l_g .
4. Accel grid thickness, t_a .
5. Accel grid hole diameter, d_a .
6. Accel grid voltage, V_a .

There are several values that are set dependently for each chromosome prior to analysis by the ffx code. These are:

1. Aperture center-to-center spacing, $l_{cc} = \sqrt{\frac{\pi d_s^2}{2\sqrt{3} \phi_s}}$. The center-to-center spacing is determined from the input screen grid physical transparency, ϕ_s , and the screen grid hole diameter, d_s , obtained from each chromosome.
2. Beamlet current, $J_b = A_{hex} \cdot j = \frac{\sqrt{3}}{2} l_{cc}^2 j$. The beamlet current is dependent on the area per aperture, A_{hex} , and the downstream specified current density, j .
3. Perveance fraction, $f_p = \frac{j_p}{P_{max}} \frac{l_e^2}{V_T^{3/2}}$. The chromosome provides values for t_s , d_s , l_g , and V_a , which then lead to f_p .

6.3.5 Variable Ranges

Intelligent choices for the variable ranges can be determined using the perveance equation, Eq. 6.8, which relates current density to aperture size. The effective acceleration length, l_e , can be determined after choosing values for the downstream current density (j), propellant type (which sets P_{max}), and net accelerating voltage (V_N) with a few additional approximations.

$$f_p = \frac{j_p}{P_{\max}} \frac{l_e^2}{V_T^{3/2}} \quad \text{Eq. 6.8}$$

$$l_e^2 = (l_g + t_s)^2 + \frac{d_s^2}{4} = \frac{f_p P_{\max} V_T^{3/2}}{j_p}$$

The current density in the perveance equation, j_p , is the current density that actually passes through the grids. It is related to the current density downstream of the grids (j), which is simply the current of ions (J_b) divided by the total aperture area (A_{hex}), by the actual grid transparency to ions, ϕ , in Eq. 6.9. A good approximation is to set the ion transparency to be equal to the physical screen grid transparency, ϕ_s .

$$j_p = \frac{j}{\phi} \approx \frac{j}{\phi_s} \quad \text{Eq. 6.9}$$

To approximate the expected total accelerating voltage, V_T , an expected R ratio of 0.85 is chosen, as in Eq. 6.10.

$$V_T = \frac{V_N}{R_{\text{exp}}} \approx \frac{V_N}{0.85} \quad \text{Eq. 6.10}$$

Setting the perveance fraction, f_p , to unity determines the maximum possible acceleration length, $l_{e\text{max}}$, required to obtain the given current density. The apertures can be made smaller to achieve the current density, but not larger.

The effective acceleration length (l_e) is a relationship between the grid spacing (l_g), screen grid thickness (t_s), and screen grid hole diameter (d_s). The upper limit on any individual variable is found by setting the other two variables to zero. This gives $l_{g\text{max}} = t_{s\text{max}} = l_{e\text{max}}$ and $d_{s\text{max}} = 2l_{e\text{max}}$. From experience, the screen grid thickness will turn out to be relatively small in order to have a reasonable transparency to ions. For this reason, $t_{s\text{max}}$ is set to be $\frac{1}{2}l_{e\text{max}}$.

There are three remaining variables: accel grid thickness (t_a), accel grid hole diameter (d_a), and accel grid voltage (V_a). There is technically no upper limit on t_a . However, increasing the accel grid thickness eventually decreases the operating range whereby the crossover limit moves toward the perveance limit. Arbitrarily, $t_{a\text{max}}$ can be set to be $\frac{1}{2}l_{e\text{max}}$ or $l_{e\text{max}}$. The accel grid hole diameter will be held less than or equal to the screen grid hole diameter, $d_{a\text{max}} = d_{s\text{max}} = 2l_{e\text{max}}$.

The accel grid voltage, V_a , is expected to lead to an R ratio that is within a certain range, $0.85 \leq R \leq 0.9$. This specifies both the upper and lower limits on V_a , using Eq. 6.11.

$$V_a = V_N \left(1 - \frac{1}{R} \right) \quad \text{Eq. 6.11}$$

The lower limits on the geometric variables are set somewhat more arbitrarily than the upper limits. For the grid thickness and hole diameter values, the lower limits are set to be one-tenth of the maximum acceleration length, $l_{e\max}$.

The lower limit for the grid spacing, l_g , can be set according to the maximum allowed electric field, E_{\max} , between the screen and accel grids, as in Eq. 6.12.

$$l_{g\min} = \frac{\frac{V_N}{R_{\min}} - V_d}{E_{\max}} \quad \text{Eq. 6.12}$$

The variable ranges are summarized in Table 6.3. The flow chart in Figure 6.16 shows how the directly input parameters (bold) are related to the indirectly calculated values.

Table 6.3 Variable limits.

Variable	Symbol	Lower Limit	Upper Limit
Screen grid thickness	t_s	$\frac{1}{10} l_{e\max}$	$\frac{1}{2} l_{e\max}$
Screen grid hole diameter	d_s	$\frac{1}{10} l_{e\max}$	$2l_{e\max}$
Grid spacing	l_g	$\frac{\frac{V_N}{R_{\min}} - V_d}{E_{\max}}$	$l_{e\max}$
Accel grid thickness	t_a	$\frac{1}{10} l_{e\max}$	$\frac{1}{2} l_{e\max}$
Accel grid hole diameter	d_a	$\frac{1}{10} l_{e\max}$	$2l_{e\max}$
Accel grid voltage	V_a	$V_N \left(1 - \frac{1}{R_{\min}} \right)$	$V_N \left(1 - \frac{1}{R_{\max}} \right)$

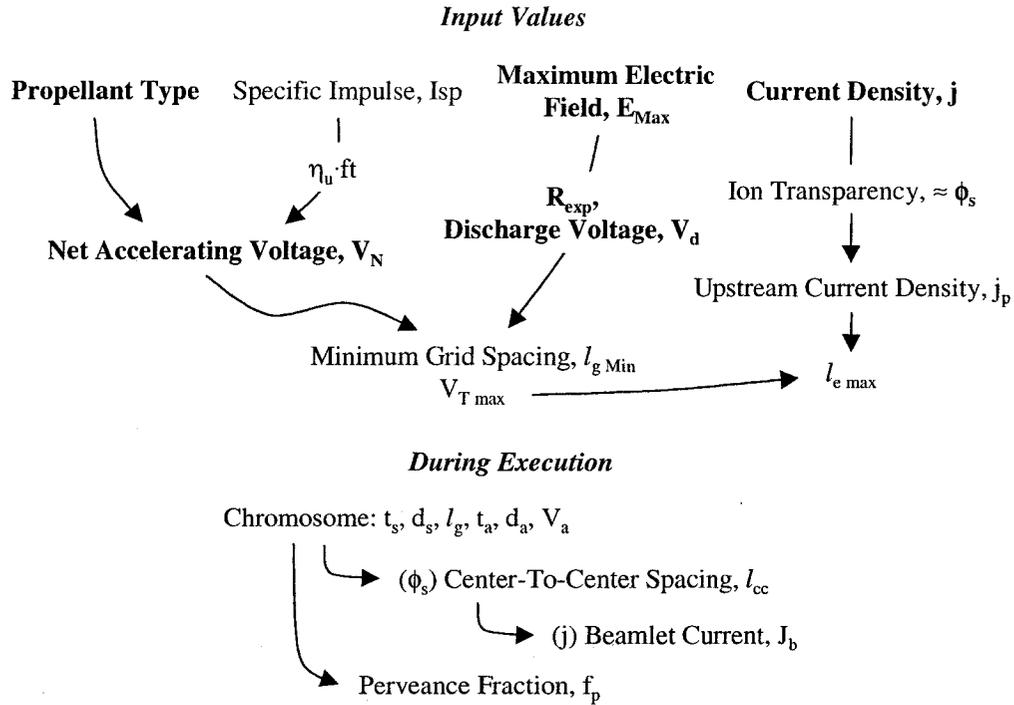


Figure 6.16 Flow chart of calculated values for the evolutionary algorithm.

6.3.6 Fitness Value

The goal of the algorithm is to optimize the impulse per unit area provided by the grids. The change in momentum per unit grid area is given in Eq. 6.13. The net accelerating voltage (V_N), propellant type (m_i), and double to single current ratio ($\frac{J_b^{++}}{J_b^+}$) are all chosen constants. The current per area, $\frac{J_b}{A_{Aperture}}$, is the downstream current density, j . The fitness value that the algorithm will maximize is f , given in Eq. 6.14, which is the product of the current density (j), predicted accel grid lifetime (L), and thrust factor (ft). The predicted accel grid lifetime is calculated as the time it would take to erode away 50% of the initial accel grid mass per aperture.

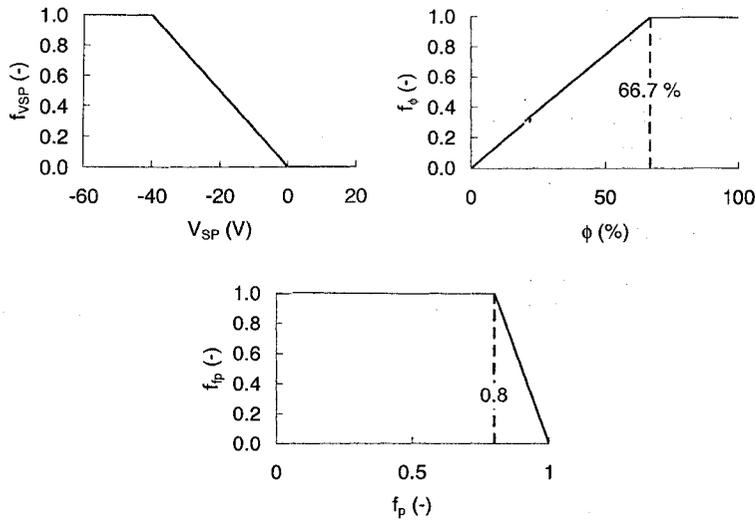
$$\frac{\Delta p}{A} = \frac{F_{Aperture} L}{A_{Aperture}} = \frac{J_b}{A_{hex}} L \frac{m_i}{e} \frac{1 + \frac{\sqrt{2} J_b^{++}}{2 J_b^+}}{1 + \frac{J_b^{++}}{J_b^+}} ft \sqrt{\frac{2eV_N}{m_i}} \quad \text{Eq. 6.13}$$

$$f = j \cdot L \cdot ft \quad \text{Eq. 6.14}$$

A few additional constraints are placed on the fitness function to ensure that good chromosomes are practical. The constraints are applied through conditional multiplying factors, f_i . If a constraint is met, no adjustment to the fitness value is made ($f_i = 1$). If a constraint is not met, the fitness value is linearly reduced.

First, the centerline saddle point potential is required to be “sufficiently” negative. If the saddle point potential is, for example, more than 40 V negative of the downstream beam plasma potential, ϕ_d , the solution is deemed acceptable. Second, to validate the assumption that ion transparency is near the physical screen grid transparency, the grid solution is said to be reasonable when $\phi \geq \phi_s$. This also serves to ensure that electrical efficiency is reasonable. Third, the beamlet is restricted to operate at a perveance fraction less than about 0.8. This, admittedly, is a relatively arbitrary constraint intended to keep the aperture well proportioned. The new fitness function is given in Eq. 6.15.

$$f = j \cdot L \cdot f_t \cdot f_{V_{SP}} \cdot f_{\phi} \cdot f_{f_p}$$



Eq. 6.15

6.3.7 Algorithm Results

Results are presented for four cases: two net voltages, 1000 and 3000 V, operating at two current densities, 25 and 50 A/m². These conditions are congruent to the operating conditions of the NSTAR and NEXT thrusters.

A detailed study of the 1000 V, 25 A/m² case is presented first to illustrate the evolutionary algorithm. Table 6.4 shows the setup for this case.

Table 6.4 Input parameter values and variable ranges for the first case.

Variable	Symbol	Units	Value	
<i>Input Quantities</i>				
Net accelerating voltage	V_N	V	1000	
Current density	j	A/m ²	25	
R ratio range	$R_{\min} \leq R \leq R_{\max}$	-	$0.85 \leq R \leq 0.9$	
Maximum allowed electric field	E_{\max}	kV/mm	3	
<i>Derived Quantities</i>				
Maximum effective acceleration length	$l_{e\max}$	mm	2.266	
Variable	Symbol	Units	Lower Limit	Upper Limit
Screen grid thickness	t_s	mm	0.227	1.133
Screen grid hole diameter	d_s	mm	0.227	4.532
Grid spacing	l_g	mm	0.384	2.266
Accel grid thickness	t_a	mm	0.227	1.133
Accel grid hole diameter	d_a	mm	0.227	4.532
Accel grid voltage	V_a	V	-176	-111

Figure 6.17 shows the fitness value of the best chromosome in a generation compared to the average fitness value of all chromosomes in the generation. As the algorithm progresses, the best and average fitness values should be increasing, eventually reaching steady state.

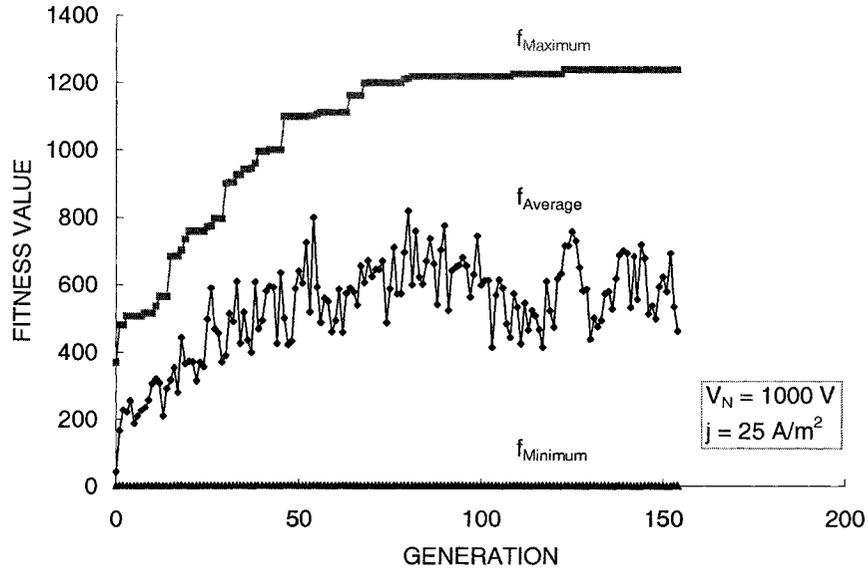


Figure 6.17 Convergence history of the evolutionary algorithm.

Figure 6.18 shows the range of chromosome fitness values on three generations. In this algorithm, the actual fitness values are not used for reproduction. Instead, the chromosomes are ranked by fitness and scaled fitness values are assigned to determine selection probability.

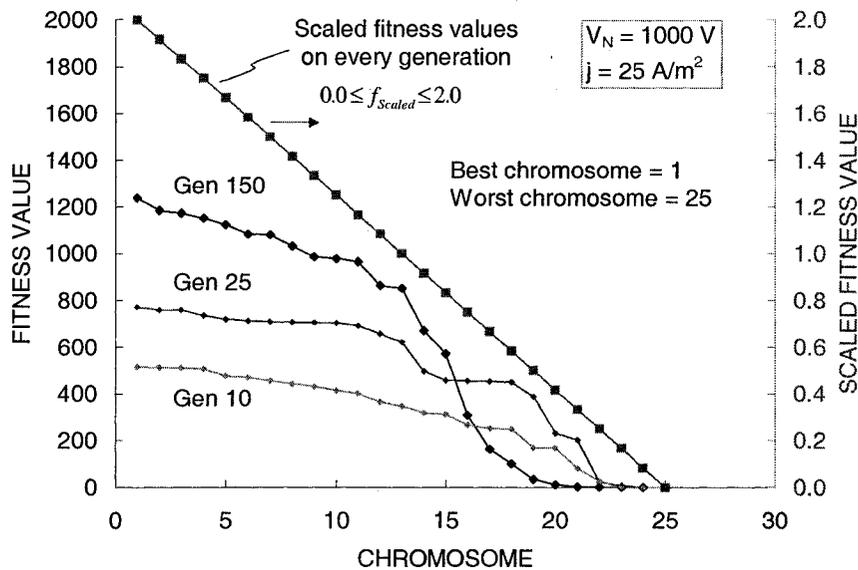


Figure 6.18 Population fitness value as a function of generation.

Figure 6.19 shows the actual chromosome variable values on generation 150 ranked by fitness. Most chromosomes are similar, indicating convergence. There are also a few vastly different chromosomes, created either through heavy mutation or randomly from scratch after a poor chromosome was selected for reproduction. The allowed variable ranges are shown to the left of the chromosomes.

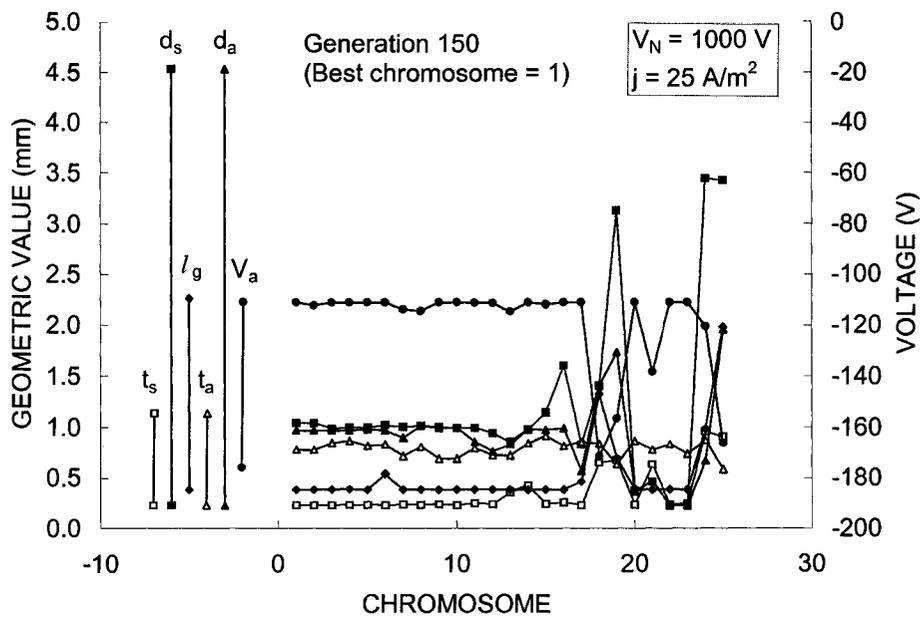


Figure 6.19 Variable values within a generation.

Finally, Figure 6.20 shows the variable values of the best chromosome in every generation. Often, the best chromosome in a generation changes only slightly through crossover or mutation to become the best chromosome in the next generation. Early in the algorithm, better chromosomes may crop up unexpectedly. In this case, the allowed domain ranges eventually limit the screen grid thickness, t_s , and accel grid voltage, V_a .

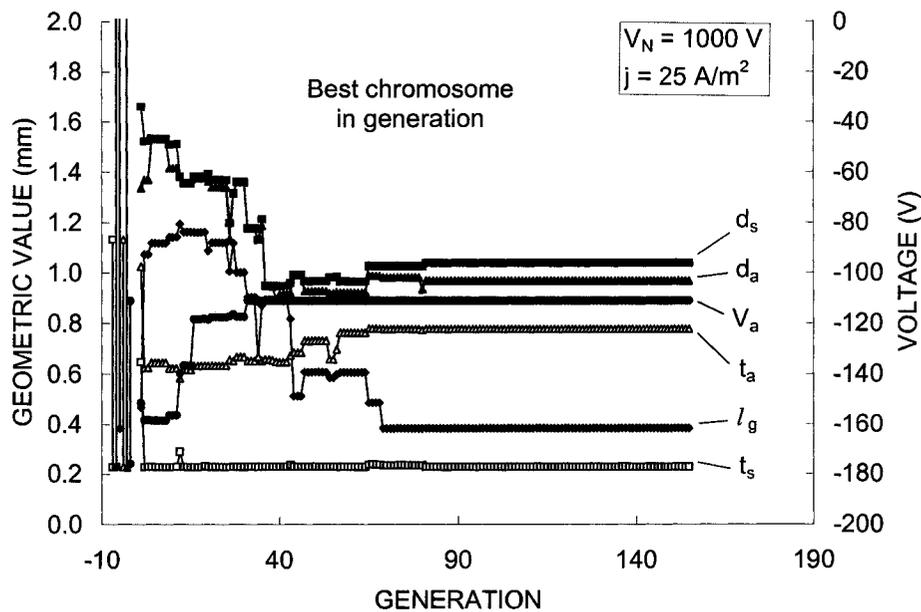


Figure 6.20 The best chromosome in every generation.

It turned out that in the particular case of $V_N = 1000$ V, the algorithm generally converged to one of two solutions, depending on the initial set of randomly generated chromosomes and those generated during execution. The solutions are referred to here by their relative permeance fraction, f_p , values. One solution converged to a low f_p , and the other solution essentially maximized f_p .

The low f_p solution has a higher fitness value than the high f_p solution. Ideally then the algorithm should always find the low f_p solution. However, at mid f_p values the fitness is lower, creating a valley between the end solutions. Once the algorithm moves toward one of the solutions it becomes increasingly unlikely that the algorithm will switch to the other solution.

Solutions with permeance fractions near 0.5 can be found by changing the fitness value multiplying factor f_{f_p} to emphasize such chromosomes. For example, when f_p is 0.0 and 0.8 the factor is zero and when f_p is 0.5 the factor is unity, with linear variations in between. The low, medium, and high f_p solutions are presented in Figure 6.21 (sized relatively) for the 50 A/m² condition. All three beamlets in this figure are operating at the same current density. Thus, many more of the low f_p beamlets fit into the same area as the high f_p beamlet, producing the same amount of current.

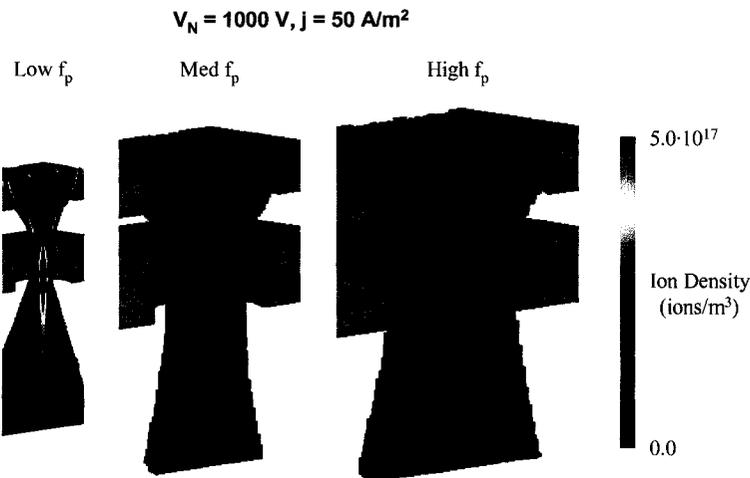


Figure 6.21 Three results of the evolutionary algorithm for a net voltage of 1000 V.

The algorithm results for the 1000 V net voltage cases are presented in Table 6.5. Low, medium, and high f_p solutions were found for both current density cases. Variables are highlighted if they are within 10 % of the domain minimum or maximum.

For the low permeance fraction solution, t_s and l_g were minimized and V_a maximized. Also, d_a was large, presumably limited by the restriction of being less than or equal to d_s . This solution is understandable for several reasons. First, V_a is maximized to minimize the ion sputter yield. Second, operation at a low permeance fraction focuses a higher fraction of charge exchange ions through the accel grid than operation at a high permeance fraction. Third, d_a is maximized to reduce the neutral density near the grids by providing less resistance to neutral flow.

The low perveance fraction solutions have a few disadvantages. First, the grid apertures are smaller, increasing difficultly in manufacturing and alignment. Second, the divergence angles are worse. Finally, the beamlets are operating near the crossover limit; the beamlet current can be increased, but not decreased.

With the medium or high perveance fraction solutions, l_g is minimized and t_a is maximized. With these solutions, the accel grid voltage needs to be more negative to prevent electron backstreaming. Also, d_a is not maximized with respect to d_s . Decreasing the accel grid hole diameter adds accel grid mass, which extends lifetime, as a tradeoff to increased neutral density.

Table 6.5 Algorithm results for a net voltage of 1000 V.

Input Values		1000					
V_N	V				50		
j	A/m ²	25			50		
$l_{e\max}$	mm	2.266			1.602		
Results		L	M	H	L	M	H
l_{cc}	mm	1.208	2.545	3.566	0.866	1.856	2.495
t_s	mm	0.227	0.598	0.824	0.160	0.408	0.526
d_s	mm	1.038	2.187	3.065	0.744	1.595	2.144
l_g	mm	0.384	0.529	0.436	0.390	0.384	0.384
t_a	mm	0.778	1.094	1.132	0.522	0.800	0.801
d_a	mm	0.969	1.533	1.988	0.705	1.294	1.410
V_a	V	-111.0	-129.0	-163.4	-111.0	-164.0	-163.9
		At or near domain <i>minimum</i>					
		At or near domain <i>maximum</i>					
Generation	-	155	161	200	133	54	178
Fitness	y-A/m ²	1238	651	892	572	308	445
J_b	mA	0.032	0.141	0.277	0.033	0.150	0.271
f_p	-	0.14	0.51	0.78	0.19	0.50	0.78
ϕ	%	82.4	67.0	59.4	76.9	65.6	62.7
V_{SP}	V	-59	-38	-39	-58	-44	-41

A slightly different set of solutions was obtained for the cases at a net voltage of 3000 V. At the higher net voltage, the algorithm always wanted to go to the low f_p solution, for the same reasons outlined previously.

For the medium f_p solution, a significantly different grid solution was obtained, referred to as M1 in the tables. At 1000 V, the grid spacing l_g was minimized, requiring a slightly more negative accel grid to counteract electron backstreaming. At 3000 V however, the M1 solution maximized V_a and kept d_s small, increasing l_g to make f_p go to 0.5. This solution is odd because it leads to difficulties in grid alignment. A second medium f_p solution was obtained by fixing l_g at its minimum value, referred to as M2 in the tables. This solution was similar to the solution obtained at 1000 V.

Figure 6.22 shows the three perveance fraction solutions obtained for a net voltage of 3000 V and a current density of 50 A/m². The algorithm results for the cases at a net voltage of 3000 V are listed in Table 6.6.

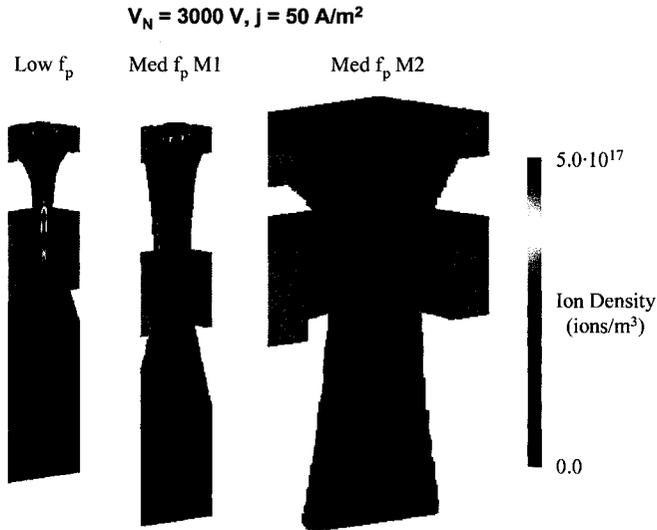


Figure 6.22 Grid solutions obtained for a net voltage of 3000 V.

Table 6.6 Algorithm results for a net voltage of 3000 V.

Input Values		3000					
V_N	V	3000					
j	A/m ²	25			50		
$l_{e \max}$	mm	5.165			3.652		
Results		L	M1	M2	L	M1	M2
l_{cc}	mm	1.900	2.036	6.275	1.360	1.348	4.000
t_s	mm	0.516	0.518	1.275	0.365	0.365	0.702
d_s	mm	1.633	1.750	5.394	1.169	1.158	3.438
l_g	mm	1.168	2.554		1.168	1.944	
t_a	mm	2.032	1.919	2.558	1.501	1.424	1.826
d_a	mm	1.618	1.618	4.074	1.118	1.158	2.982
V_a	V	-345.0	-333.0	-384.4	-333.0	-333.0	-400.5
At or near domain <i>minimum</i> .							
At or near domain <i>maximum</i> .							
Generation	-	34	32	60	54	36	67
Fitness	y-A/m ²	622	352	117	301	187	63
J_b	mA	0.079	0.090	0.857	0.080	0.079	0.696
f_p	-	0.14	0.42	0.53	0.22	0.46	0.51
ϕ	%	70.9	55.0	66.8	67.8	47.8	67.9
V_{SP}	V	-300	-281	-31	-289	-289	-65

The evolutionary algorithm uses initial erosion rates to predict grid lifetime. The ffx code was used to perform full lifetime simulations for the algorithm solutions. For each net voltage and current

density combination, a constant downstream length was used to fairly compare grid lifetime. The erosion predictions are shown in Figure 6.23, where end of life is taken to occur at 50 % accel grid mass loss. Uncolored cells indicate remaining grid material, while any colored cells have been completely eroded away.

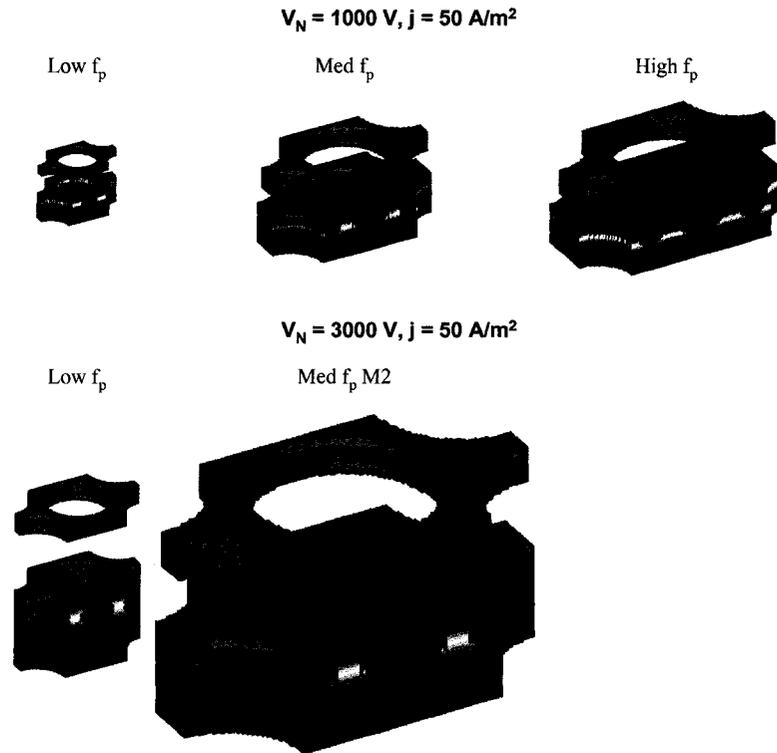


Figure 6.23 Evolutionary algorithm solutions near 50 % accel grid mass loss.

The lifetime predictions from the algorithm and the full lifetime simulations are compared in Table 6.7. Note that the lifetime results are highly dependent on the listed accel grid impingement current ratios, J_a/J_b .

Table 6.7 Full lifetime simulation results.

V_N	V	1000						3000			
		25			50			25		50	
j	A/m^2	L	M	H	L	M	H	L	M2	L	M2
f_p	-										
<i>From genetic algorithm</i>											
Generation	-	155	161	200	133	54	178	34	60	54	67
Fitness	$\text{yrs} \cdot A/m^2$	1238	651	892	572	308	445	622	117	301	63
Life	yrs	50.6	27.5	36.0	11.6	6.2	9.0	25.3	4.7	6.2	1.3
<i>From full lifetime simulations</i>											
Life	yrs	74.6	28.2	18.9	15.7	4.6	4.5	8.0	6.1	4.7	1.6
ft	-	0.978	0.994	0.991	0.984	0.994	0.991	0.982	0.996	0.988	0.996
Fitness	$\text{yrs} \cdot A/m^2$	1825	701	468	773	230	223	195	152	233	80
<i>Average values over grid lifetime</i>											
J_a/J_b	%	0.14	0.26	0.25	0.19	0.30	0.35	0.28	0.46	0.42	0.53
ϕ	%	79.8	65.1	59.6	75.6	65.8	61.5	72.1	65.6	67.0	65.3
V_{SP}	V	-52	-28	-20	-49	-32	-22	-282	-10	-279	-29
α	deg	21	12	13	18	11	12	18	10	16	9

Finally, there are several ways to compare the grid solutions. One way is to normalize each solution by screen grid hole diameter, d_s . This results in similar normalizations with respect to perveance fraction, and the averages of these normalizations are given in Table 6.8. The normalized NSTAR grid geometry has a relatively thin, closed, accel grid compared to the algorithm solutions.

Table 6.8 Average geometric values normalized to d_s .

		Low f_p	Medium f_p	High f_p	NSTAR
l_{cc}/d_s	-	1.16	1.16	1.16	1.16
t_s/d_s	-	0.27	0.24	0.26	0.20
d_s/d_s	-	1.00	1.00	1.00	1.00
l_g/d_s	-	0.65	0.26	0.16	0.35
t_a/d_s	-	0.99	0.50	0.37	0.27
d_a/d_s	-	0.96	0.78	0.65	0.60

For design purposes, one could combine the maximum expected current density with a medium f_p solution grid. Then, other beamlets elsewhere on the grid surface operating at lower beamlet currents, still above the crossover limit, would not be life limiting. Alternatively, one could conceivably design for the minimum expected current density by selecting a low f_p grid solution, ensuring that all beamlets operate above the expected current density.

Previously, optimizations were performed manually for the example grid set at several beamlet currents. The evolutionary algorithm was used to perform these optimizations as well, and the results are compared in Table 6.9. In these cases, only two parameters were varied at a time, maximizing grid lifetime under the requirement that the centerline saddle point potential be held at least 40 V negative of the downstream beam plasma potential. The manual and algorithm optimizations turned out similarly.

Table 6.9 Optimization results for the example grid set.

J_b mA	Optimization Method	l_{cc} mm	t_s mm	d_s mm	l_g mm	t_a mm	d_a mm	V_a V
	Reference	2.327	0.4	2	1	0.8	1.2	-400
0.3	Manual	2.327	0.4	2	0.925	0.8	1.2	-197
0.3	Algorithm	2.327	0.4	2	0.881	0.8	1.2	-190
0.3	Manual	2.327	0.4	2	1	0.8	1.545	-245
0.3	Algorithm	2.327	0.4	2	1	0.8	1.881	-295

7 Code Validation

The ffx code has been compared to several other ion optics codes and to experimental data, including both sub-scale gridlet tests and full thruster tests.

7.1 Gridlet Tests

There are several instances where it is beneficial to test sub-scale grid sets, called “gridlets”, rather than full grid sets. The idea behind gridlet tests is that the test conditions can be controlled more easily than when working with large grid sets.

Gridlets are typically operated using a relatively large discharge chamber compared to the area of ion extraction. This implies that the plasma is approximately uniform upstream of the apertures. Additionally, the grid spacing, l_g , for example can be controlled more accurately in gridlet tests compared to large grid tests where thermal expansion can be significant and unpredictable. As a result, predictions concerning the onset of electron backstreaming with variation in accel grid voltage can be made more accurately as well.

7.2 Carbon Based Ion Optics (CBIO) Grid Set

The CBIO project had the goal of determining the applicability of carbon-carbon material for ion thruster optics. Specifically, the project assessed grid performance, vibration survival, voltage standoff capability, manufacturability, and erosion resistance.

Several gridlets have been constructed and experimentally tested using CBIO-similar grid geometry. Gridlets with 1, 7, 19, and 37 apertures have been tested to determine crossover and perveance limits, Figure 7.1, as well as the accel voltages required to prevent electron backstreaming for a range of net accelerating voltages. The CBIO grid geometry as tested is listed in Table 7.1, and the operating conditions are listed in Table 7.2 [Wang].

Table 7.1 CBIO grid geometry.

Parameter	Symbol	mm	≈ in
Aperture center to center spacing	l_{cc}	2.674	0.105
Screen grid hole diameter	d_s	2.305	0.091
Screen grid thickness	t_s	0.461	0.018
Grid spacing	l_g	0.813	0.032
Accel grid hole diameter	d_a	1.396	0.055
Accel grid thickness	t_a	1.016	0.040

Table 7.2 CBIO operating conditions.

Net Accelerating Voltage	Accel Grid Voltage
V_N (V)	V_a (V)
800	-140
1000	-150
1200	-166
1420	-170
1600	-176
1800	-190
2000	-195
2200	-200

$$J_b = 0.379 \text{ mA}, f_p = 0.75$$

$$V_N = 1600 \text{ V}, V_a = -176 \text{ V}$$

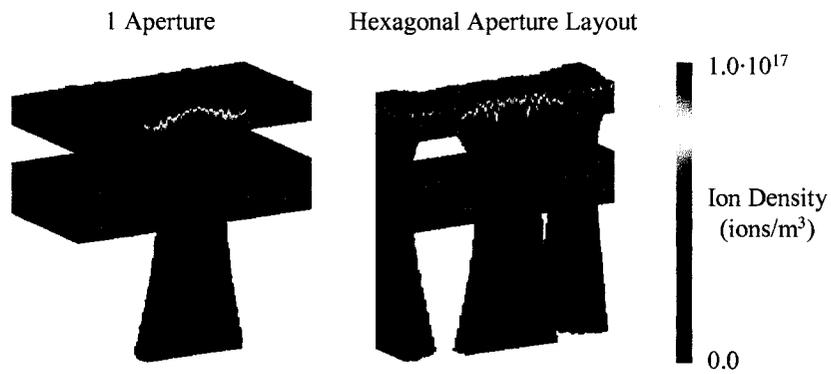


Figure 7.1 CBIO grids operating near the perveance limit.

Perveance limits, found both experimentally and numerically, are shown in Figure 7.2 over a range of total accelerating voltages. Likewise, the corresponding crossover limits are shown in Figure 7.3. The code data is shown as found with perfectly aligned apertures and as found when using slightly misaligned holes. The hexagonal aperture layout is essentially the case of an infinite number of apertures, thus it is most similar to the 37 aperture gridlet.

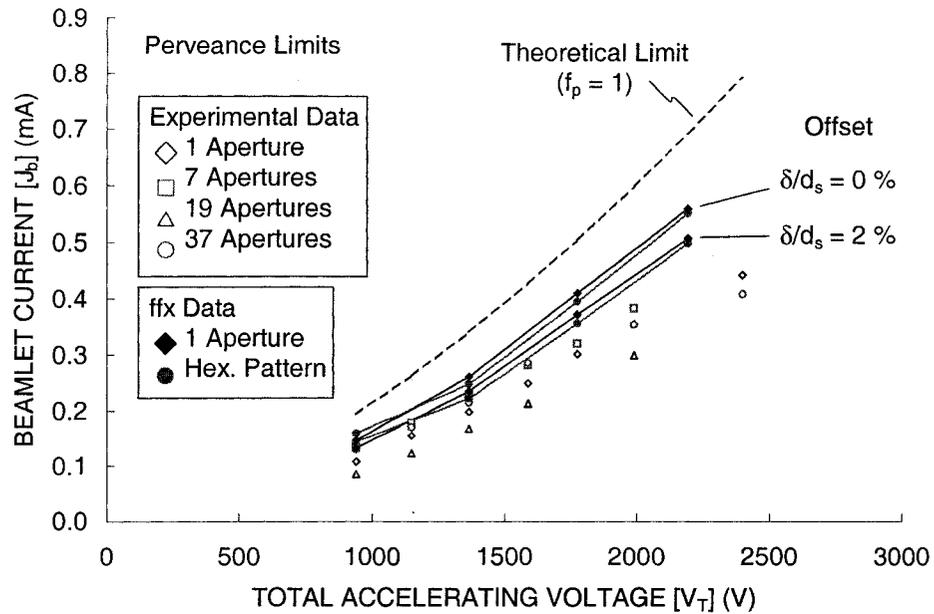


Figure 7.2 CBIO perveance limits as a function of total accelerating voltage.

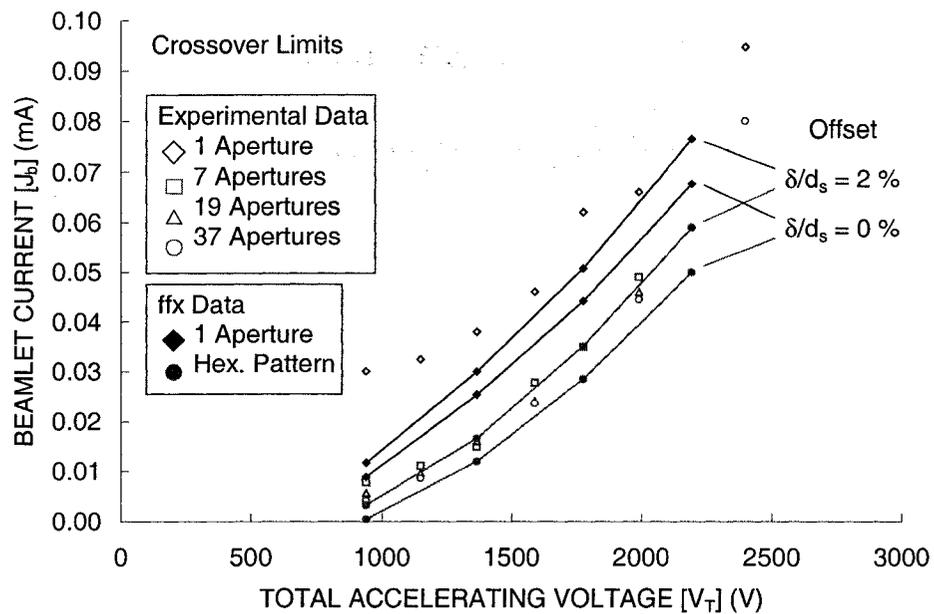


Figure 7.3 CBIO crossover limits as a function of total accelerating voltage.

In general, the code predicts greater perveance limits and lesser crossover limits than found experimentally in the gridlet tests. As seen elsewhere, the perveance limits predicted in the 1 aperture and hexagonal aperture layout cases are essentially the same as a result of little to no sheath interaction. The crossover limits, however, are different due to significant sheath interaction in the hexagonal layout case.

Two factors that have been seen to affect the code crossover and perveance limit predictions significantly are a) aperture misalignment and b) variation of the upstream plasma potential relative to the screen grid potential.

Aperture misalignment decreases the perveance limits and increases the crossover limits, bringing them more in line with the gridlet data. Decreasing the upstream plasma potential increases the crossover limits, but doing so also increases the perveance limits. Decreasing the electron temperature was seen to increase the crossover limits, but only slightly compared to the aforementioned effects. Figure 7.4 compares aligned grids to grids where the accel grid is offset by 5 % of the screen grid hole diameter. It is unlikely that an offset greater than 5 % is not visually detectable in the experimental gridlet tests.



Figure 7.4 CBIO grid misalignment.

The crossover and perveance limits in the previous figures were corrected using the offset data, found using the code in the 1 aperture case, listed in Table 7.3. The crossover and perveance limits were found to approximately vary by a constant perveance fraction, Δf_p , for any total accelerating voltage. A constant Δf_p results in a non-constant change in beamlet current, ΔJ_b , because of the nonlinear affect of total voltage.

Table 7.3 CBIO offset hole crossover and perveance limit predictions.

V_N	V_a	V_T	Offset	Crossover Limit	Perveance Limit
V	V	V	δ/d_s %	J_b mA	J_b mA
800	-140	940	0	0.007	
800	-140	940	5	0.014	
1200	-166	1366	0	0.026	0.314
1200	-166	1366	5	0.037	0.250
1200	-166	1366	10	0.054	0.166
2000	-195	2195	0	0.075	
2000	-195	2195	5	0.097	

Near the crossover limit in this case, as with many other observed cases, the plasma sheath moves far upstream away from the screen grid as the plasma density is very low. It is unknown if it is possible that the upstream plasma potential is lower near the grids than it is when the plasma density is greater, where it is thought that the upstream plasma potential is likely close to the anode potential. Nevertheless, it is observed in the code that lowering the upstream plasma potential near the crossover limit improves

agreement with the experimental gridlet tests. In this particular case, reducing the upstream plasma potential to about 18 V above the screen grid potential, at any net voltage, will bring the crossover limits into agreement without using any misalignment assertion.

Backstreaming data were also recorded for the four CBIO gridlets. Code predictions of the backstreaming limit variation as a function of net accelerating voltage are compared to the experimentally determined backstreaming limits in Figure 7.5.

Experimentally, the accel grid voltage is reduced until the indicated beamlet current begins to rise, signaling the onset of electron backstreaming. In the code, the accel grid voltage is reduced until the centerline saddle point potential rises to within a few volts of the downstream plasma potential, where higher energy electrons will likely begin to backstream. The experimental downstream plasma potential is unknown in this case, but it is not unreasonable to take it to be about 5 V. The code predictions of backstreaming limit were assigned when the saddle point potential became 0 V. Using these definitions, the experimental and numerical backstreaming limits are in very good agreement.

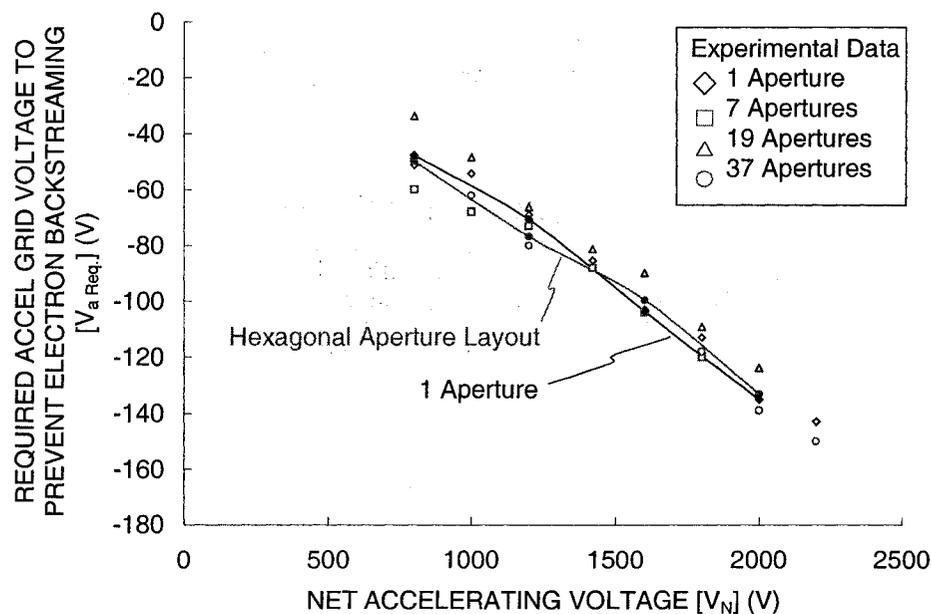


Figure 7.5 Good agreement between the numerically predicted and gridlet determined CBIO backstreaming limits.

7.3 NSTAR

The NSTAR ion engine is a 30 cm diameter thruster capable of operating over a power range of 0.5 to 2.3 kW, or a specific impulse range of 1900 s to 3100 s. NSTAR stands for the NASA Solar Technology Application Readiness program, which developed the thruster. An NSTAR ion engine was used on the highly successful Deep Space 1 mission, which visited the asteroid Braille in July of 1999 and

comet Borelly in September of 2001 [Brophy]. Over the course of the mission, the NSTAR ion engine processed 73.4 kg of xenon propellant and operated for 16,265 hours [Anderson, Sengupta].

Several life tests using NSTAR ion thrusters have been performed. Here, the 8,192 hour Life Demonstration Test (LDT) and 30,352 hour Extended Life Test (ELT) will be reviewed. Prior to those tests, a 2,000 hour test and a 1,000 hour test were performed to identify and correct design deficiencies.

The Extended Life Test was conducted at the Jet Propulsion Laboratory from 1998 to 2003 using the Deep Space 1 mission flight spare thruster, FT2. During the Extended Life Test, the NSTAR thruster processed 235.1 kg of xenon and operated for 30,352 hours under continuous vacuum.

The Life Demonstration Test used the second Engineering Model Thruster (EMT2) constructed at the NASA Glenn Research Center [Polk]. It operated for a total of 8,192 hours, processing 88 kg of xenon propellant at a power level of 2.3 kW.

Extensive data has been collected on all aspects of the NSTAR ion engine through both experimental and numerical tests, making it a useful validation tool [Foster, Soulas].

The NSTAR grid geometry is listed in Table 7.4 [Diaz]. The apertures in the grids are chemically etched, a process that results in cusps midway through the apertures. The inner diameter listed for the screen and accel grid holes describes the smallest diameter of an aperture, which occurs at the cusp tips. The outer diameter is the hole diameter at the upstream or downstream edge of an aperture.

Table 7.4 NSTAR grid geometry.

Parameter	Symbol	mm/mm
Aperture center to center spacing	l_{cc}/d_s	1.16
Screen grid hole diameter	d_s/d_s	
Inner diameter, at cusp tips		1
Outer diameter, at entrance/exit		1.08
Screen grid thickness	t_s/d_s	0.2
Grid spacing	l_g/d_s	
Cold gap, not operating		0.347
Hot gap, during operation		0.16
Accel grid hole diameter	d_a/d_s	
Inner diameter		0.6
Outer diameter		0.707
Accel grid thickness	t_a/d_s	0.267

Although the NSTAR thruster is 30 cm in diameter, the active diameter with apertures is about 28.5 cm. This results in about 15085 total apertures over the grid surface.

For the Extended Life Test, 16 throttle points, TH0 to TH15, were selected from 50 throttle points available to the Deep Space 1 thruster. A partial listing of the throttle table is given in Table 7.5.

Table 7.5 Partial NSTAR throttle table.

Throttle Level	Flow Rates			Propellant Utilization Efficiency	
	Main	Cathode	Neutralizer	Discharge	Total
TH	η_u				
-	sccm	sccm	sccm	%	
0	5.89	2.42	2.33	86	67
3	6.77	2.42	2.33	93	74
6	11.15	2.42	2.33	93	80
9	15.72	2.42	2.33	92	82
12	19.52	2.82	2.74	93	83
15	23.25	3.62	3.52	91	81

TH	V_N	V_a	J^{++}/J^+	J_B	J_A	F
-	V	V	%	A	A	-
0	650	-150	4	0.51	0.9	0.37
3	1100	-180	8	0.61	1.1	0.36
6	1100	-180	12	0.91	2.0	0.40
9	1100	-180	14	1.20	3.2	0.45
12	1100	-180	15	1.49	4.6	0.48
15	1100	-180	17	1.76	6.5	0.50

The flight spare thruster was operated at seven throttle points over the course of the Extended Life Test, usually for 4500 to 5500 hours per segment. A summary of the conditions at each of the segments is given in Table 7.6.

Table 7.6 Extended Life Test operation summary.

Segment	TH	Power	Operation Time	Propellant Use	V_N	V_a	J^{++}/J^+
-	-	kW	hours	kg	V	V	%
1	12	2.0	447	3.9	1100	-180	15
2	15	2.3	4246	44.0	1100	-180	17
3	8	1.5	5758	39.7	1100	-180	13
4	15	2.3	5166	54.3	1100	-250	17
5	0	0.5	5689	22.4	650	-150	4
6	15	2.3	4400	46.5	1100	-250	17
7	5	1.1	4646	24.3	1100	-250	11
Total:			30352	235.1			

Segment	J_B	J_A	J_b avg	J_a avg	F	J_b peak	J_a peak
-	A	A	mA	mA	-	mA	mA
1	1.49	4.6	0.099	0.00030	0.48	0.206	0.00046
2	1.76	6.5	0.117	0.00043	0.50	0.233	0.00063
3	1.10	2.8	0.073	0.00019	0.43	0.168	0.00030
4	1.76	6.5	0.117	0.00043	0.50	0.233	0.00063
5	0.51	0.9	0.034	0.00006	0.37	0.091	0.00011
6	1.76	6.5	0.117	0.00043	0.50	0.233	0.00063
7	0.81	1.7	0.054	0.00011	0.39	0.139	0.00020

There are several things to observe about the Extended Life Test. The thruster was operated at the highest power level, TH15, for three of the test segments. The greatest amount of grid erosion takes place at

that level because the impingement current is greatest and the accel grid voltage is most negative. Additionally, the accel grid voltage magnitude was raised during the test in order to prevent electron backstreaming. The accel grid hole diameter was seen to increase on the thruster centerline, creating the need for a more negative accel grid voltage. The extended life test was terminated not because of structural failure, but because the accel grid voltage could no longer be kept sufficiently negative to prevent electron backstreaming.

Individual NSTAR Simulations

Several simulations were performed for apertures operating with the conditions of the greatest throttle point, TH15. Using the hot grid gap (0.014 inches smaller than the cold grid gap) the crossover and perveance limits obtained for four types of gridlets are shown in Figure 7.6. The average beamlet current for TH15 is 0.117 mA, and the peak beamlet current, predicted using a flatness parameter of 0.50, is 0.233 mA. These beamlet currents correspond to perveance fractions of 0.26 and 0.52 respectively.

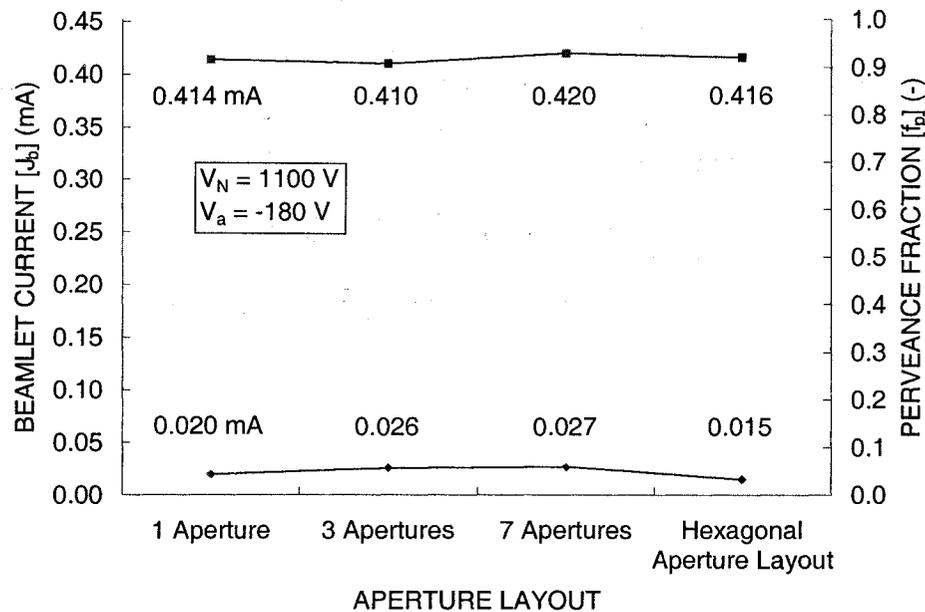


Figure 7.6 Crossover and perveance limits for the NSTAR grid set.

In the case of the NSTAR grid, the sheath upstream of the screen grid moves well away from the apertures near the crossover limit. Figure 7.7 shows gridlets operating at a beamlet current of 0.010 mA, which is below the crossover limit.

$$J_b = 0.01 \text{ mA}, f_p = 0.022$$

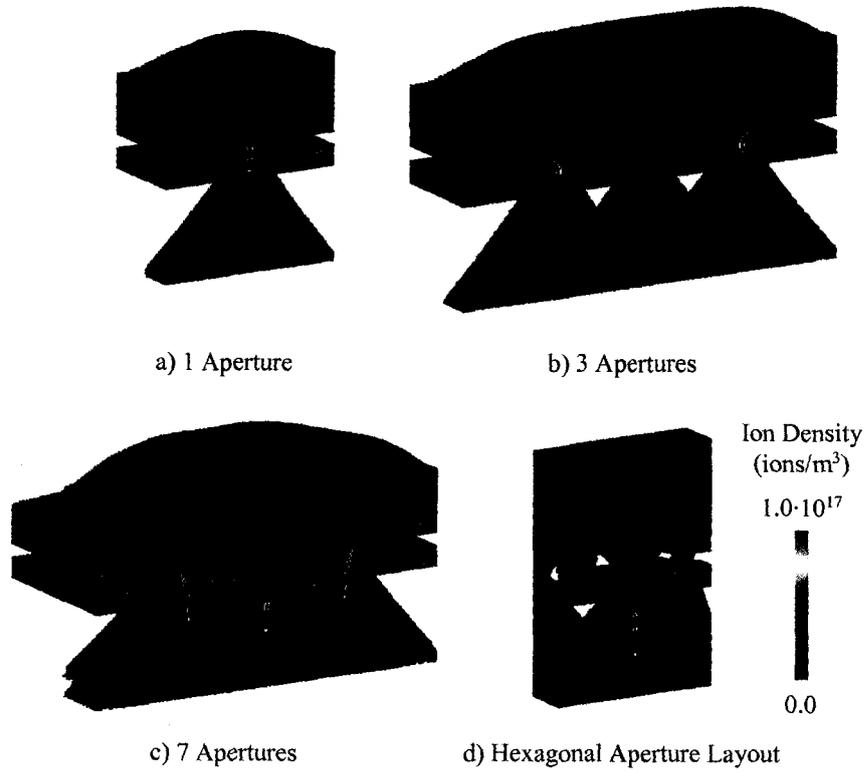


Figure 7.7 NSTAR gridlets operating below the crossover limit.

Cross sections of the beamlets taken through a plane at the downstream edge of the accel grid are shown in Figure 7.8. The scale of the ion density has been reduced greatly to show the details of the cross sections.

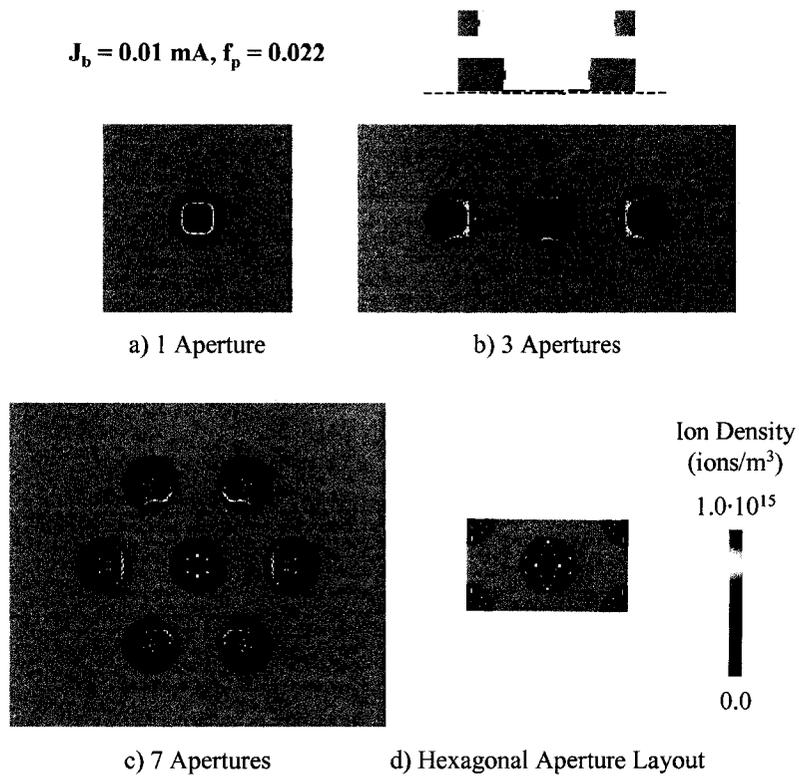


Figure 7.8 Ion density cross section at the downstream plane of the accel grid when operating below the crossover limit.

Figure 7.9 shows additional cross sections taken through a plane 1.27 mm downstream of the accel grid, where beamlets from adjacent holes are interacting. Interesting patterns are seen in the 3 and 7 aperture cases, as well as the hexagonal aperture layout case.

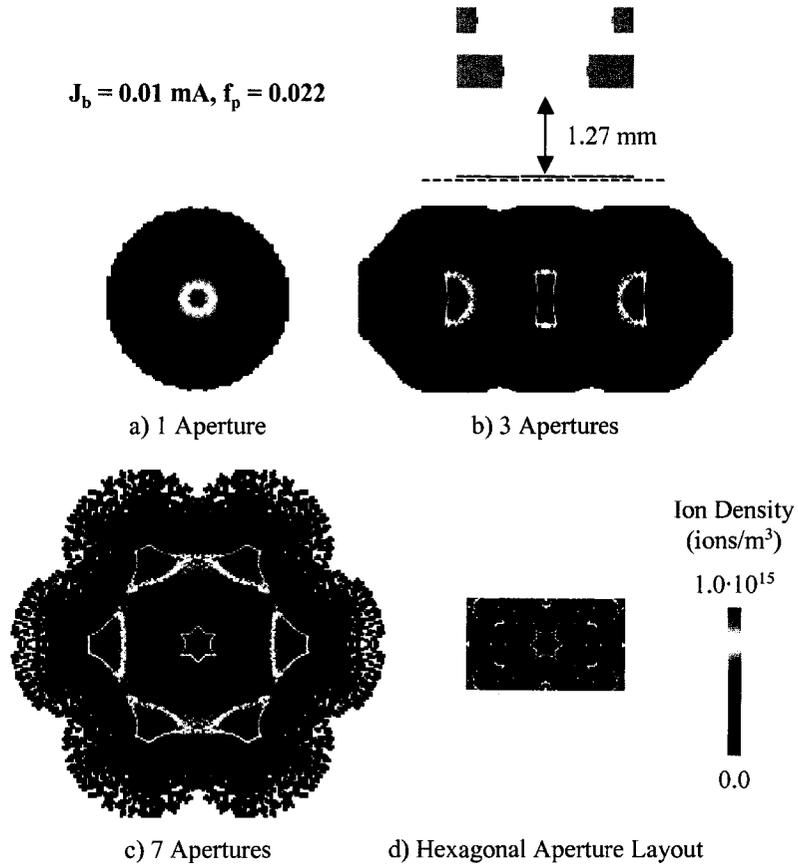


Figure 7.9 Ion density cross section 1.27 mm downstream of the accel grid when operating below the crossover limit.

Figure 7.10 shows ion injection locations on the upstream surface of the simulation volumes colored according to which surface the ion macro particles end up striking. An ion is colored green if it strikes the screen grid, red if it strikes the accel grid, and blue if it passes through the grids. The screen grid location and hole diameter is shown in black. This representation was first depicted by Malone. Note that only a fraction of the true number of particles simulated are shown.

In the 3 aperture case, ions from only two sides of the aperture, away from the end apertures, end up passing through the grids. This results in the rectangular beamlet cross section downstream of the accel grid seen before. It can be seen in general that the area from which ions are extracted is much larger than the screen grid hole.

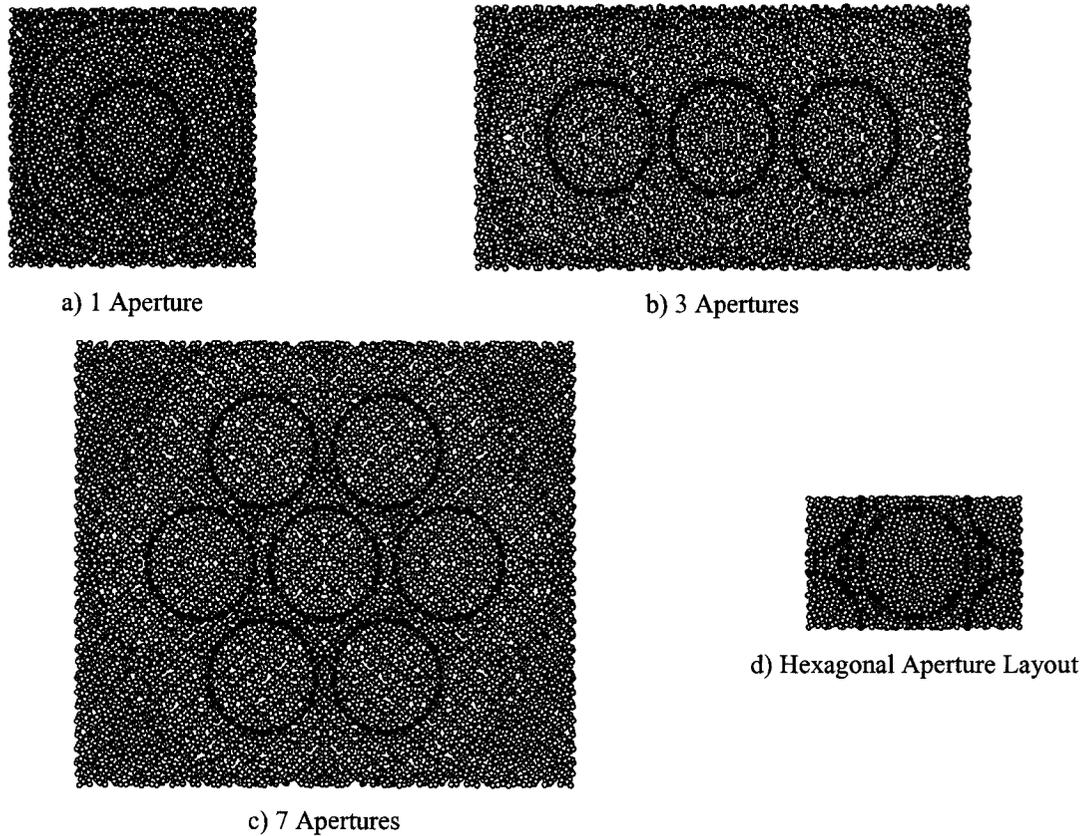


Figure 7.10 Ion injection points colored according to termination surface. $J_b = 0.010$ mA, $f_p = 0.022$.

NSTAR Current Density Profiles

Scans of the current density downstream of an NSTAR ion thruster have been made experimentally at several axial locations. This provides an opportunity for validation of the current density profile program, the intent of which is to determine the beamlet current variation across the thruster face given downstream current density measurements.

The ion macro particle trajectory distributions at several beamlet currents were fit to the current density equation, and the resulting coefficient values are shown in Figure 7.11. The distributions were well modeled with a single curve, whereby the constant A_2 is simply set to zero.

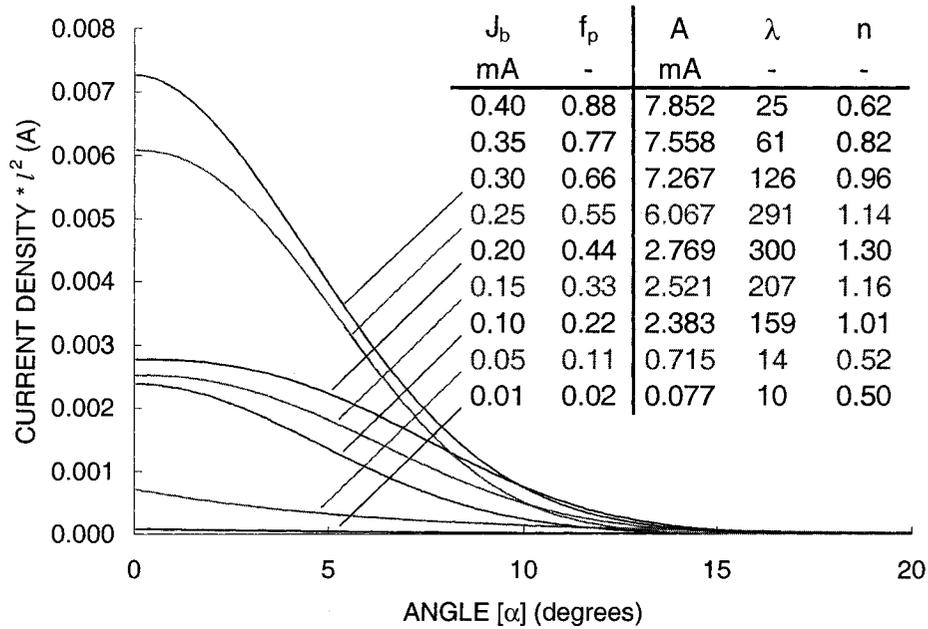


Figure 7.11 NSTAR coefficients for the beamlet current density equation.

The current density profiles measured experimentally were each adjusted to give an integrated beam current of 1.76 A. These are shown in Figure 7.12.

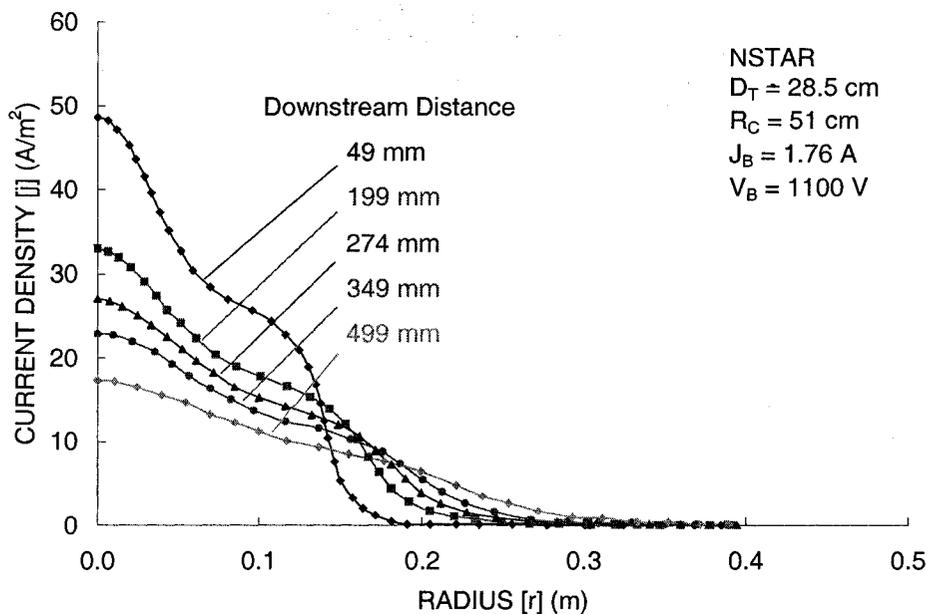


Figure 7.12 Current density profiles downstream of the NSTAR thruster.

The current density profile program was used to find the beamlet current variation on the face of the thruster for each of the five measured current density curves. The beamlet current profiles give “matched” curves that agree well with the measured curves, as shown in Figure 7.13.

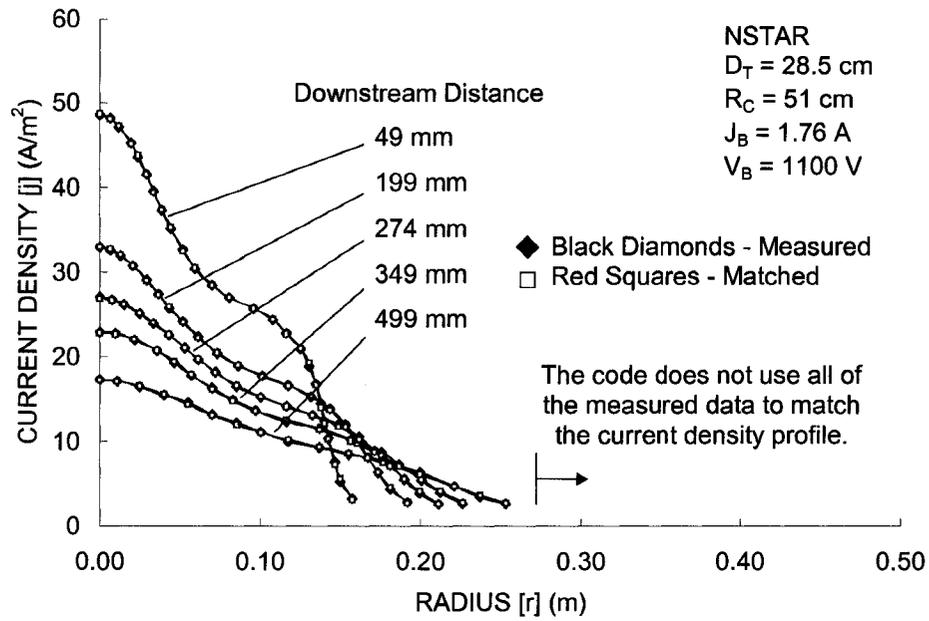


Figure 7.13 Code determined current density profiles.

The beamlet current profiles that correspond to the matched current density profiles are shown in Figure 7.14. As can be seen here, although the current density profiles match well, the beamlet current profiles do not perfectly agree. Ideally, one beamlet current profile would be found that would work well for all five current density profiles.

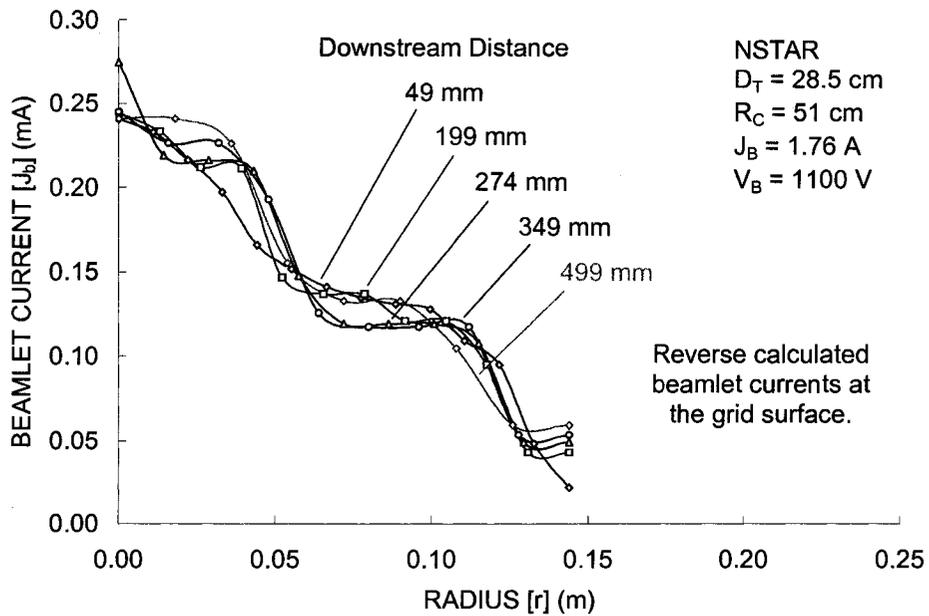


Figure 7.14 Beamlet current profiles corresponding to the matched current density profiles.

The best beamlet current profile is likely the one found using the current density profile closest to the grid face, at 49 mm downstream. Using this profile, the current densities at three downstream locations are shown in Figure 7.15. Here, reasonable agreement with the experimentally measured current density profiles is obtained.

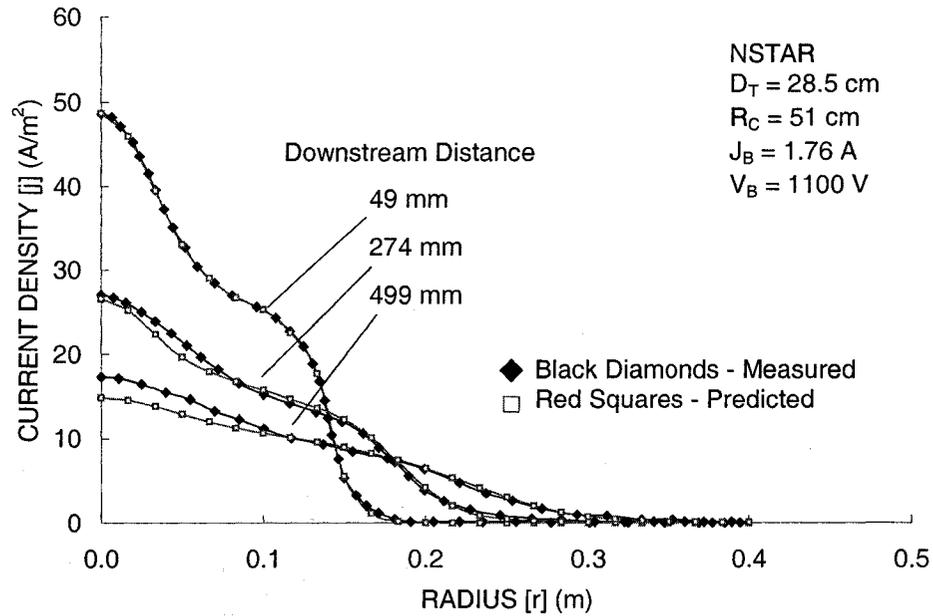


Figure 7.15 Current density profiles determined using the beamlet current profile found 49 mm downstream.

The peak, centerline, beamlet current in the 49 mm downstream profile was 0.240 mA. At TH15, the flatness parameter is 0.50, which implies a peak beamlet current of 0.233 mA. These values are sufficiently similar, within 3 %, to reasonably take the peak beamlet current to be 0.233 mA.

NSTAR Extended Life Test

The ffx code was used to simulate the peak (centerline) and average beamlet current apertures over the course of the 30352 hour NSTAR Extended Life Test. The erosion on the downstream side of the accel grid as a function of time is shown in Figure 7.16.

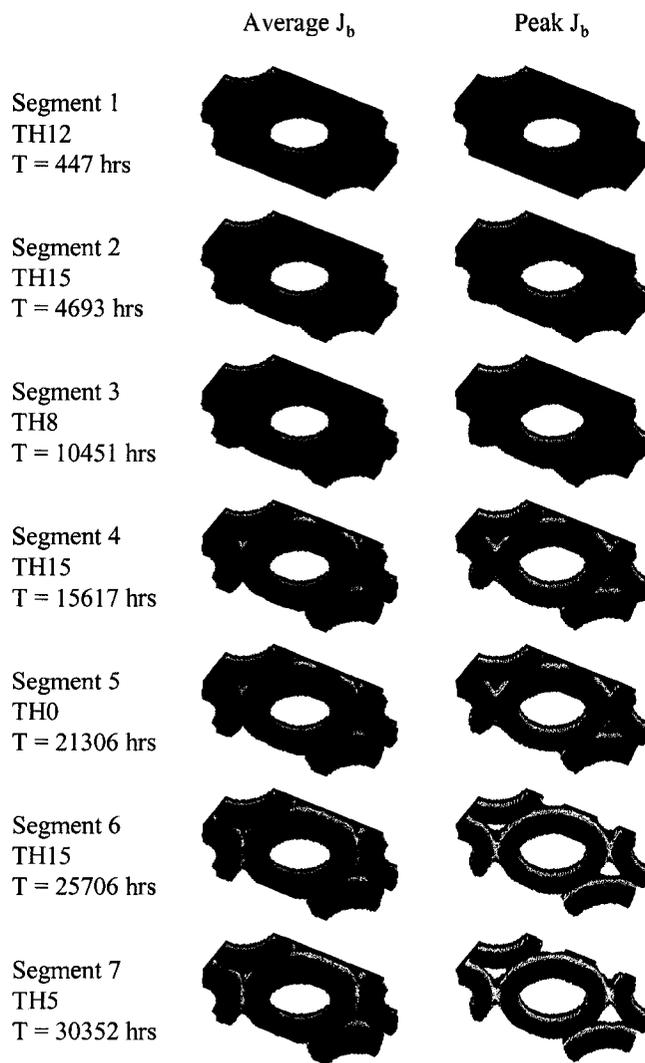


Figure 7.16 Erosion depictions for the average and peak beamlet currents during the Extended Life Test.

One thing immediately noticeable about the centerline aperture is that the pits wore completely through the accel grid. Excellent photographs of both the upstream and downstream sides of the accel grid taken at the conclusion of the life test are compared to the code predicted erosion pattern in Figure 7.17.

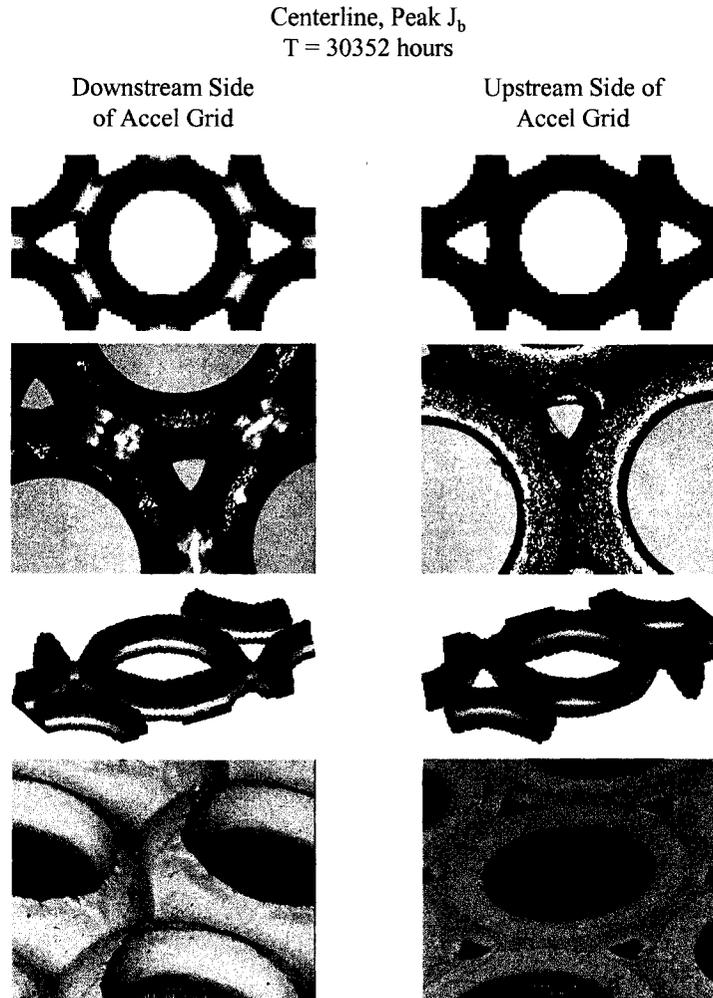


Figure 7.17 Views of the upstream and downstream sides of the centerline aperture at the end of the life test.

An initial attempt at simulating the centerline aperture in the Extended Life Test resulted in too much accel grid erosion. In that simulation, the impingement to beamlet current ratio of the beamlet, J_a/J_b , was taken to be the same as the overall impingement ratio, J_A/J_B . A different approach is to model two components of the accel grid current: a local component that is related directly to the beamlet current, and a global component that is related to charge exchange ion current originating farther downstream of the aperture.

The local accel grid current is taken to be the current which originates upstream of the neutralization surface, close to the accel grid. The global accel grid current comes from charge exchange ions that originate downstream of the neutralization surface. The global current is modeled as flowing to every aperture on the grid evenly. This approximation comes from the idea that charge exchange ions created in the downstream beam plasma will flow in all directions equally. This model tends to decrease the impingement current ratio on the centerline and increase it at the outer radii.

The impingement current to an aperture, J_a , is given by Eq. 7.1. The total impingement current, J_A , is given by the integral in Eq. 7.2. Simulations showed that the local impingement current ratio, $\left. \frac{J_a}{J_b} \right|_{local}$, stayed nearly constant over the beamlet current range of interest. Also, the global current per aperture, J_{global} , is defined as constant. Using these approximations, the integral simplifies to Eq. 7.3 where H is the number of apertures and J_B is the total beam current. The global current per hole can be found using that equation knowing the other parameters from experiment or simulation. Once the global current per hole is known, Eq. 7.1 can be used to obtain the accel grid current for any aperture given a beamlet current. This model was used in the NSTAR life test simulations.

$$J_a = J_b \left. \frac{J_a}{J_b} \right|_{local} + J_{global} \quad \text{Eq. 7.1}$$

$$J_A = 2\pi \frac{1}{A_g} \int_0^{R_T} r J_a dr \quad \text{Eq. 7.2}$$

$$\left. \frac{J_a}{J_b} \right|_{local} \approx \text{constant and } J_{global} = \text{constant}$$

$$J_A = J_B \left. \frac{J_a}{J_b} \right|_{local} + H J_{global} \quad \text{Eq. 7.3}$$

The electron backstreaming limit, or the accel grid voltage required to prevent electron backstreaming, was measured during the Extended Life Test. Similarly, the peak beamlet current was used to predict the backstreaming limit during the ELT simulation. These results are shown in Figure 7.18.

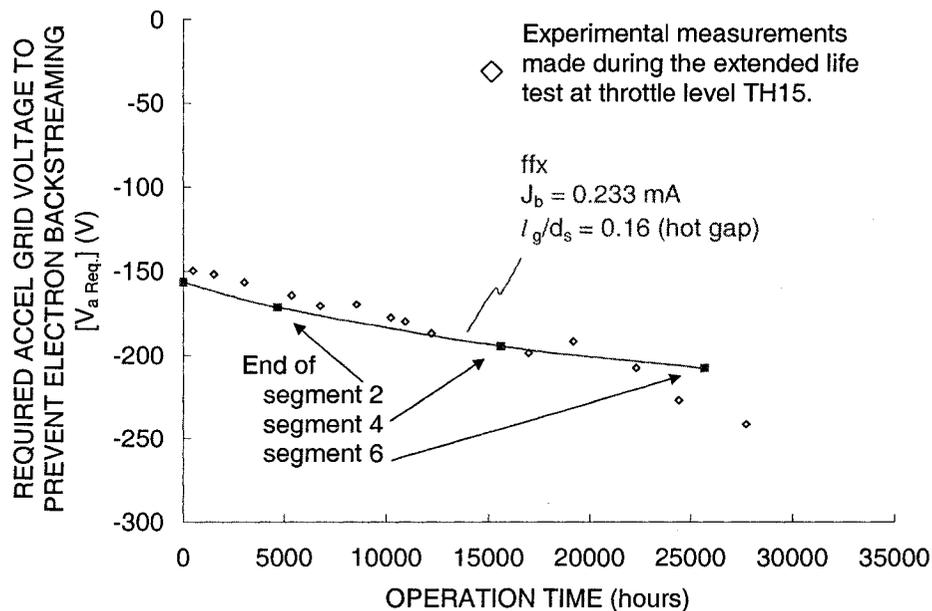


Figure 7.18 Electron backstreaming limit as a function of time at TH15 during the ELT.

Toward the end of the test, the code predicts a required accel grid voltage less negative than what was required experimentally. Post test measurements showed that the cold grid gap reduced during the life test by about 0.2 mm. This suggests that the hot grid gap also might have been reduced during the test to a value less than the spacing used in the code simulation. A more negative accel grid voltage would be required to prevent electron backstreaming in that case.

The pit wear depth was measured during the Extended Life Test, with reliable results being obtained up to the end of the third test segment. Figure 7.19 compares these measurements with the approximate pit and groove wear depths from the simulation of the centerline aperture recorded at the end of each test segment. The simulation showed that the pits wore nearly through the accel grid thickness by the end of the fourth test segment.

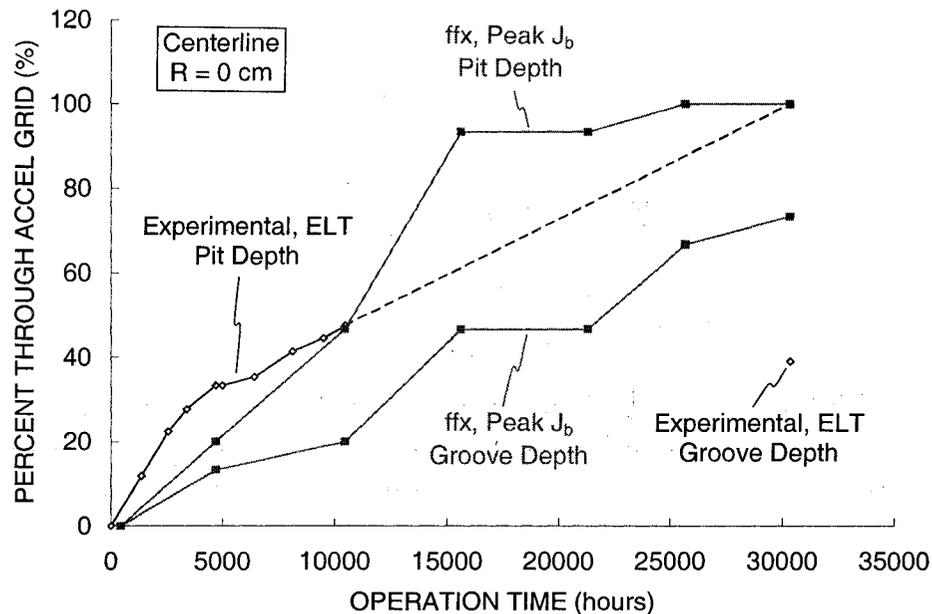


Figure 7.19 Pit and groove wear during the Extended Life Test.

The groove depth seen in the simulation was quite a bit more than what was seen experimentally. In the test, the grooves wore through about 39 % of the accel grid thickness, and in the simulation the grooves wore through about 73 % of the thickness. This difference could be due to the simulation method regarding charge exchange ions originating far downstream of the neutralization surface.

Another thing that can be compared is the accel grid aperture diameter as a function of time, as shown in Figure 7.20. In this figure, the rate of diameter expansion in the centerline aperture is very similar between experiment and simulation. However, the experimentally measured diameter was offset by about 0.06 mm compared to the simulation. The overall change in diameter was similar, 0.24 mm in the experiment compared to 0.20 mm in the simulation.

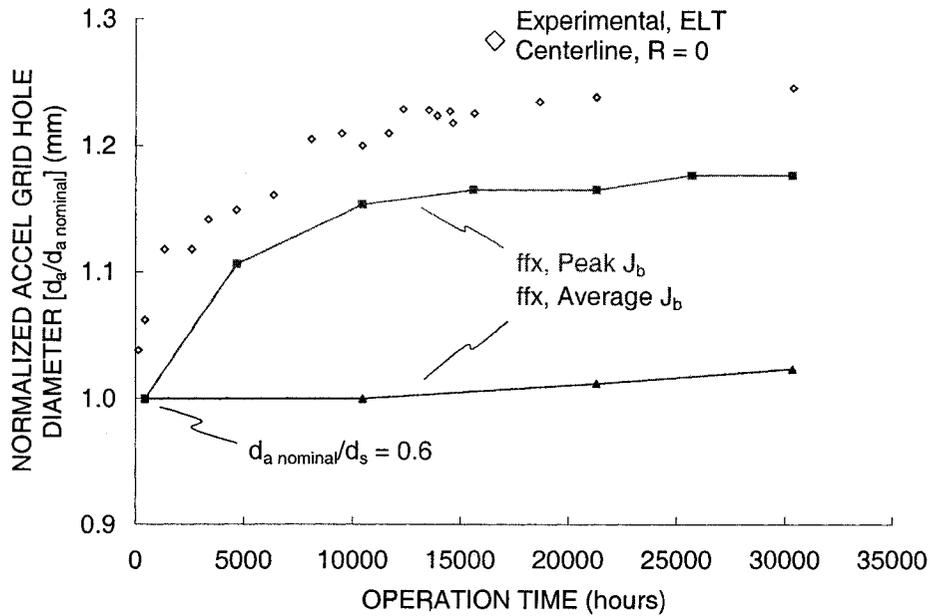


Figure 7.20 Accel grid hole diameter as a function of time during the Extended Life Test.

NSTAR 8200 Hour Life Demonstration Test

The Life Demonstration Test was operated at a single throttle point, very similar to TH15 in the Extended Life Test. Initial simulations comparing the backstreaming limit of the 8200 hour life test to the centerline aperture indicated that the grid spacing might have been larger in the Life Demonstration Test than in the Extended Life Test.

Figure 7.21 shows backstreaming limit comparisons over time for a hot grid spacing 0.010 inches smaller than the cold grid spacing, rather than the 0.014 inch reduction used in the Extended Life Test simulation. The experimental backstreaming limits are about 9 V positive of the centerline aperture simulation predictions.

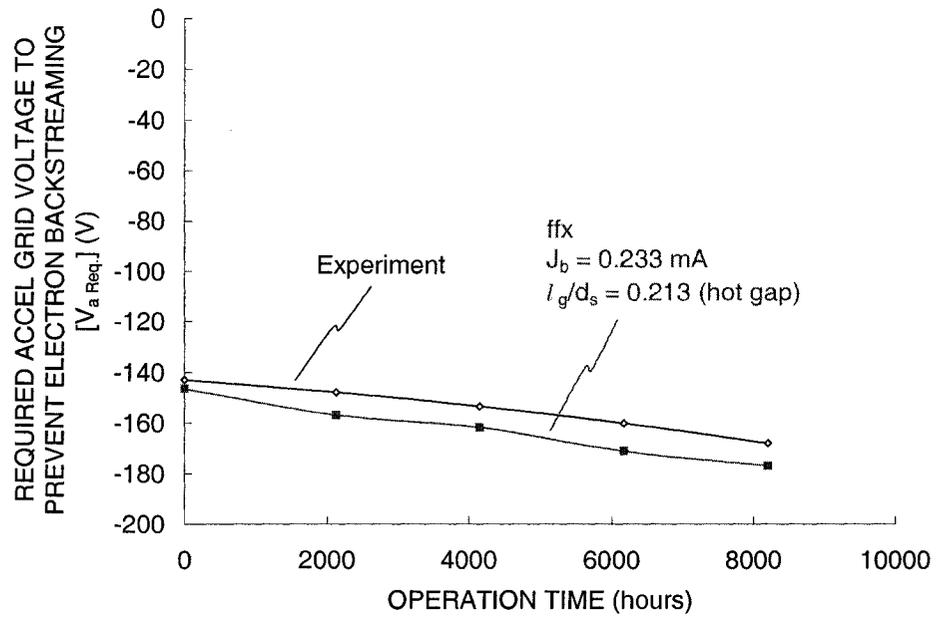


Figure 7.21 Backstreaming limit comparison during the Life Demonstration Test.

Figure 7.22 shows the downstream side of the accel grid at five points in time during the LDT. Expectedly, the pit and groove wear pattern is similar to that observed following the first few segments of the Extended Life Test.

$$J_b = 0.233 \text{ mA}$$

$$\eta_u = 0.91$$

$$J_a/J_b = 0.28 \%$$

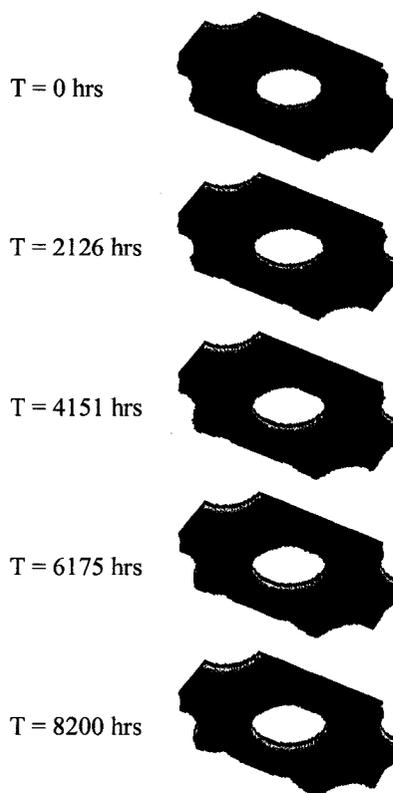


Figure 7.22 Simulation results for the 8200 hour Life Demonstration Test.

A comparison of the centerline pit and groove depths at the conclusion of the LDT is given in Table 7.7. As with the ELT, the simulated groove depth was greater than observed experimentally, while the pit depth was accurate.

Table 7.7 Pit and groove depth measurements for the Life Demonstration Test.

	Pit (% of t_a)	Groove (% of t_a)
Experimental Depth	45.2	8.2
Simulation Depth	50.0	28.6

The centerline accel grid hole diameter from experiment and simulation are compared in Table 7.8. Very good agreement is seen in this case, with the final hole diameters being similar to within 2.6 %.

Table 7.8 Centerline accel grid hole diameter following the LDT.

	Accel Grid Hole Diameter Final $d_a/d_{a \text{ nominal}}$
Experimental d_a	1.19
Simulation d_a	1.16
% Difference	-2.6

Finally, Table 7.9 compares the screen grid film thickness on the downstream side of the screen grid after the LDT. Good agreement is obtained, with the simulation giving a film thickness about 8.7 % lower than experimentally measured.

Table 7.9 Screen grid film thickness for the centerline aperture of the LDT.

	Screen Grid Film Thickness (μm)
Experimental Film Thickness	8.81
Simulation Film Thickness	8.04
% Difference	-8.7

Charge exchange ion production is broken up into four regions in Figure 7.23. Region 1 is the region upstream of the accel grid exit plane, Region 2 is the region downstream of the accel grid and upstream of the neutralization surface, and Region 3 is the region between the neutralization surface and an arbitrarily chosen plane further downstream. Regions 1 through 3 consist of charge exchange ions produced the normal way, by a volumetric production rate equation which is a product of the ion density, neutral density, ion velocity, and cross section per cell.

Region 4 is considered to consist of charge exchange ions produced far downstream. In this region, charge exchange ion production is uniform, and the initial ion vectors are directed entirely axially upstream. The amount of charge exchange ion production is chosen to agree with the experimentally measured accel grid impingement current.

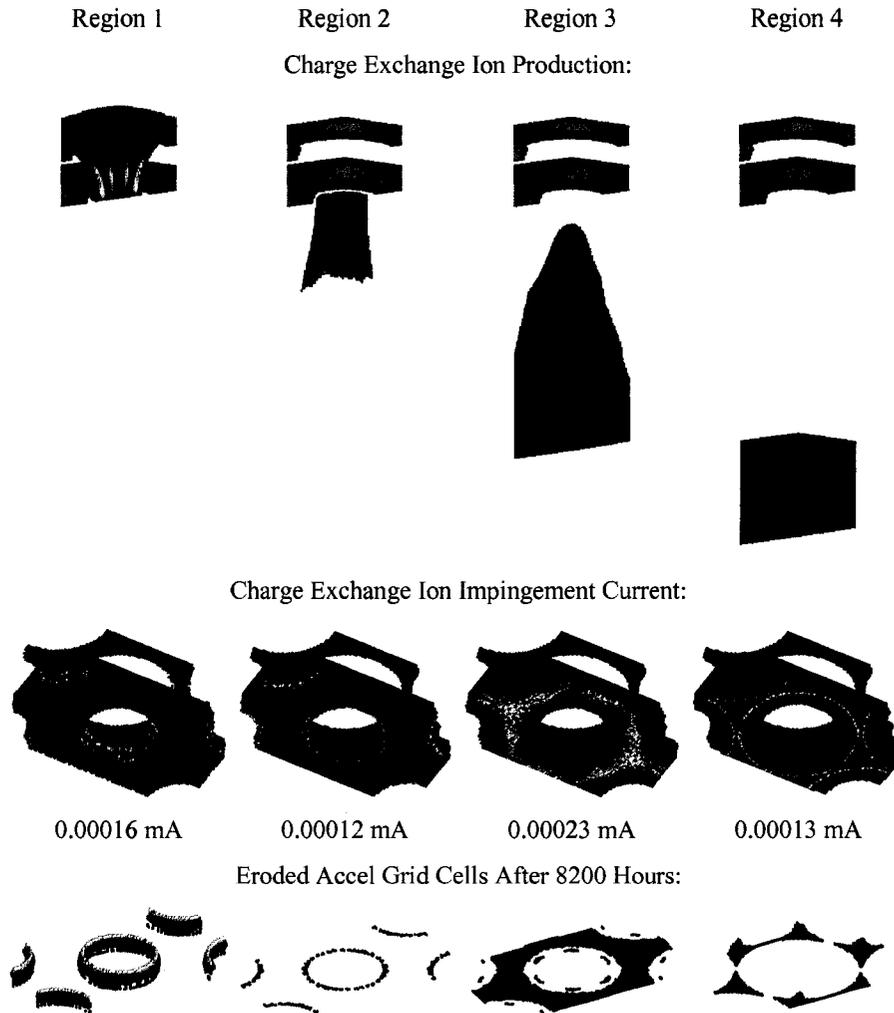


Figure 7.23 Accel grid erosion from charge exchange ions originating in different regions of the beamlet.

The charge exchange ions originating within region 1 primarily erode the accel grid hole barrel. In region 2, the charge exchange ions mostly strike the downstream accel grid hole edge. The charge exchange ions in region 3 erode away the downstream side of the accel grid. At this beamlet current, these ions erode away the grooves more than the pits. The charge exchange ions in region 4 erode away the pits much more than the grooves.

7.4 NEXT

NASA's Evolutionary Xenon ion Thruster, or NEXT, is a larger, higher power advancement of the NSTAR ion thruster. The NEXT thruster uses 36 cm (active diameter) molybdenum grids, which results in about 60 % more ion extraction area than NSTAR. Using similar current densities, the beam current is

doubled to 3.52 A from 1.76 A. Also, using a maximum net accelerating voltage of 1800 V, the maximum specific impulse is increased to roughly 4050 s.

A 2000 hour wear test has been completed using a 40 cm diameter version of the NEXT thruster [Kamhawi, Soulas]. Currently, a Long Duration Test is being conducted at the NASA Glenn Research Center using a 36 cm diameter thruster that will demonstrate the ability to process 450 kg of xenon propellant [Frandina, Soulas]. The baseline mission requirement for the thruster is 185 kg, while the normal processing objective is 300 kg. As of June, 2005, the thruster had operated for 493 hours, processing 10.2 kg of xenon.

The NEXT grid geometry is identical to the NSTAR grid geometry with the exception that the accel grid thickness, t_a , was increased by 50 %. The NEXT grids are referred to as Thick Accelerator Grid, or TAG, optics. Additionally, improved manufacturing techniques have been used to reduce the height of the cusps on the apertures. The NEXT grid geometry is given in Table 7.10.

Table 7.10 NEXT grid geometry.

Parameter	Symbol	mm/mm
Aperture center to center spacing	l_{cc}/d_s	1.16
Screen grid hole diameter	d_s/d_s	
Inner diameter, at cusp tips		1
Outer diameter, at entrance/exit		1.053
Screen grid thickness	t_s/d_s	0.2
Grid spacing (cold)	l_g/d_s	
Cold gap, not operating		0.347
Hot gap, during operation		0.16
Accel grid hole diameter	d_a/d_s	
Inner diameter		0.6
Outer diameter		0.653
Accel grid thickness	t_a/d_s	0.4

Several ion optics numerical simulation codes, summarized in Table 7.11, are in currently in use. All of these codes have been used to simulate NEXT thruster grid apertures [Emhoff, Kafafy, Malone, Petillo].

Table 7.11 Ion optics numerical simulation codes.

Code Name	Place of Development / Primary Use	Simulation Type	Simulation Volume
ffx	Colorado State University	Steady State / Ray	3D, Rectangular Prism
igx		Tracing	3D, Wedge
CEX2D	Jet Propulsion Laboratory	Steady State / Ray	2D, Rectangle
CEX3D		Tracing	3D, Wedge

erode	University of Michigan	Time Based / Particle In Cell	2D, Rectangle
HG-IFE-PIC	Virginia Polytechnic Institute & State University	Time Based / Particle In Cell	3D, Rectangular Prism
MICHELLE	Science Applications International Corporation / NASA Glenn Research Center	Steady State / Ray Tracing (time based also available)	3D, Rectangular Prism

It is difficult to directly compare all of the simulation codes because all of the codes use different algorithms. Additionally, different researchers may use different input values when simulating the same problem.

ffx, igx, CEX2D, CEX3D, and erode

The ffx, igx, CEX2D, CEX3D, and erode codes were all used to simulate the NEXT grid impingement and backstreaming limits utilizing as many of the same input parameters as possible, listed in Table 7.12.

Table 7.12 NEXT numerical code comparison parameters.

Parameter	Symbol	Value
Net Accelerating Voltage	V_N (V)	1800
Discharge Voltage	V_d (V)	25
Accel Grid Voltage	V_a (V)	-210
Beam Plasma Potential	(V)	10
Discharge Propellant Utilization Efficiency	η_u (%)	90.0
Double To Single Current Ratio	J^{++}/J^+ (%)	0.0
Electron Temperature	T_e (eV)	
Upstream		6.0
Downstream		1.0
Grid Spacing	l_g/d_s (mm/mm)	0.347

Impingement limit predictions from the five codes are shown in Figure 7.24. With the exception of the CEX2D code, the perveance limit predictions are all very similar. The ffx and erode codes also show a crossover limit at a very low beamlet current.

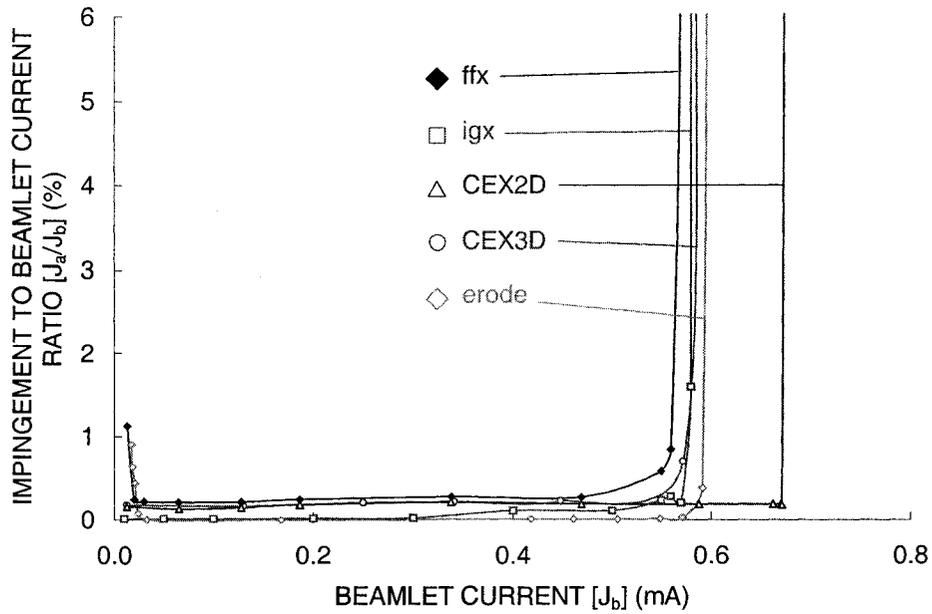


Figure 7.24 NEXT impingement limit predictions.

Predictions of the centerline saddle point potential, $V_{\text{SaddlePoint}}$, as a function of beamlet current from the five codes are shown in Figure 7.25. The ffx and igx codes show a slightly greater margin against electron backstreaming compared to the CEX2D, CEX3D, and erode codes.

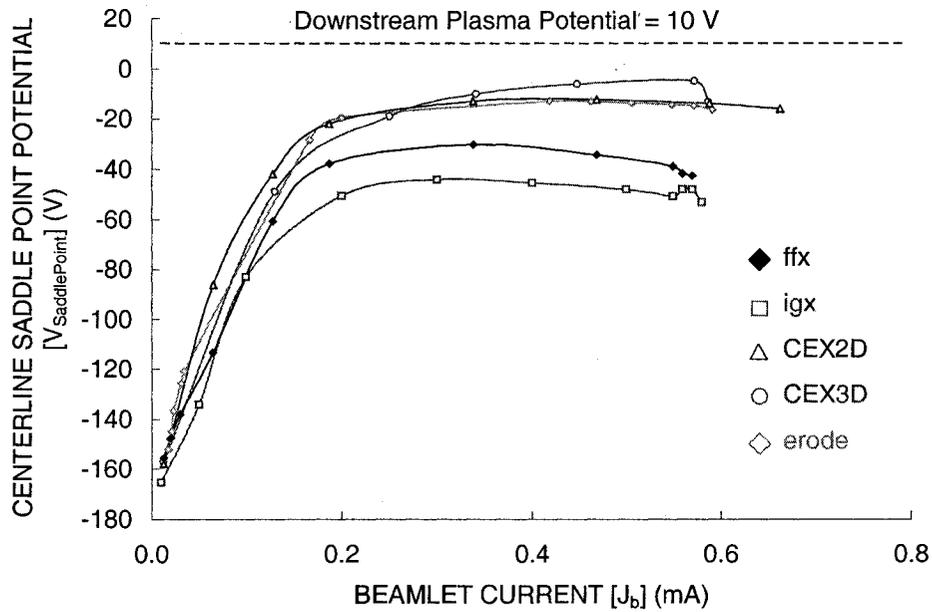


Figure 7.25 NEXT centerline saddle point potential predictions.

Grid ion transparencies are shown in Figure 7.26 as predicted by the four of the five simulation codes. Again, the ffx and igx codes predict slightly lower ion transparencies in general compared to the CEX2D and CEX3D codes.

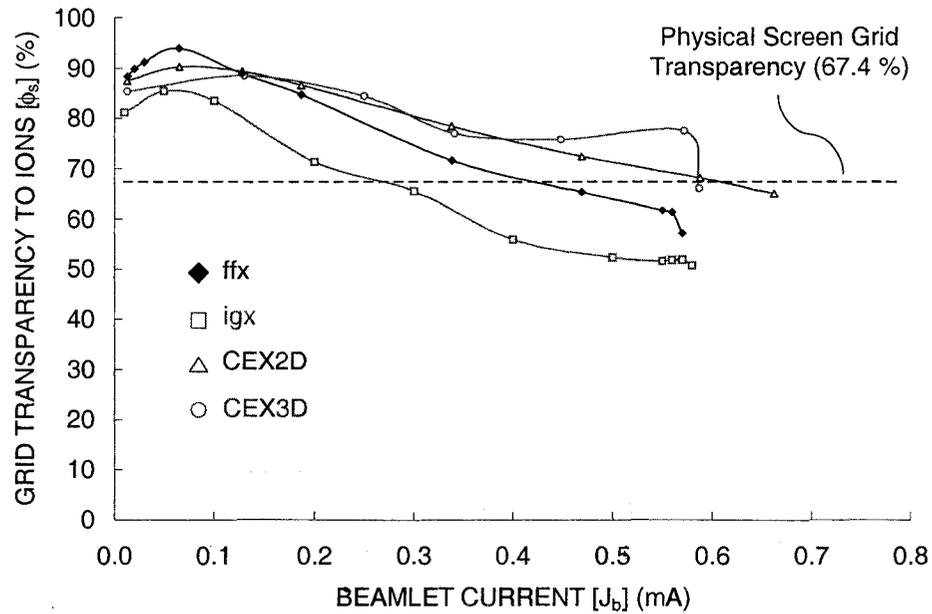


Figure 7.26 NEXT ion transparency predictions.

As a final comparison, accel grid erosion rates as a function of beamlet current are shown in Figure 7.27 as calculated using the ffx, CEX2D, and CEX3D optics codes. It is encouraging that the ffx and CEX3D codes predict similar accel grid hole barrel and downstream erosion rates over a range of beamlet currents. It is noted that the ffx erosion rate suddenly increases at the greatest beamlet current in this figure as it is very close to the onset of direct ion impingement.

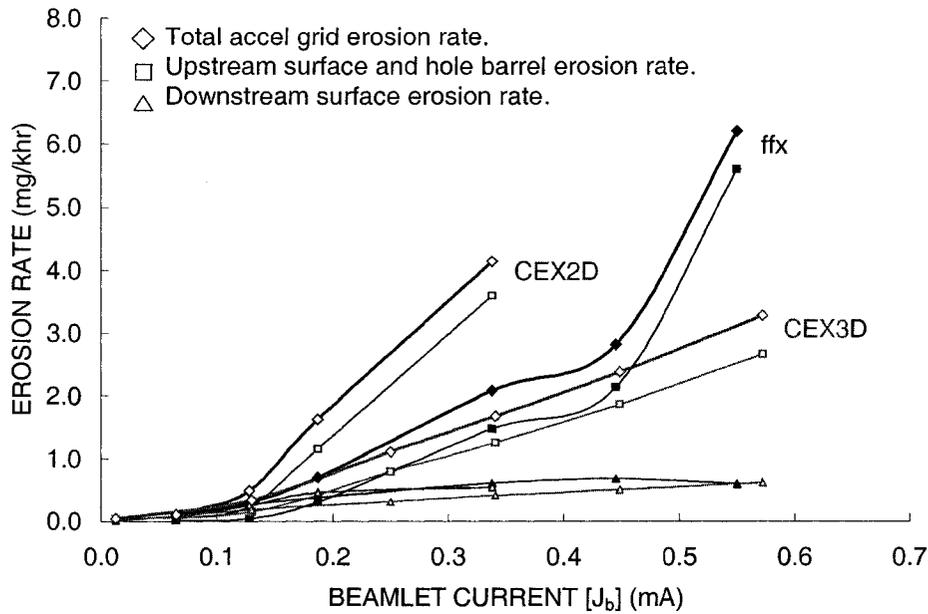


Figure 7.27 ffx and CEX3D code erosion rate predictions on the accel grid hole barrel and downstream surfaces.

ffx and HG-IFE-PIC

The ffx and HG-IFE-PIC codes were compared using a grid setup similar to NEXT, as summarized in Table 7.13. The primary difference is a smaller screen grid hole diameter.

Table 7.13 NEXT grid geometry and operating conditions for the comparison of the ffx and HG-IFE-PIC codes.

Parameter	Symbol	mm/mm
Aperture center to center spacing	$l_{cc}/d_{s\text{NEXT}}$	1.165
Screen grid hole diameter	$d_s/d_{s\text{NEXT}}$	0.793
Screen grid thickness	$t_s/d_{s\text{NEXT}}$	0.2
Grid spacing	$l_g/d_{s\text{NEXT}}$	0.347
Accel grid hole diameter	$d_a/d_{s\text{NEXT}}$	0.573
Accel grid thickness	$t_a/d_{s\text{NEXT}}$	0.4
Parameter	Symbol	Value
Net Accelerating Voltage	V_N	1800 V
Screen Grid Voltage	V_s	1780 V
Accel Grid Voltage	V_a	-210 V
Beam Plasma Potential		0 V

Electron Temperature

T_e

Upstream

5.0 eV

Downstream

1.5 eV

Propellant Temperature

300 K

Figure 7.28 compares the crossover and perveance limits calculated using the ffx and HG-IFE-PIC codes for the modified NEXT grid geometry. Also shown in this figure is experimental gridlet data obtained near the crossover limit using a 37 aperture gridlet. Both codes show consistent perveance limits in terms of the number of apertures being simulated, while the ffx code perveance limits are somewhat lower than those predicted by the HG-IFE-PIC code.

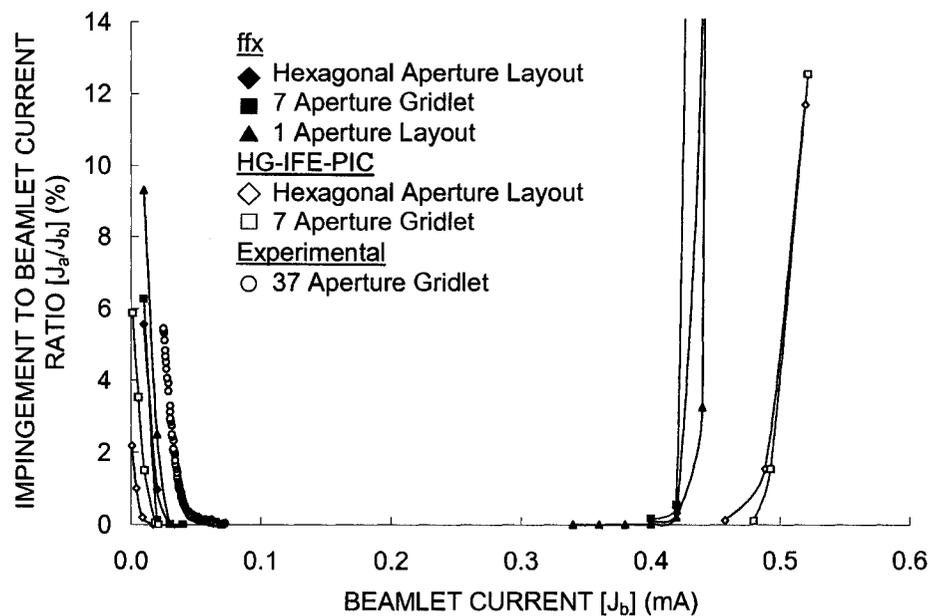


Figure 7.28 Crossover and perveance limits predicted by the ffx and HG-IFE-PIC codes.

Figure 7.29 shows the crossover limits of the previous figure in greater detail. The ffx code crossover predictions are greater than the HG-IFE-PIC code predictions, but are less than the experimentally determined crossover limit. Also, while the 7 aperture ffx data falls between the hexagonal aperture layout and 1 aperture data at a beamlet current of 0.01 mA, it falls below the hexagonal aperture layout data at 0.02 mA. This is in contrast to the HG-IFE-PIC code which shows the 7 aperture curve lying entirely above the hexagonal aperture layout curve. Investigations into several simulation parameters did not explain the trend in the 7 aperture data. The 1 aperture ffx data did lie entirely above the hexagonal aperture layout data as anticipated. It is useful to note that the rate of rise of the impingement current is similar between the codes and the experimental data.

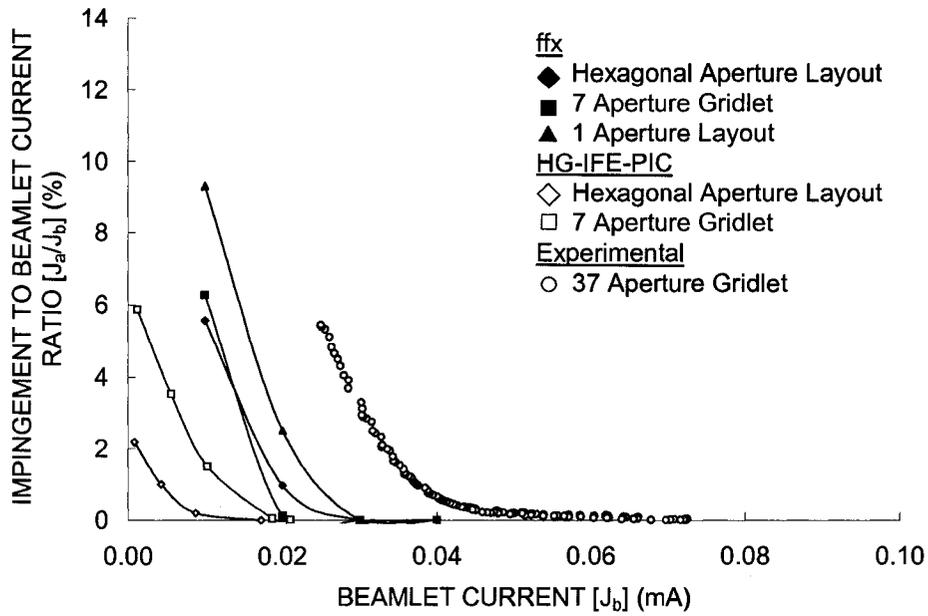


Figure 7.29 Crossover limits for the modified NEXT grid geometry.

Finally, a comparison of the centerline saddle point potential as a function of accel grid voltage is shown in Figure 7.30. In this case, the upstream plasma density was held constant, at $1.0 \cdot 10^{17}$ ions/m³, and single aperture gridlets were simulated. The code agreement at this plasma density is very good. It is noted that this plasma density corresponded to a relatively low beamlet current, near 0.08 mA, where all of the codes improve in agreement with regard to saddle point potential in general.

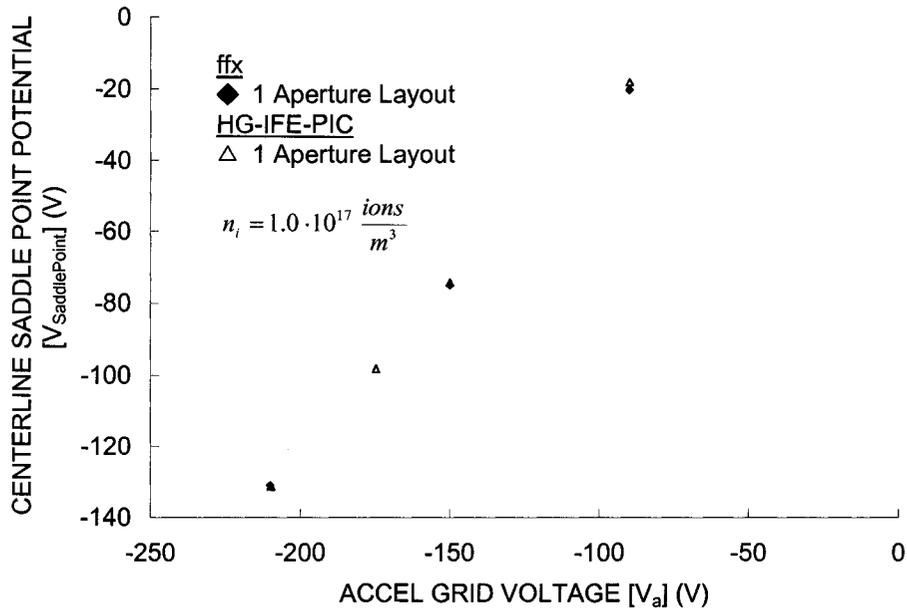


Figure 7.30 Agreement between the ffx and HG-IFE-PIC code predictions of saddle point potential.

ffx and MICHELLE

The ffx and MICHELLE codes were compared using a modified NEXT grid geometry that corresponds to the outer edge geometry of the 40 cm thruster used in the 2000 hour wear test. Malone first used the MICHELLE code to study crossover erosion that took place near the thruster edge, presumably during operation at a low beam current. The test geometry and conditions at radii of 180, 195, and 199 mm are given in Table 7.14. The beamlet currents in this table were estimated using the given current densities and a ffx code calculated ion transparency of 86 %.

Table 7.14 NEXT grid geometry and operating conditions for the comparison of the ffx and MICHELLE codes.

Parameter	Symbol	Radial Position		
		180 mm	195 mm	199 mm
Screen grid hole diameter	$d_s/d_{s\text{NEXT}}$, inner (mm/mm)		0.971	
	$d_s/d_{s\text{NEXT}}$, outer (mm/mm)		1.055	
Grid spacing (hot)	$l_g/d_{s\text{NEXT}}$ (mm/mm)		0.2	
Accel grid hole diameter	$d_a/d_{s\text{NEXT}}$, inner (mm/mm)	0.535	0.520	0.509
	$d_a/d_{s\text{NEXT}}$, outer (mm/mm)	0.667	0.651	0.640
Beam current	J_B (A)		1.20	
Net Accelerating Voltage	V_N (V)		1800	
Screen Grid Voltage	V_s (V)		1775	
Accel Grid Voltage	V_a (V)		-250	
Beam Plasma Potential	(V)		15	
Electron Temperature	T_e , upstream (eV)		6.0	
	T_e , downstream (eV)		1.0	
Current density	j (mA/cm ²)	0.423	0.161	0.114
Ion transparency	ϕ (%)		86	
Beamlet current	J_b (mA)	0.015	0.006	0.004

First, grids with apertures in a hexagonal layout pattern were simulated, the results of which are shown in Figure 7.31. Direct, crossover, ion impingement can be seen at all three radii. In the last set of photos in this figure, the grid cells that have been completely eroded away due to crossover impingement after 500 hours of operation are shown colored according to their axial location with respect to the accel grid.

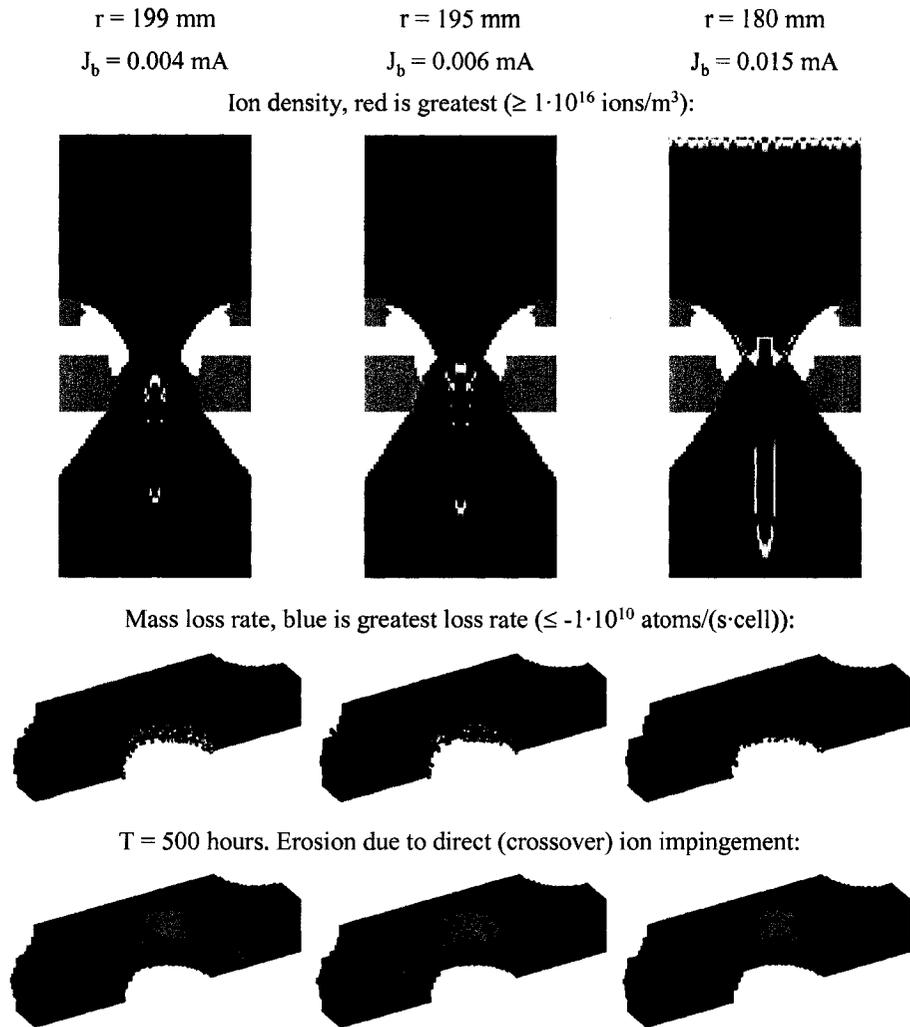


Figure 7.31 Hexagonal aperture layout grids showing crossover impingement at three radial positions.

Next, grids with 1 and 7 aperture patterns were simulated at the greatest radial position of 199 mm in order to better approximate edge aperture erosion. The results of these simulations are shown in Figure 7.32. In going from the hexagonal grid to the 7 aperture grid to the 1 aperture grid, the impingement to beamlet current ratio, J_a/J_b , increases.

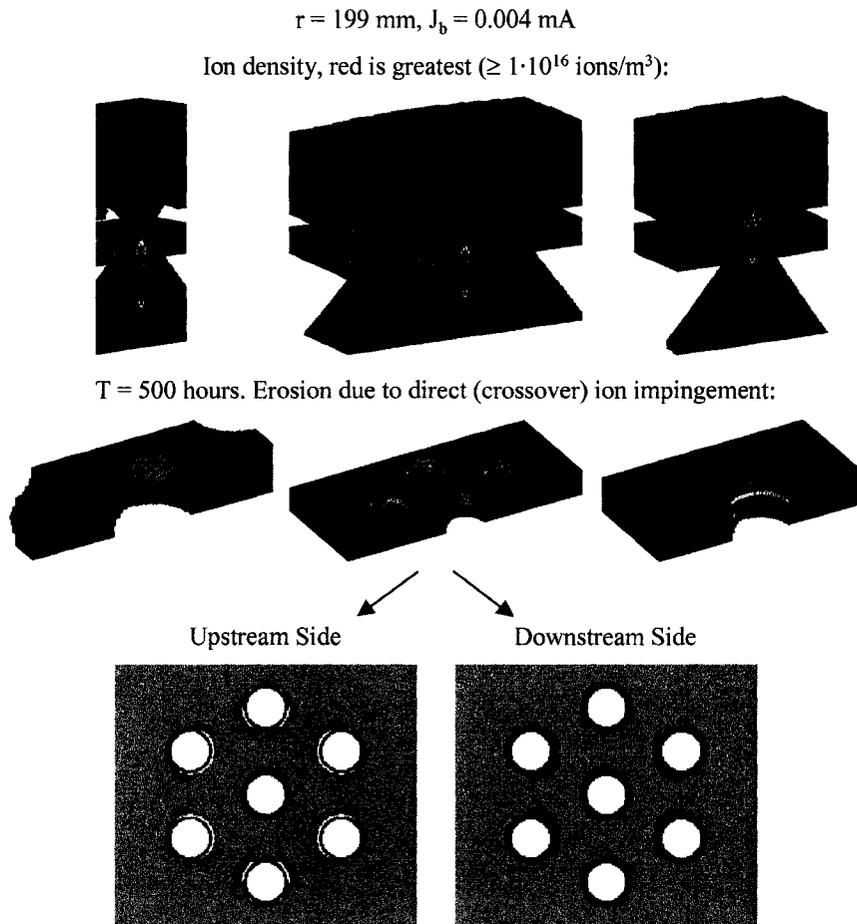


Figure 7.32 Crossover ion erosion among the 1, 7, and hexagonal aperture layout grids.

It is important to remember that symmetry conditions are used on the sides of each of these simulations. Edge aperture erosion is best approximated by the 7 aperture simulation. Figure 7.33 shows the 7 aperture grid reflected several times to show the actual geometry being simulated.

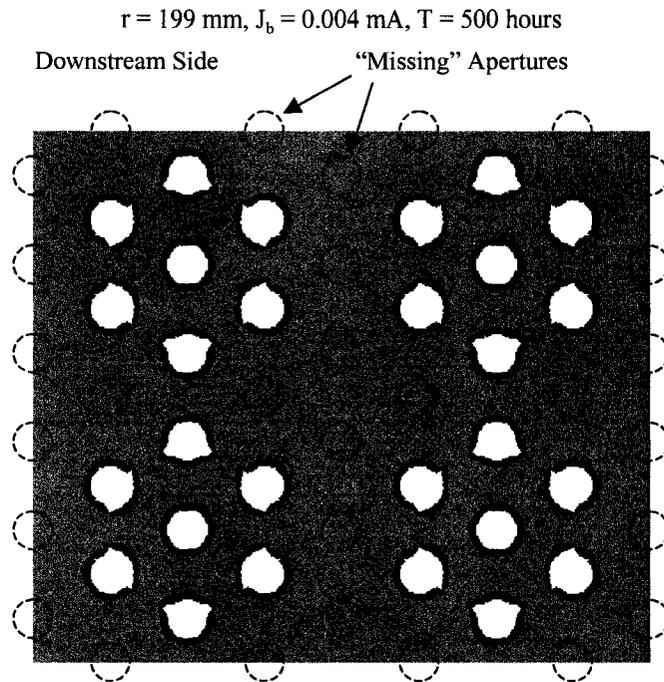


Figure 7.33 Reflection of the 7 aperture grid illustrating the actual geometry being simulated.

Figure 7.34 shows the erosion at three radial positions that occurred during the NEXT thruster 2000 hour wear test. The erosion at the 199 mm radial position is similar to the erosion predicted by the ffx code for the "edge" aperture in the 7 aperture grid, where three of the six adjacent apertures are missing. In general, the impingement current patterns found using the ffx code are similar to those found using the MICHELLE code.

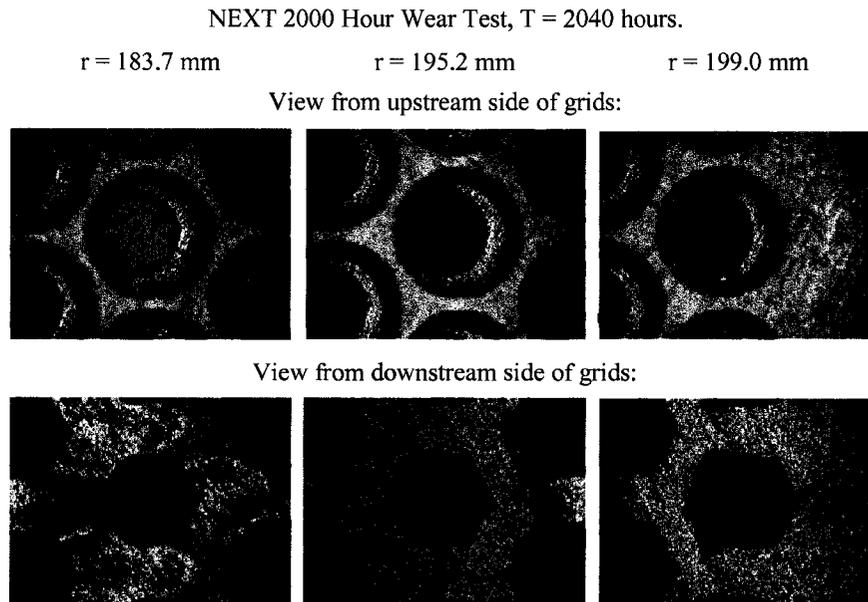


Figure 7.34 Erosion near the edge of the grids following the NEXT 2000 hour wear test.

Table 7.15 gives the impingement to beamlet current ratios, J_a/J_b , that were seen in the code simulations. These ratios are averaged over the duration of the 500 hour simulations.

Table 7.15 Average impingement to beamlet current ratios from the ffx code simulations for the NEXT edge aperture study.

Grid Type	Radial Position (mm)	Impingement to Beamlet Current Ratio [J_a/J_b] (%)
Hexagonal	180	2.4
Hexagonal	195	4.2
Hexagonal	199	4.6
Hexagonal	199	4.6
7 Aperture	199	13.5
1 Aperture	199	20.2

It should be pointed out that simulating erosion due to crossover impingement is troublesome given the very high ion incidence angles that are involved. In order to maintain non-zero erosion rates in unusual situations where the surface normal vectors are highly orthogonal to the ion trajectories, the ion incidence angles are not allowed to be greater than 1.3 radians (74 degrees) in terms of the sputtering yield equations. This provides a minimum sputter yield, ensuring that erosion can take place, if only at a small rate, in all cases.

NEXT 2000 Hour and Long Duration Tests

The operating conditions, as simulated, for the 2000 hour life test (using a 40 cm NEXT thruster) and the Long Duration Test (using a 36 cm thruster) are given in Table 7.16. Most conditions are the same for both tests. The grid spacing, l_g , and beamlet current, J_b , values are uncertain. The beamlet currents are simply derived from the beam current, flatness parameter, and beam diameter values. The grid spacing for the 2000 hour wear test was used to obtain good agreement with the experimentally observed erosion patterns. The grid spacing for the LDT was used to be in good agreement with the experimentally measured electron backstreaming limit at beginning of life.

Table 7.16 NEXT operating conditions for the 2000 hour and Long Duration life tests.

Parameter	Symbol	Value
Grid spacing (hot)		
2000 hour (40 cm)	l_g/d_s NEXT (mm/mm)	0.267
LDT (36 cm)	l_g/d_s NEXT (mm/mm)	0.16
Accel grid hole diameter		
2000 hour (40 cm)	d_a/d_s NEXT, inner (mm/mm)	0.667

	d_a/d_s NEXT, outer (mm/mm)	0.72
	d_a/d_s NEXT, inner (mm/mm)	0.6
LDT (36 cm)	d_a/d_s NEXT, outer (mm/mm)	0.653
Net accelerating voltage	V_N (V)	1800
Screen grid voltage	V_s (V)	1775
Accel grid voltage	V_a (V)	-210
Beam plasma potential	(V)	10
Electron Temperature	$T_{e, \text{upstream}}$ (eV)	6.0
	$T_{e, \text{downstream}}$ (eV)	1.0
Beam current	J_B (A)	3.52
Flatness parameter	F	0.703
Holes		
2000 hour (40 cm)	H	29722
LDT (36 cm)	H	24058
Centerline (peak) beamlet current		
2000 hour (40 cm)	J_b (mA)	0.168
LDT (36 cm)	J_b (mA)	0.208
Accel grid current		
2000 hour (40 cm)	J_A (mA)	12.5
LDT (36 cm)	J_A (mA)	14.2
Flow Rates		
Main	(sccm)	49.64
Discharge chamber cathode	(sccm)	4.87
Neutralizer cathode		
2000 hour (40 cm)	(sccm)	5.16
LDT (36 cm)	(sccm)	4.01

Figure 7.35 compares the pit and groove erosion patterns of the centerline aperture following the NEXT 2000 hour wear test found experimentally and through simulation. Higher than usual mesh resolution was used in the simulations to more accurately resolve the small amount of erosion that occurs. Note that very little erosion was seen on the aperture barrel region at any radius over this time span.

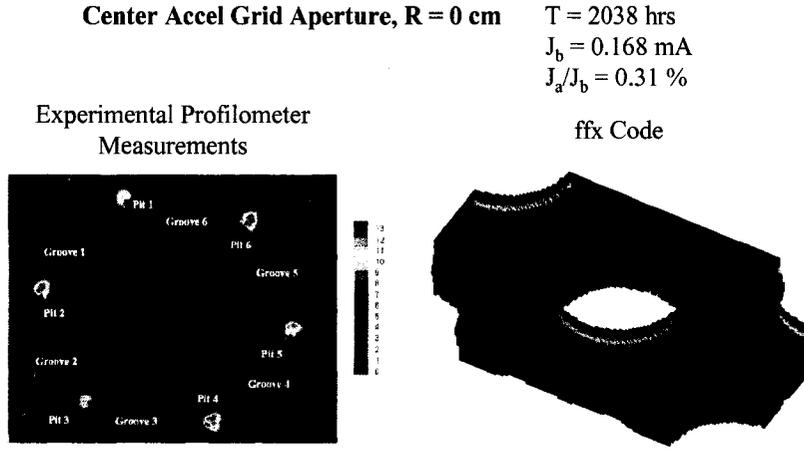


Figure 7.35 Erosion patterns from experiment and simulation for the center aperture of the NEXT 2000 hour wear test.

Figure 7.36 compares the pit and groove depths measured experimentally, with a laser profilometer, to depths found through simulation for apertures at three radii: centerline (0 cm), one-fourth radius (5 cm), and half radius (10 cm). The agreement is good at all three radii for the simulated beamlet currents and grid spacing. Note that the erosion patterns were seen to be quite sensitive to both beamlet current and spacing in this region.

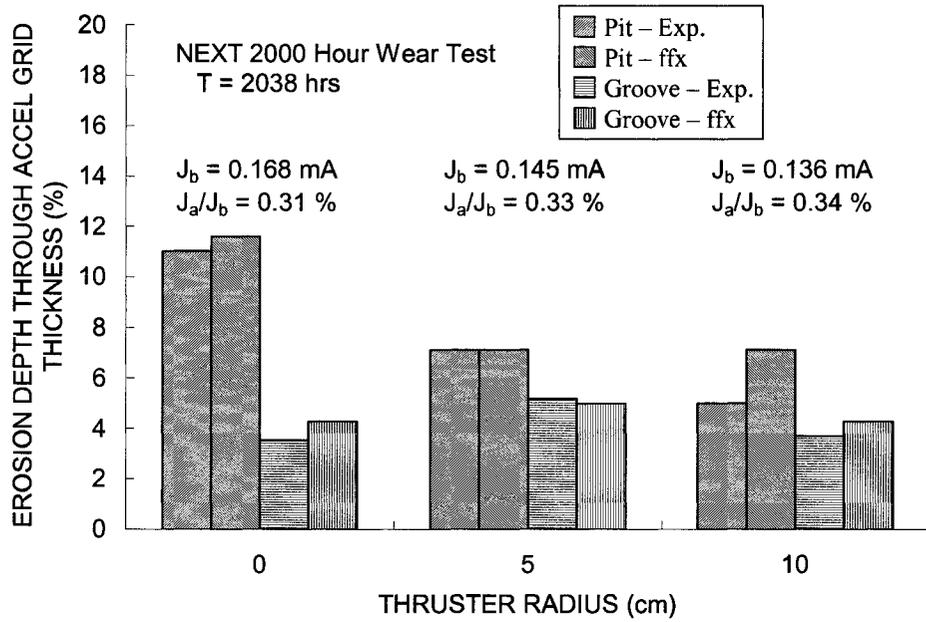
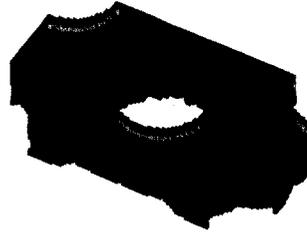


Figure 7.36 Pit and groove depths at three radii for the NEXT 2000 hour wear test.

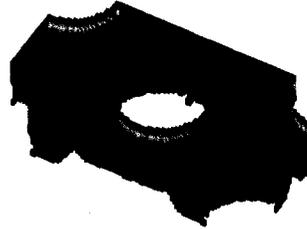
Figure 7.37 shows the predicted erosion that would take place over the first 300 kg of propellant use during the NEXT Long Duration Test. Note that the erosion is highly specific to the beamlet current, impingement to beamlet current ratio, and grid spacing used in the simulation.

$J_b = 0.208 \text{ mA}$
 $J_a/J_b = 0.38 \%$
 $l_g = 0.3048 \text{ mm}$

103 kg
5022 hours



202 kg
9804 hours



300 kg
14587 hours



Figure 7.37 Predicted erosion for the centerline aperture of the NEXT Long Duration Test.

Figure 7.38 shows the pit and groove erosion depths observed as a function of operation time over the first 300 kg of the NEXT Long Duration Test. The pit and groove erosion depths were extrapolated to wear through. End of life is indicated by groove wear through, which occurs at about 483 kg of propellant use, assuming thruster operation remains constant. Using the centerline saddle point potential as a function of time (not shown) over the first 300 kg, the onset of electron backstreaming (for the centerline aperture) is expected to occur at a propellant throughput value of about 446 kg or greater. Again, these values are specific to the values of J_b , J_a/J_b , and l_g used in the simulation.

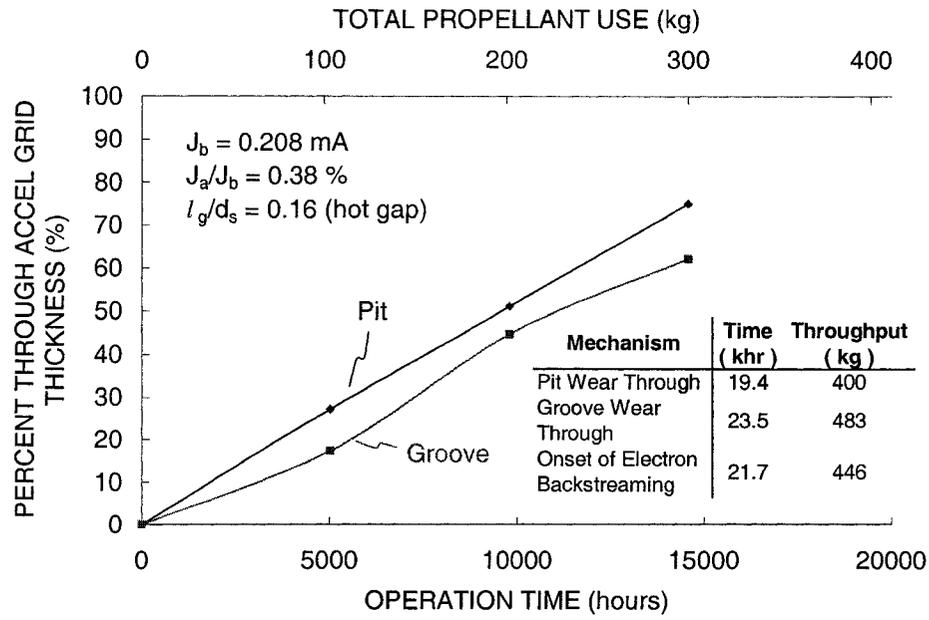


Figure 7.38 Pit and groove erosion depths for the centerline aperture of the NEXT LDT.

8 ffx Ion Optics Simulation Code

The ffx code is a three-dimensional simulation code designed primarily for the analysis of ion thruster optics. The code analyzes a three-dimensional, rectangular prism volume. A Cartesian mesh is used which greatly enhances simulation speed and the ease with which algorithms are implemented. The upstream and downstream surfaces are held at fixed potential values, while symmetry conditions are applied on the four remaining sides of the volume.

The mesh spacing in each coordinate direction is determined by the specified volume size and the number of nodes in each direction. The mesh spacing is uniform along each individual direction, but can be different among the three directions. Normally, the number of cells in the x direction is specified, which sets the mesh spacing in the x direction. The code then automatically chooses the number of cells in the y and z directions to approximately result in cubic cells.

An outline of the code algorithm is given in Figure 8.1. First, an input file is read by the code that sets the following:

- Grid geometry (l_{cc} , d_s ...)
- Applied voltages of the grids, upstream and downstream plasmas (V_s , V_a ...)
- Grid material and propellant type (Molybdenum, xenon)
- Beamlet current (J_b) or upstream ion density (n_i)
- Electron (T_e) and ion (T_i) temperatures
- Mesh size
- Double to single current ratio (J^{++}/J^+)
- Iteration scheme and number of ion macro particles
- Propellant utilization efficiency (η_u)

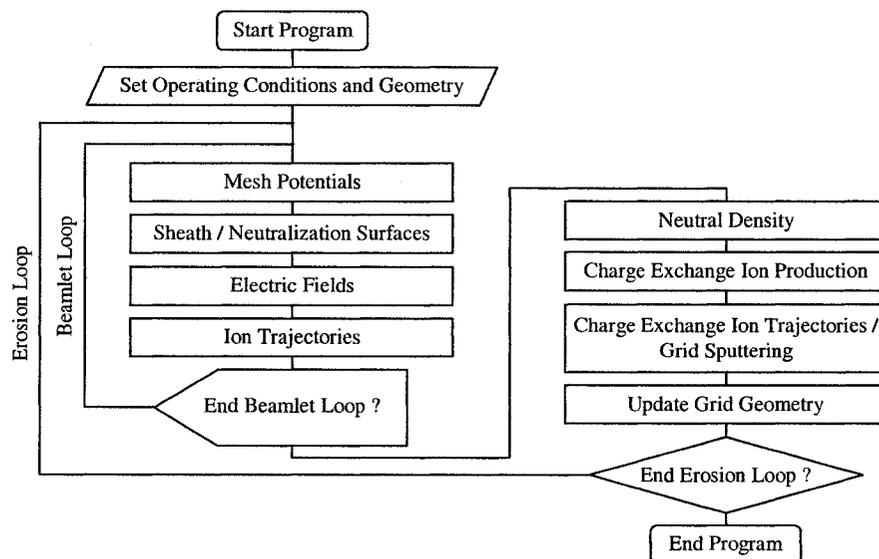


Figure 8.1 Flowchart for the ffx simulation code.

The first main iteration procedure is the beamlet loop, which can be thought of as the portion of the code that determines the beamlet shape. In the beamlet loop, the code alternates between 1) solving for mesh potential values and 2) tracking ion trajectories.

The region very close to the grids is considered to be electrostatic, where electric fields are very large compared to any magnetic fields that might be present [Brown]. In solving for the mesh potential values, the code solves the Poisson equation, which is a second order elliptical partial differential equation. This equation essentially relates the second derivative of potential, ϕ , to ion (n_i) and electron (n_e) number density. On the first iteration, no ions or electrons are present, thus the Poisson equation reduces to the Laplace equation when ρ is zero. The Poisson equation is given as Eq. 8.1.

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0} \quad \text{or} \quad \frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = -\frac{\rho}{\epsilon_0} \quad \text{Eq. 8.1}$$

The space charge, ρ , is comprised of ion and electron space charge, as given in Eq. 8.2. In the case of ion thrusters, only singly and doubly charged ions normally have significant populations in the plasmas. In this equation, ϵ_0 is the permittivity of free space.

$$-\frac{\rho}{\epsilon_0} = -\frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} = -\frac{q_i^+ n_i^+ + q_i^{++} n_i^{++} + q_e n_e}{\epsilon_0} \quad \text{Eq. 8.2}$$

$$\epsilon_0 \approx 8.8542 \cdot 10^{-12} \frac{F}{m} = 8.8542 \cdot 10^{-12} \frac{C^2}{Nm^2}$$

Once the Poisson equation has been solved to determine the scalar potential values, the electric field at each node is obtained by taking the negative of the gradient of the potential function, as shown in Eq. 8.3. In a Cartesian coordinate system, the electric field is essentially the first derivative of potential in each direction.

$$\vec{E} = -\nabla \phi \quad \text{or} \quad \vec{E} = -\left(\frac{\partial \phi}{\partial x} \hat{i} + \frac{\partial \phi}{\partial y} \hat{j} + \frac{\partial \phi}{\partial z} \hat{k} \right) \quad \text{Eq. 8.3}$$

Finite difference equations are used to approximate the derivatives in the Poisson and electric field equations [Chapra]. In using finite difference equations, several choices for the solution of the Poisson equation become available. Second order and fourth order accurate central divided difference equations are given in Eq. 8.4 for the x direction. Similar equations can be written for the y and z directions.

Second order accurate formulas:

$$\frac{\partial \phi}{\partial x} = \frac{\phi_{i+1} - \phi_{i-1}}{2\Delta x}$$

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{\phi_{i+1} - 2\phi_i + \phi_{i-1}}{(\Delta x)^2}$$

Eq. 8.4

Fourth order accurate formulas:

$$\frac{\partial \phi}{\partial x} = \frac{-\phi_{i+2} + 8\phi_{i+1} - 8\phi_{i-1} + \phi_{i-2}}{12\Delta x}$$

$$\frac{\partial^2 \phi}{\partial x^2} = \frac{-\phi_{i+2} + 16\phi_{i+1} - 30\phi_i + 16\phi_{i-1} - \phi_{i-2}}{12(\Delta x)^2}$$

Once the electric fields are known, ion macro particles are tracked through the simulation volume. An ion macro particle is a particle that behaves in electric and magnetic fields as a single ion, but carries the weight of many ions. As long as a sufficient number of macro particles are used, the simulation time is greatly reduced without sacrificing accuracy.

Ions are typically injected into the upstream surface at the Bohm velocity, which they acquire upstream of the sheath, in the pre-sheath region, as described previously. The Lorentz force law describes the forces on charged particles due to electric and magnetic fields. The method of particle tracking, called the Leapfrog method, used in the code is described in greater detail in a separate section.

As ion macro particles move through the simulation volume, they “deposit” space charge on the surrounding mesh nodes by the method of volume weighting, as shown in Figure 8.2. Eq. 8.5 shows the space charge (C/m^3) that is added to node 1 from the macro particle located at (x,y,z) . The macro particle carries a current J_{Particle} ($A = C/s$) and is considered to be at that location for a time step of Δt (s). The cell volume is V_{cell} (equal to $\Delta x \cdot \Delta y \cdot \Delta z$), and an additional volume fraction is used to divide up the space charge. Seven additional equations are used with different volume fractions for the remaining nodes. Note that as the particle moves closer to node 8, less space charge is deposited to node 1.

$$\rho_1 = \rho_1 + \frac{J_{\text{Particle}} \Delta t}{V_{\text{cell}}} \frac{(x_2 - x)(y_2 - y)(z_2 - z)}{\Delta x \cdot \Delta y \cdot \Delta z}$$

Eq. 8.5

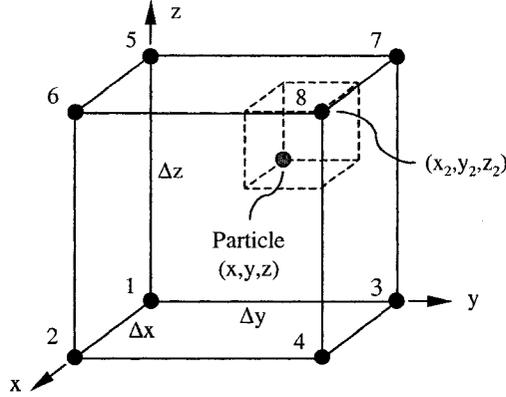


Figure 8.2 Method of volume weighting used to deposit ion space charge on mesh nodes.

Only one macro particle is tracked at a time. Ion macro particles are injected until either 1) the specified beamlet current is met or 2) the upstream plasma density is met.

Once all the ions have flown through the simulation volume, the Poisson equation can be solved again using the new estimates for the space charge at every node. Following that solution, the ion macro particles are tracked again to obtain better space charge values. The iteration procedure within the beamlet loop continues in this way until the potential values at all nodes change by a small amount, which signifies convergence.

To this point, the method of obtaining the electron space charge contribution has been omitted. The electrons within the discharge and beam plasmas upstream and downstream of the ion thruster grids are considered to be much more mobile than the ions. As a result, the electrons can be treated with equations that relate the electron density at any node to the average ion density within the plasma using an expression that involves the node potential relative to the plasma potential.

Two equations are used for each plasma, one for the case where the node potential is greater than the plasma potential and one where it is less than the plasma potential. These equations are given in Eq. 8.6 for the upstream discharge chamber plasma. Here, $(\rho_i^+ + \rho_i^{++})_{u \text{ Average}}$ is the average ion space charge in the upstream plasma, ϕ_u is the plasma potential, and T_{eu} is the electron temperature. Two similar equations are used for the downstream beam plasma with different plasma density, potential, and electron temperature values.

$$\rho_e = -(\rho_i^+ + \rho_i^{++})_{u \text{ Average}} \exp\left(\frac{\phi - \phi_u}{T_{eu}}\right) \quad \text{for } \phi \leq \phi_u$$

$$\rho_e = -(\rho_i^+ + \rho_i^{++})_{u \text{ Average}} \cdot \left(1 + \frac{\phi - \phi_u}{T_{eu}}\right) \quad \text{for } \phi \geq \phi_u$$

Eq. 8.6

Once the beamlet loop is finished, the code goes into the erosion loop where it deals with processes related to grid erosion.

First, the code calculates the neutral density variation throughout the grid region. This is done using analytical equations, the process of which is described in a separate section.

The neutral density variation is needed to calculate the volumetric charge exchange ion production rates, which is the next step in the erosion loop. The charge exchange ion production rate ($1/[s \cdot m^3]$) within a cell is a product of the neutral density ($1/m^3$), ion density ($1/m^3$), relative particle speed (m/s), and reaction cross section (m^2).

Once the rate of charge exchange ion production has been calculated for every cell, charge exchange ion macro particles are tracked through the volume. These particles originate from the centers of the cells where production takes place.

An erosion model is used to calculate the erosion rates caused by charge exchange ions that strike grid surfaces. Ions from the discharge chamber, tracked during the beamlet loop, can also add to the grid erosion rates.

The number of atoms sputtered from a grid cell is a product of the sputter yield and the number of ions carried by the impacting macro particle. The sputter yield is dependent on the impacting macro particle 1) energy and 2) angle of incidence. The angle of incidence is found using the local surface normal vector of the impacted cell. Erosion constantly changes the grid surface. As a result, normal vectors are calculated at the start of every erosion loop.

One way to obtain a surface normal vector is by calculating the location of the regional center of mass of the impacted cell relative to its geometrical center. The surface normal of the impacted cell is then defined as the line that extends from the center of mass in the region through the center of the impacted cell, as shown in Figure 8.3. These surface normal vectors give a much more reasonable representation of the true surface shape than do the cell faces.

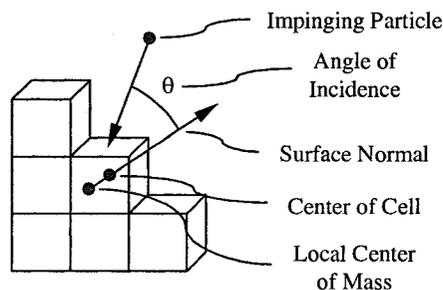


Figure 8.3 Calculation of the local surface normal vector of any cell.

After all charge exchange ions have been tracked, and erosion rates calculated, the deposition of sputtered grid material back onto the grids is taken into account. Vectors, sampled from a cosine distribution, are selected for every grid cell from which erosion takes place. These vectors are followed from the cells, taking into account the individual surface normal vectors, to termination upon a grid or the upstream or downstream surfaces. Sputtered grid material is taken to move through the simulation volume in straight lines. All sputtered grid atoms that end up terminating upon a grid cell are assumed to become part of that cell.

After the deposition of sputtered grid material has been considered, the overall cell mass change rates are calculated. These are simply the mass loss and mass gain rates added together.

An erosion time step is selected prior to the start of erosion simulations. This time step is used to calculate how many atoms are lost and gained from each cell using the overall mass change rates. Generally, this time step is on the order of 50 to 500 hours. The time step can be changed during the simulation either manually or automatically depending on the cell loss rates.

The erosion time step is multiplied by the overall mass change rate in all cells to determine the actual number of atoms lost or gained. There are several situations to consider. For illustration purposes, take the maximum number of atoms in a cell to be N_{Filled} , calculated given a certain cell size and grid material density. Furthermore, take the current number of atoms in a cell to be N . Situations, and the actions taken, are reported in Table 8.1 following a time step application. The code actions are designed to maintain the best possible atom conservation.

Table 8.1 Situations where cells in the simulation are either losing or gaining atoms.

Situation		Resolution
Cell is losing atoms.	$0 < N < N_{\text{Filled}}$	The cell has lost some cells, but it is not down to zero yet. The cell remains a grid cell with no modification.
Cell is losing atoms.	$N < 0$	The cell has lost more atoms than it had to give up. In this case, cells are taken evenly from surrounding grid cells until $N = 0$. Then, the cell is turned into a free cell, no longer part of a grid.
Cell is losing atoms.	$N > N_{\text{Filled}}$	Although the cell is losing mass, it still has more atoms than a completely filled cell of its size. This cell remains a grid cell with no additional modification.
Cell is gaining atoms.	$0 < N < N_{\text{Filled}}$	There are two possibilities. This cell is either a grid cell that has previously lost some mass or it is a free cell that is being filled with atoms. In either case, no action is taken.
Cell is gaining atoms.	$N > N_{\text{Filled}}$	<p>1) If the cell is already part of a grid, the excess cells are distributed to adjacent cells if $N \geq 2 \cdot N_{\text{Filled}}$. First, excess atoms are given to cells part of a grid if they have atom deficiencies. Next, excess cells start filling free cells evenly. Atom distribution is not done until $N \geq 2 \cdot N_{\text{Filled}}$ to make sure the cell is consistently gaining mass in order to avoid inadvertent atom loss.</p> <p>2) If the cell is a free cell, not part of a grid, it has now been completely filled with atoms so it becomes a grid cell. The last adjacent cell that helped fill the cell with atoms determines which grid the cell becomes a part of.</p>

After all special cell cases are dealt with, the code begins the next erosion loop. Because the grid shape can change, the beamlet loop is used to update the beamlet shape. Usually, grid changes are slight and the beamlet loop becomes much shorter after the first iteration.

The overall ffx code algorithm is typical of steady-state algorithms used in optics simulation codes. The steady state simplification allows the potential solution and particle tracking sections to be separated. The main advantage of this type of simulation is greater speed. A single run of the ffx code

typically takes less than an hour even for a relatively high beamlet current. A full simulation of grid erosion takes on the order of one day.

In time dependent simulations, particles are all tracked together, alternately moving particles and solving for potentials. A simulation of this type typically takes longer to run, but also has other inherent advantages. For instance, particle collisions can be taken into account if desired, and charge exchange ion space charge is included automatically. Charge exchange ion space charge can be taken into account in the ffx code using multiple erosion loops with an erosion time step of zero.

8.1 Particle Tracking

In the ffx code, particle pushing is performed using what is commonly called the Leapfrog method [Hut]. It is called the leapfrog method because particle positions and accelerations are essentially defined at integer values (t_i, t_{i+1}, t_{i+2} , etc.) while velocities are defined at half integer values ($t_{i-1/2}, t_{i+1/2}, t_{i+3/2}$, etc.). The two main equations for the Leapfrog method are given in Eq. 8.7.

$$\begin{aligned}\vec{r}_{i+1} &= \vec{r}_i + \vec{v}_i \Delta t + \frac{1}{2} \vec{a}_i (\Delta t)^2 \\ \vec{v}_{i+1} &= \vec{v}_i + \frac{1}{2} (\vec{a}_i + \vec{a}_{i+1}) \Delta t\end{aligned}\tag{Eq. 8.7}$$

The Leapfrog method is a second order accurate method. This is in contrast to the first order accurate Euler method or the fourth order accurate Runge-Kutta method for example. The Leapfrog method gives much better accuracy than the Euler method with only slightly more programming complexity. Also, the Leapfrog method requires less computations than the fourth order Runge-Kutta method, which helps preserve simulation speed.

The above equations for the Leapfrog method give the position and velocity on integer time steps. A different, but equivalent, way of writing this method is given in Eq. 8.8. Here, velocities are defined on half integer time steps. A special case is the velocity $v_{1/2}$, which is defined in Eq. 8.9 in terms of v_0 instead of $v_{-1/2}$.

$$\begin{aligned}\vec{r}_i &= \vec{r}_{i-1} + \vec{v}_{i-\frac{1}{2}} \Delta t \\ \vec{v}_{i+\frac{1}{2}} &= \vec{v}_{i-\frac{1}{2}} + \vec{a}_i \Delta t\end{aligned}\tag{Eq. 8.8}$$

$$\vec{v}_{\frac{1}{2}} = \vec{v}_0 + \vec{a}_0 \frac{\Delta t}{2}\tag{Eq. 8.9}$$

When injecting ion macro particles into the simulation volume, the initial position of the particle (\vec{r}_0) is determined using a random number generator. Additionally, the initial particle velocity (\vec{v}_0) is

axial, with the particle speed being the Bohm velocity. During particle movement, only the acceleration (\vec{a}_i) and time step (Δt) vary.

The force on a particle in the presence of magnetic and electric fields is given by the Lorentz force equation, Eq. 8.10. Here, q is the particle charge, \vec{v} is the velocity, \vec{B} is the magnetic field, and \vec{E} is the electric field. This force is equal to the particle's mass times its acceleration, \vec{a} .

$$\vec{F} = q(\vec{v} \times \vec{B} + \vec{E}) = m\vec{a} \quad \text{Eq. 8.10}$$

With the electrostatic simplification, a particle's acceleration at any time step is given in Eq. 8.11.

$$\vec{a}_i = \frac{q}{m} \vec{E}_i \quad \text{Eq. 8.11}$$

The location of a particle in the simulation volume is easily determined using the regularly spaced Cartesian mesh. Eq. 8.12 gives the x direction index number, x_{cell} , of the cell that contains the particle, where Δx is the mesh spacing in the x direction. Two additional equations are used for the y and z indices.

$$x_{cell} = \text{int}\left(\frac{x}{\Delta x}\right) \quad \text{Eq. 8.12}$$

The electric fields at every node are calculated prior to particle movement. To obtain the electric field at the particle's actual location, the electric fields at the 8 nodes of the cell are volume weighted in the reverse manner that space charge from the particle is applied to the nodes, as given in Eq. 8.13.

$$\vec{E} = \sum_{n=1}^8 \vec{E}_n \cdot V_n \quad \text{Eq. 8.13}$$

Here, V_n is the volume fraction appropriate to node n . One way to obtain the volume fractions is to use the cell index values. First, six length fractions, x_a through z_b , can be defined that describe the fractional position of the particle within the cell. Two of these fractions for the x direction are given in Eq. 8.14. Next, these length fractions are combined six different ways to obtain the volume fractions. The electric field in the x direction is given in Eq. 8.15 for example. Note that the closer a particle is to a particular node, the more that node influences the electric field at the particle's location.

$$x_a = \frac{x}{\Delta x} - x_{cell} \quad \text{Eq. 8.14}$$

$$x_b = 1 - \left(\frac{x}{\Delta x} - x_{cell}\right) = 1 - x_a$$

$$\begin{aligned}
E_x &= E_{x_{ijk}} \cdot x_b \cdot y_b \cdot z_b \\
&+ E_{x_{i+1jk}} \cdot x_a \cdot y_b \cdot z_b \\
&+ \vdots \\
&+ E_{x_{ij+1k+1}} \cdot x_b \cdot y_a \cdot z_a \\
&+ E_{x_{i+1j+1k+1}} \cdot x_a \cdot y_a \cdot z_a
\end{aligned} \tag{Eq. 8.15}$$

Conditions developed by Courant, Friedrichs, and Lewy (commonly called CFL conditions) describe the maximum time step that is required in order to accurately track particles through a certain mesh size. In the ffx code, the particle time step used in the Leapfrog method is adjusted constantly such that particles do not move more than one-third of the way through any cell in a single time step, as given by Eq. 8.16. Here, h is taken as the minimum mesh size among the three Cartesian directions, and $|\vec{v}|$ is the particle speed at any time. Reducing the time step requirement further was not seen to change simulation results.

$$\Delta t \leq \frac{1}{3} \frac{h}{|\vec{v}|} \tag{Eq. 8.16}$$

The total energy of a particle should ideally be conserved as it moves through the simulation volume. However, small changes in energy can arise when using numerical methods. To combat this effect, energy compensation is used to adjust the particle's velocity vector in order to keep a constant total energy. The initial total energy is set by the initial particle velocity, usually the Bohm velocity, and the potential at the location where the particle is injected, usually the upstream surface. At any location the potential is ϕ , and there is a corresponding conserved velocity $v_{conserved}$ at that location, as given by Eq. 8.17. The uncorrected particle velocity following a time step is v , given in Eq. 8.18. Defining α to be the ratio of the conserved to actual velocity magnitudes, Eq. 8.19, the particle velocity is corrected by multiplying the three velocity components by α to achieve the desired (conserved) velocity magnitude, as shown in Eq. 8.20.

$$\begin{aligned}
E_{Total} &= E_{Potential} + E_{Kinetic} \\
&= q\phi + \frac{1}{2}mv^2
\end{aligned}$$

$$v_{conserved} = \sqrt{\frac{2}{m}(E_{Total} - q\phi)} \tag{Eq. 8.17}$$

$$v = \sqrt{v_x^2 + v_y^2 + v_z^2} \tag{Eq. 8.18}$$

$$\alpha = \frac{v_{conserved}}{v} \tag{Eq. 8.19}$$

$$v_{conserved} = \sqrt{(\alpha v_x)^2 + (\alpha v_y)^2 + (\alpha v_z)^2} \tag{Eq. 8.20}$$

Note that the potential, ϕ , at any location is obtained in the same way the electric fields are obtained at any location. Specifically, volume weighting is used with the node potentials instead of the electric fields.

8.2 Solution of the Poisson Equation

The Poisson equation, Eq. 8.21, generally describes the relationship between potential, ϕ , and charged particle space charge, ρ , in the electrostatic case. Using a Cartesian coordinate system and considering the plasma to consist of singly and doubly charged ions as well as electrons, this equation expands to Eq. 8.22.

$$\nabla^2 \phi = -\frac{\rho}{\epsilon_0} \quad \text{Eq. 8.21}$$

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial y^2} + \frac{\partial^2 \phi}{\partial z^2} = -\frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} \quad \text{Eq. 8.22}$$

There are two ways to treat the electron populations in the upstream, discharge chamber, and downstream, beam, plasmas. One way is referred to as the cold sheath approximation, which essentially assumes a zero volt electron temperature. Here, electrons are assumed to be present at any location where the potential is at (or above) the reference plasma potentials. During the potential solution procedure, no node potential is allowed to be greater than the appropriate reference plasma potential. This assumes that the electron and ion space charges are exactly equal at or above the plasma potentials.

The other way to describe the electron populations is by using Boltzmann electron relationships. Here, the electron populations are described in terms of the bulk, average, ion densities within the discharge chamber and beam plasmas. The electron density at any node depends on the node potential and electron temperature of the plasma. Eq. 8.23 gives these relationships for the upstream discharge chamber plasma. The electron temperature determines how fast the electron density drops off as the potential decreases in moving toward the screen and accel grids.

$$\begin{aligned} \rho_e &= -(\rho_i^+ + \rho_i^{++})_{u \text{ Average}} \exp\left(\frac{\phi - \phi_u}{T_{eu}}\right) && \text{for } \phi \leq \phi_u \\ &= -\rho_{iu \text{ Avg}} \exp\left(\frac{\phi - \phi_u}{T_{eu}}\right) \\ \rho_e &= -(\rho_i^+ + \rho_i^{++})_{u \text{ Average}} \cdot \left(1 + \frac{\phi - \phi_u}{T_{eu}}\right) && \text{for } \phi \geq \phi_u \\ &= -\rho_{iu \text{ Avg}} \cdot \left(1 + \frac{\phi - \phi_u}{T_{eu}}\right) \end{aligned} \quad \text{Eq. 8.23}$$

In using the electron equations, a boundary is set along the axial distance of the simulation, usually just above the accel grid, that separates the discharge chamber and beam plasma regions. One set of potential and electron temperature values are used upstream of the boundary, and a different set downstream of the boundary.

The average ion densities in the upstream and downstream regions are found during the code simulation by averaging the ion densities over nodes close to the upstream and downstream boundary surfaces.

One important result of using the Boltzmann electron equations occurs just downstream of the accel grid. Here, the beamlet often converges on the centerline, making the ion density, and consequently the potential, greater there than the average ion density found farther downstream. This subtle increase in potential tends to direct charge exchange ions into pit and groove erosion patterns that better match experimentally seen erosion patterns.

One significant disadvantage of using the Boltzmann electron relationship is that it makes the Poisson equation nonlinear as a result of the exponential term. This has several implications with respect to solution techniques.

Here, second order finite centered difference formulas can be used to approximate the second order partial derivatives in the Poisson equation, which gives Eq. 8.24.

$$\frac{\phi_{i+1} - 2\phi_{ijk} + \phi_{i-1}}{(\Delta x)^2} + \frac{\phi_{j+1} - 2\phi_{ijk} + \phi_{j-1}}{(\Delta y)^2} + \frac{\phi_{k+1} - 2\phi_{ijk} + \phi_{k-1}}{(\Delta z)^2} = -\frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} \quad \text{Eq. 8.24}$$

After the finite difference equations are substituted into the Poisson equation, either explicit or implicit solution techniques can be used to solve the equations. The goal in either case is essentially to invert the matrix formed by combining all of the equations written for every node in the mesh. This matrix is of the form given in Eq. 8.25, where A is a coefficient matrix, u is a vector of unknown variables, and f is a vector of constants.

$$Au = f \quad \text{Eq. 8.25}$$

Explicit solution of the matrix involves inverting the matrix directly. Inverting matrices is usually efficient for small matrices, but very time consuming and memory intensive for large matrices. Additionally, explicit methods are usually only used for linear sets of equations, requiring some sort of linear approximation to be performed on the Boltzmann electron equations.

Implicit methods are more often used with large sets of equations, as in this case. Additionally, both linear and nonlinear solvers can be developed. The Gauss-Seidel and Jacobi methods are discussed first, which can be used with linear sets of equations. Newton's method is discussed next, which works for both linear and nonlinear sets of equations.

8.2.1 Gauss-Seidel and Jacobi Methods

The Gauss-Seidel and Jacobi iteration methods are similar. The Poisson equation written for a general node can be solved for the unknown node potential, ϕ_{ijk} , resulting in Eq. 8.26 for example. To solve the entire set of equations for all nodes, one simply iterates through every node solving for new values of ϕ_{ijk} using the currently stored values for the surrounding node potentials (ϕ_{i+1} , ϕ_{i-1} , etc.).

$$\phi_{ijk} = \left(\frac{1}{-2 \left(\frac{1}{(\Delta x)^2} + \frac{1}{(\Delta y)^2} + \frac{1}{(\Delta z)^2} \right)} \right) \times \left(-\frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} - \frac{\phi_{i+1} + \phi_{i-1}}{(\Delta x)^2} - \frac{\phi_{j+1} + \phi_{j-1}}{(\Delta y)^2} - \frac{\phi_{k+1} + \phi_{k-1}}{(\Delta z)^2} \right) \quad \text{Eq. 8.26}$$

Initially, guess values for all of the node potentials are chosen. Usually, using $\phi = 0$ for all nodes is acceptable, although faster convergence can be obtained if better initial guesses are used. Nodes that are at known potentials, for instance those nodes located inside grids and on the upstream and downstream boundaries, are simply not calculated. For a linear set of equations, this method converges to one set of potential values after iterating through all nodes multiple times.

In this case, a linear set of equations arises when the electron charge density ρ_e is linear. Recall that ρ_i^+ and ρ_i^{++} are constants during the potential solution, obtained during particle tracking, where each node can have a different charge density value. The Boltzmann expression can be made linear by approximating the exponential function by a linear function. Then, the part of the electron equation that depends on ϕ_{ijk} can be factored out, once again obtaining an equation similar to the one above where ϕ_{ijk} is solved for explicitly.

The difference between the Gauss-Seidel and Jacobi methods lies in which node potential values are used to find new potential values. Gauss-Seidel uses the most recently calculated node values always. For example, in calculating the new potential ϕ_{ijk} , a new potential ϕ_{i-1} was just calculated while ϕ_{i+1} will be calculated next. Gauss-Seidel uses the new value for ϕ_{i-1} to calculate ϕ_{ijk} , while Jacobi would use the old value for ϕ_{i-1} to calculate ϕ_{ijk} . The Jacobi method does not use the new set of potential values until all of the nodes have been calculated.

Using the most recent potential values usually results in Gauss-Seidel converging slightly faster than Jacobi. However, the node potentials become slightly directional when using Gauss-Seidel because the new potential values depend on the order in which the nodes are updated. For example, the potential values will be different when iterating from $i = 0$ to $i = n$ than when iterating from $i = n$ to $i = 0$ unless exact convergence is achieved. The Jacobi method is not directional because it updates the entire set of potentials only after all new node potentials have been calculated.

One way to make Gauss-Seidel non-directional is to use what is commonly called a red-black updating scheme. In this method, all of the “red” nodes are updated first, followed by the “black” nodes. In using the second order finite difference equations, only the six adjacent node potentials (ϕ_{i+1} , ϕ_{i-1} , etc.) are used in updating the current node potential (ϕ_{ijk}). Therefore, the node potentials ϕ_{ijk} , ϕ_{i+2jk} , ϕ_{i+4jk} , etc. can be updated successively without affecting each other for example. Once the red nodes are updated, the new values for the red nodes can be used for the black nodes. The red-black checkerboard analogy is really for two dimensions, but the same concept can be used in three dimensions without much more complexity.

8.2.2 Newton’s Method

Newton iteration can be used in the case where the system of equations is either linear or nonlinear. Newton’s method treats the system of equations as a root finding problem. To turn the elliptic partial differential equation into a root finding problem, all of the terms are moved to one side. The new function will be called $L(\phi)$, and the goal is to determine all of the values of ϕ such that $L(\phi) = 0$ as in Eq. 8.27. In terms of a single general node, ϕ_{ijk} , Eq. 8.28 is obtained using the second order finite difference equations.

$$L(\phi) = \nabla^2 \phi + \frac{\rho}{\epsilon_0} \quad \text{Eq. 8.27}$$

$$L(\phi_{ijk}) = \frac{\phi_{i+1} - 2\phi_{ijk} + \phi_{i-1}}{(\Delta x)^2} + \frac{\phi_{j+1} - 2\phi_{ijk} + \phi_{j-1}}{(\Delta y)^2} + \frac{\phi_{k+1} - 2\phi_{ijk} + \phi_{k-1}}{(\Delta z)^2} + \frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} \quad \text{Eq. 8.28}$$

Figure 8.4 illustrates Newton’s method. To obtain a new estimate for ϕ_{ijk} that makes $L(\phi_{ijk}) = 0$, the slope of the function is calculated at ϕ_{ijk}^{old} . Using this slope, Eq. 8.29 can be derived by noting that $L(\phi_{ijk}^{new}) = 0$.

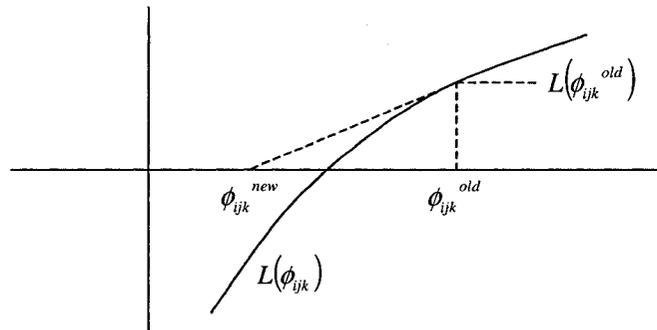


Figure 8.4 Illustration of Newton’s method.

$$\frac{\partial L(\phi_{ijk}^{old})}{\partial \phi_{ijk}} = \frac{L(\phi_{ijk}^{old}) - L(\phi_{ijk}^{new})}{\phi_{ijk}^{old} - \phi_{ijk}^{new}} = \frac{L(\phi_{ijk}^{old}) - 0}{\phi_{ijk}^{old} - \phi_{ijk}^{new}}$$

$$\phi_{ijk}^{new} = \phi_{ijk}^{old} - \frac{L(\phi_{ijk}^{old})}{\frac{\partial L(\phi_{ijk}^{old})}{\partial \phi_{ijk}}}$$
Eq. 8.29

In this case, the partial derivative of L can be found easily, although two equations are obtained for each plasma because of the electron equations. Eq. 8.30 gives the derivative generally, without substitution of an electron equation. Eq. 8.31 gives the derivatives of the electron equations for the upstream plasma depending on the relationship between ϕ_{ijk}^{old} and ϕ_u .

$$\frac{\partial L(\phi_{ijk})}{\partial \phi_{ijk}} = \frac{-2}{(\Delta x)^2} + \frac{-2}{(\Delta y)^2} + \frac{-2}{(\Delta z)^2} + \frac{1}{\epsilon_0} \frac{\partial \rho_e}{\partial \phi_{ijk}}$$
Eq. 8.30

$$\frac{\partial \rho_e}{\partial \phi_{ijk}} = -\rho_{iu \text{ Avg}} \frac{1}{T_{eu}} \exp\left(\frac{\phi_{ijk}^{old} - \phi_u}{T_{eu}}\right) \quad \text{for } \phi_{ijk}^{old} \leq \phi_u$$
Eq. 8.31

$$\frac{\partial \rho_e}{\partial \phi_{ijk}} = -\rho_{iu \text{ Avg}} \frac{1}{T_{eu}} \quad \text{for } \phi_{ijk}^{old} \geq \phi_u$$

For any node, one of four equations is used in Newton's method depending on 1) the location of the node within the upstream or downstream plasmas and 2) the value of ϕ_{ijk}^{old} relative to either ϕ_u or ϕ_d . These two criteria essentially determine which equation for ρ_e to use. Iteration through the nodes is done using the red-black iteration scheme as would be used with the Gauss-Seidel method.

8.2.3 Successive Overrelaxation

The rate of convergence of the Gauss-Seidel method can often be improved by using successive overrelaxation. Relaxation tries to predict a new value of ϕ_{ijk}^{new*} based on past calculated values ϕ_{ijk}^{new} and ϕ_{ijk}^{old} , as in Eq. 8.32. The parameter ω is called the relaxation parameter. For $0 < \omega < 1$ the method is called successive underrelaxation, and for $1 < \omega < 2$ the method is called successive overrelaxation.

$$\phi_{ijk}^{new*} = \omega \phi_{ijk}^{new} + (1 - \omega) \phi_{ijk}^{old} \quad \text{where } 0 < \omega < 2$$
Eq. 8.32

As an example, consider $\phi_{ijk}^{old} = 1000$ and $\phi_{ijk}^{new} = 1010$, where ϕ_{ijk}^{new} was just calculated using a Gauss-Seidel iteration. Looking at these values, one might expect ϕ_{ijk} to continue to increase on

the next iteration. Using $\omega = 1.5$, the value of ϕ_{ijk}^{new*} would be immediately set to 1015, likely improving the estimate of ϕ_{ijk} with a single calculation. The optimum value of ω is problem specific.

Using the nonlinear electron equations, the Gauss-Seidel method can be used to solve the Poisson equation employing underrelaxation to damp out the exponential potential oscillations.

8.2.4 Linearization of the Electron Equations

One way to keep the Poisson equation linear is to approximate the nonlinear electron equations using a Taylor series expansion. A general Taylor series expansion of the function $f(x)$ around x_0 is given in Eq. 8.33. Applying the Taylor series to the electron equation around ϕ_u that pertains to $\phi \leq \phi_u$ and keeping only the first two terms gives Eq. 8.34.

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2!}f''(x_0) + \dots + \frac{(x - x_0)^n}{n!}f^n(x_0) + R \quad \text{Eq. 8.33}$$

$$\begin{aligned} \rho_e &= -\rho_{iu \text{ Avg}} \exp\left(\frac{\phi - \phi_u}{T_{eu}}\right) \\ &\approx -\rho_{iu \text{ Avg}} \exp\left(\frac{\phi_u - \phi_u}{T_{eu}}\right) - (\phi - \phi_u)\rho_{iu \text{ Avg}} \exp\left(\frac{\phi_u - \phi_u}{T_{eu}}\right) \frac{1}{T_{eu}} \\ &= -\rho_{iu \text{ Avg}} - \rho_{iu \text{ Avg}} \frac{(\phi - \phi_u)}{T_{eu}} \\ &= -\rho_{iu \text{ Avg}} \left(1 + \frac{\phi - \phi_u}{T_{eu}}\right) \end{aligned} \quad \text{Eq. 8.34}$$

The end result is that the same electron equation applies for the case where $\phi \leq \phi_u$ as it does where $\phi \geq \phi_u$. However, for $\phi \leq \phi_u$, the linearization is only appropriate when $\left(1 + \frac{\phi - \phi_u}{T_{eu}}\right) \geq 0$, or alternatively $\phi \geq \phi_u - T_{eu}$. Although not shown, the Poisson equation can be explicitly solved for ϕ_{ijk} , facilitating the use of the Gauss-Seidel and Jacobi methods. Linearization of the electron equations is not used in the ffx code.

8.2.5 Multigrid

In three dimensions, the number of mesh nodes can become very large. As a result, a large number of iterations have to be performed in order to solve the mesh potentials to the desired accuracy using the Gauss-Seidel, Jacobi, or Newton iteration methods. A typical ffx simulation mesh might contain 32 by 56 by 292 cells, or 551133 nodes.

More advanced methods can be used to improve the solution speed. One such method is the multigrid method, which is appropriate for elliptic partial differential equations, either linear or nonlinear [Press].

The multigrid method uses a series of meshes of different spacing, as shown in Figure 8.5. The finest mesh, with spacing h , is the mesh on which the problem is originally defined. In other words, the solution of ϕ on the finest mesh is what is desired. Each coarser grid typically has twice the spacing of the previous grid: $2h$, $4h$, etc.

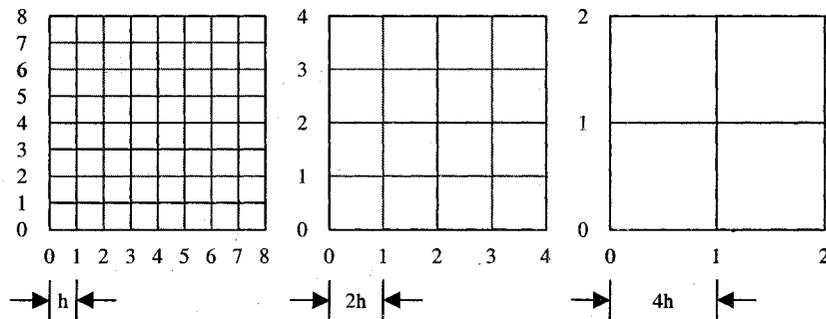


Figure 8.5 Fine grid with spacing h and the next two coarser grids. The coarser grids have spacing $2h$, $4h$, $8h$, etc.

The idea behind using grids of different spacing is that information about mesh potentials can be transferred between node points much more quickly with the coarse grids than with the fine grids. For the fine grid in Figure 8.5 for example, it would take 8 iterations through all of the nodes for information to travel from one side of the mesh to the other. However, using the grid of $4h$ spacing, it only takes 2 iterations for the information to pass across the mesh. In three dimensions, each successive coarse grid has 8 times fewer cells than the last.

There are two main ways to use multigrid methods. The linear multigrid method is, appropriately, used for linear problems. The Full Approximation Storage Algorithm (FAS) multigrid method is more general and can be used for both linear and nonlinear problems.

8.2.6 Linear Multigrid Method

A system of equations, one for every node, arises from the Poisson equation that can be thought of in Matrix form as in Eq. 8.35. Here, u is the vector of unknowns, A is the coefficient matrix, and f is the

collection of all source terms. On the finest grid, the unknowns are the node potentials, ϕ , and the source terms are the ion densities. With the linear method, the electron densities have to be linear, thus they can also be thought of as source terms here.

$$Au = f \quad \text{Eq. 8.35}$$

The multigrid method is general to more than just the Poisson equation. As a result, the equation to be solved in a more general sense is $Lu = f$. The Poisson equation is of the form $\nabla^2 u = f$.

Figure 8.6 shows the linear multigrid method using only two grids. The linear multigrid method only solves for the node potentials on the finest grid. The next coarsest grid, of spacing $2h$, is used to find corrections to the potentials on the finest grid. In this case, u_h are the potential values, ϕ . The defect, d , is the error in the approximate solution u , and it should ideally be going to zero as the solution progresses. The coarser grid with spacing $2h$ is used to find a correction, v , to the potentials, u , on the finest grid. The corrections, v , like the defect should be going toward zero as the solution progresses. Eq. 8.36 gives the definition of u_h , $L_h u_h$, and f_h for the finest grid in the linear method.

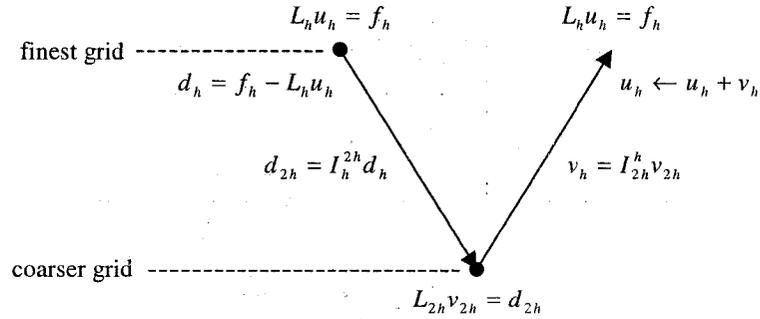


Figure 8.6 One v-cycle of the linear multigrid algorithm. The coarser grid is used to find a correction v to the potentials u on the fine grid.

$$u_h = \phi$$

$$L_h u_h = \frac{u_{i+1} - 2u_{ijk} + u_{i-1}}{(\Delta x)^2} + \frac{u_{j+1} - 2u_{ijk} + u_{j-1}}{(\Delta y)^2} + \frac{u_{k+1} - 2u_{ijk} + u_{k-1}}{(\Delta z)^2}$$

$$f_h = -\frac{\rho_i^+ + \rho_i^{++} + \rho_e}{\epsilon_0} \quad \text{Eq. 8.36}$$

$$\frac{u_{i+1} - 2u_{ijk} + u_{i-1}}{(\Delta x)^2} + \frac{u_{j+1} - 2u_{ijk} + u_{j-1}}{(\Delta y)^2} + \frac{u_{k+1} - 2u_{ijk} + u_{k-1}}{(\Delta z)^2} = f_h$$

The linear multigrid method can be summarized as follows:

1. Pre smooth the solution on the finer grid by doing a few relaxation iterations solving $L_h u_h = f_h$.
2. Calculate the defect (sometimes called the residual) $d_h = f_h - L_h u_h$.
3. Restrict the defect to the next coarser grid: $d_{2h} = I_h^{2h} d_h$.

4. Use relaxation iterations to solve $L_{2h}v_{2h} = d_{2h}$ for the correction v .
5. Prolongate the correction, v , on the coarse grid to the finer grid: $v_h = I_{2h}^h v_{2h}$.
6. Correct the finer grid solution: $u_h \leftarrow u_h + v_h$.
7. Post smooth the solution on the finer grid by doing a few relaxation iterations solving $L_h u_h = f_h$.
8. Repeat the cycle until the solution tolerance has been achieved.

In solving $L_{2h}v_{2h} = d_{2h}$ on the coarser grid, the source term in Poisson's equation becomes the defect, d , and the grid spacing is doubled in each direction, as shown in Eq. 8.37.

$$L_{2h}v_{2h} = d_{2h}$$

$$\frac{v_{i+1} + 2v_{ijk} + v_{i-1}}{(2\Delta x)^2} + \frac{v_{j+1} + 2v_{ijk} + v_{j-1}}{(2\Delta y)^2} + \frac{v_{k+1} + 2v_{ijk} + v_{k-1}}{(2\Delta z)^2} = d_{2h} \quad \text{Eq. 8.37}$$

Relaxation refers to the solution of the Poisson equation on any grid level using the Gauss-Seidel, Jacobi, or Newton methods. The nodes are updated only a few times on each level, just to smooth the unknown vector before moving to a different grid level. The largest iteration time is spent at the coarsest level where there are the least number of nodes.

The restriction and prolongation steps transfer information between grid levels. Restriction, I_h^{2h} , goes from fine to coarse and prolongation, I_{2h}^h , goes from coarse to fine. Figure 8.7 illustrates restriction and prolongation using full weighting. For node ijk , there are 27 surrounding nodes that are involved in changing levels. Restriction takes values from these 27 fine mesh nodes and combines them using a certain weighting scheme to obtain one value at node ijk in the coarse mesh. Similarly, when going from coarse to fine, the value at node ijk is distributed to the surrounding 27 nodes using a similar weighting scheme.

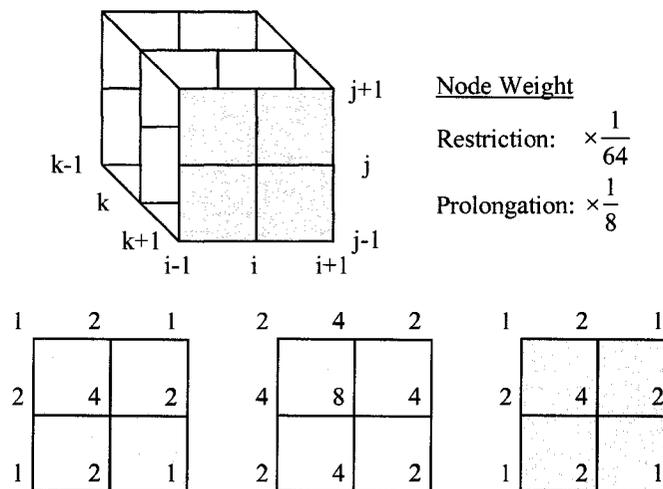


Figure 8.7 Restriction and prolongation.

Eq. 8.38 shows how restriction of the defect, d , is accomplished in going from a fine grid of spacing h to a coarse grid of spacing $2h$ for the general node ijk . In restriction, the weights add up to unity. Thus, the defect at node ijk in the coarse grid is an average defect near the node. In prolongation, the weights add up to 8. After cycling through all of the nodes in the coarse grid, all nodes in the fine grid receive a total weighted value of unity during prolongation.

$$\begin{aligned}
 d_{2h}^{ijk} &= \frac{1}{64} d_h^{i-1j-1k-1} + \frac{2}{64} d_h^{ij-1k-1} + \frac{1}{64} d_h^{i+1j-1k-1} \\
 &+ \dots + \frac{8}{64} d_h^{ijk} + \dots \\
 &+ \frac{1}{64} d_h^{i-1j+1k+1} + \frac{2}{64} d_h^{ij+1k+1} + \frac{1}{64} d_h^{i+1j+1k+1}
 \end{aligned}
 \tag{Eq. 8.38}$$

More grid levels are usually used to increase the rate of convergence. Two typical methods of using the grids are v-cycles and w-cycles. Figure 8.8 shows an example of a v-cycle and a w-cycle with three grid levels. Cycles of either method are performed until the solution on the finest grid reaches some tolerance value.

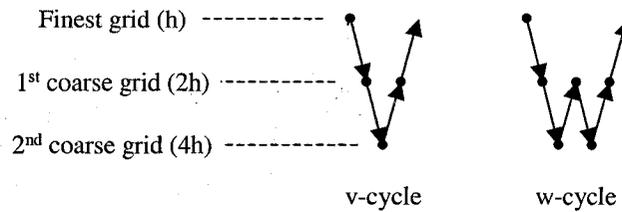


Figure 8.8 A v-cycle and a w-cycle using three grid levels.

Because it is not altogether clear how to go from the 1st coarse grid to the 2nd coarse grid, Figure 8.9 shows how the linear multigrid method works with three grids in a v-cycle. One should then be able to extend the ideas shown in this figure to a w-cycle. In the linear multigrid method, the second coarse grid is actually used to make corrections to the corrections on the first coarse grid. In trying to be consistent with the notation, the defect on the first coarse grid level has been called d'_{2h} to differentiate it from the restricted defect from the finest grid d_{2h} . The same has been done for the correction from the second coarse grid, v'_{2h} , used to correct v_{2h} .

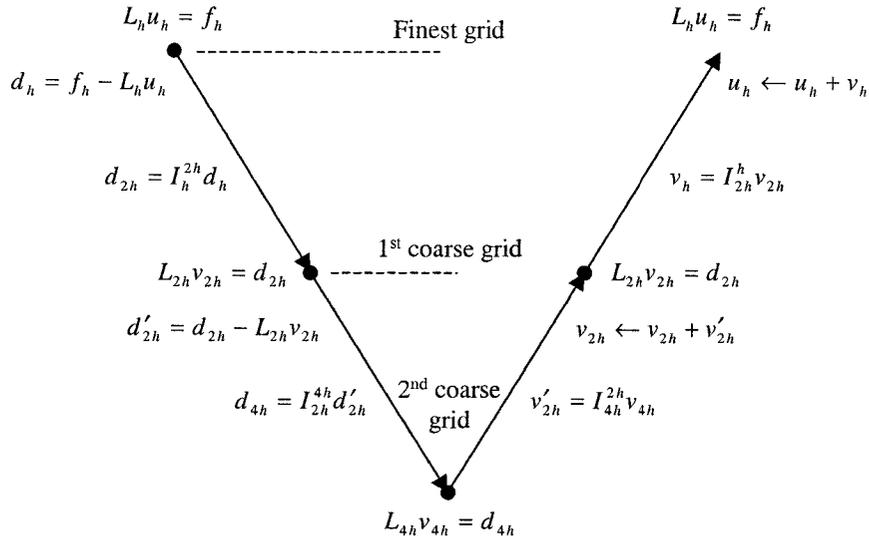


Figure 8.9 The linear multigrid method using a v-cycle with three grid levels.

Typically, each successive coarse grid has double the previous mesh spacing. As a result, in order to use two grids in the multigrid method, the number of cells in the finest mesh needs to be a multiple of 2 in each direction so that the coarse grid is well defined. Similarly, if three grids are to be used in the multigrid method, the finest mesh needs to be a multiple of 4 in each direction.

For example, if there are 32 x-direction cells in the finest mesh of a three grid setup, the first coarse grid will have 16 x-direction cells, and the second coarse grid will have 8 x-direction cells. However, 30 x-direction cells in the finest mesh will not work with a three grid setup because the second coarse grid would not have an integer number of cells.

8.2.7 Nonlinear Multigrid Method

The nonlinear version of the multigrid method is called the Full Approximation Storage Algorithm, or FAS, method. As opposed to the linear multigrid method, the nonlinear FAS multigrid method essentially solves the full problem, i.e. solves for the node potentials, on all of the grid levels.

The main drawback of the FAS method is that there are more operations that have to be done in going between levels than in the linear method. Thus, the FAS method is slower than the linear method when applied to linear equations, but both methods will solve the linear equation in the same way. In the FAS algorithm, the electron density term, which is a function of potential, is written on the left hand side of the equation, and only the constant ion density source term on the right, as in Eq. 8.39.

$$u_h = \phi$$

$$L_h[u_h] = \frac{u_{i+1} - 2u_{ijk} + u_{i-1}}{(\Delta x)^2} + \frac{u_{j+1} - 2u_{ijk} + u_{j-1}}{(\Delta y)^2} + \frac{u_{k+1} - 2u_{ijk} + u_{k-1}}{(\Delta z)^2} + \frac{\rho_e}{\varepsilon_0} \quad \text{Eq. 8.39}$$

$$f_h = -\frac{\rho_i^+ + \rho_i^{++}}{\varepsilon_0}$$

The scheme of the FAS multigrid method is shown in Figure 8.10. Here, one can see that there are several more operations in going between the finer and coarser grid levels. These extra operations can have a very noticeable affect on the speed that the algorithm solves a linear equation when compared to the linear method. One also may notice that the form of the equation being solved on both the finer and coarser grid is the same. Both u_h and u_{2h} are potential values that have nearly the same magnitude on both grid levels at the same points in space.

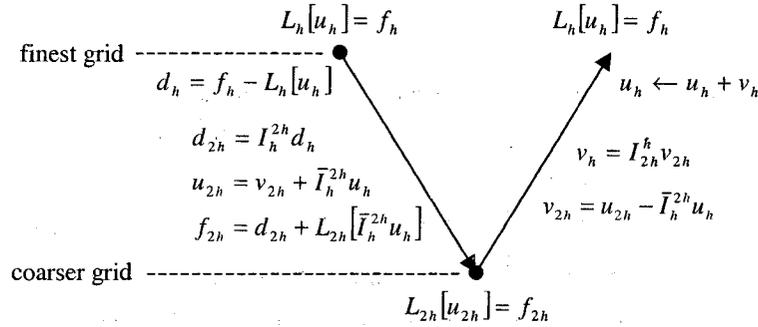


Figure 8.10 One v-cycle of the Full Approximation Scheme (FAS) multigrid method.

The nonlinear FAS multigrid method can be summarized as follows:

1. Pre smooth the solution on the finer grid by doing a few relaxation iterations solving

$$L_h[u_h] = f_h.$$

2. Calculate the defect (sometimes called the residual) $d_h = f_h - L_h[u_h]$.

3. Restrict the defect to the next coarser grid: $d_{2h} = I_h^{2h} d_h$.

4. Calculate u and f on the next coarser grid: $u_{2h} = v_{2h} + \bar{I}_h^{2h} u_h$, $f_{2h} = d_{2h} + L_{2h}[\bar{I}_h^{2h} u_h]$.

5. Use relaxation iterations to solve $L_{2h}[u_{2h}] = f_{2h}$ for u_{2h} .

6. Calculate the correction, v : $v_{2h} = u_{2h} - \bar{I}_h^{2h} u_h$.

7. Prolongate the correction, v , on the coarse grid to the finer grid: $v_h = I_{2h}^h v_{2h}$.

8. Correct the finer grid solution: $u_h \leftarrow u_h + v_h$.

9. Post smooth the solution on the finer grid by doing a few relaxation iterations solving

$$L_h[u_h] = f_h.$$

10. Repeat the cycle until the solution tolerance has been achieved.

With the FAS method, Newton's method must be used (instead of the Gauss-Seidel or Jacobi methods) for the relaxation iterations because it is a nonlinear method, capable of solving the nonlinear electron equation terms.

A potential solution for a typical NSTAR grid simulation is shown in Figure 8.11. The potential values of the nodes within the grids and on the upstream and downstream volume faces are known, as shown in this figure. Initially, all unknown node potential values are set to 0.0 V.

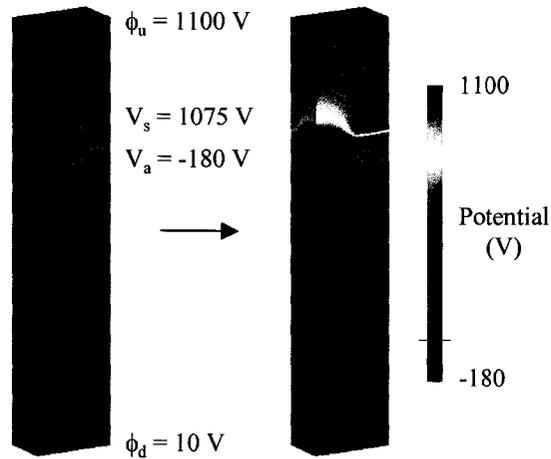


Figure 8.11 Potential solution of a typical NSTAR simulation.

The convergence histories of four solution techniques solving the typical NSTAR simulation are shown in Figure 8.12. Here, the multigrid method is not used (the solution is found using the finest grid only). On the first iteration, with no ion or electron space charge known ($\rho = 0$), the Poisson equation is the Laplace equation, $\nabla^2 \phi = 0$. The Jacobi method is compared to successive overrelaxation using three different relaxation parameters, ω . When $\omega = 1.0$, successive overrelaxation reduces to the Gauss-Seidel method. Also, with the Laplace equation, Newton's method is the same as the Gauss-Seidel method.

As expected, the Jacobi method shows slower convergence than the Gauss-Seidel method. Additionally, overrelaxation ($\omega = 1.5$) increases the convergence rate while underrelaxation ($\omega = 0.5$) decreases the convergence rate.

A single iteration involves one update of every unknown node potential in the mesh. For perspective, 5000 iterations took approximately 40 minutes to perform using a processor rated at about 2.4 GHz. It is noted that better performance could certainly be obtained by optimizing the algorithms and coding.

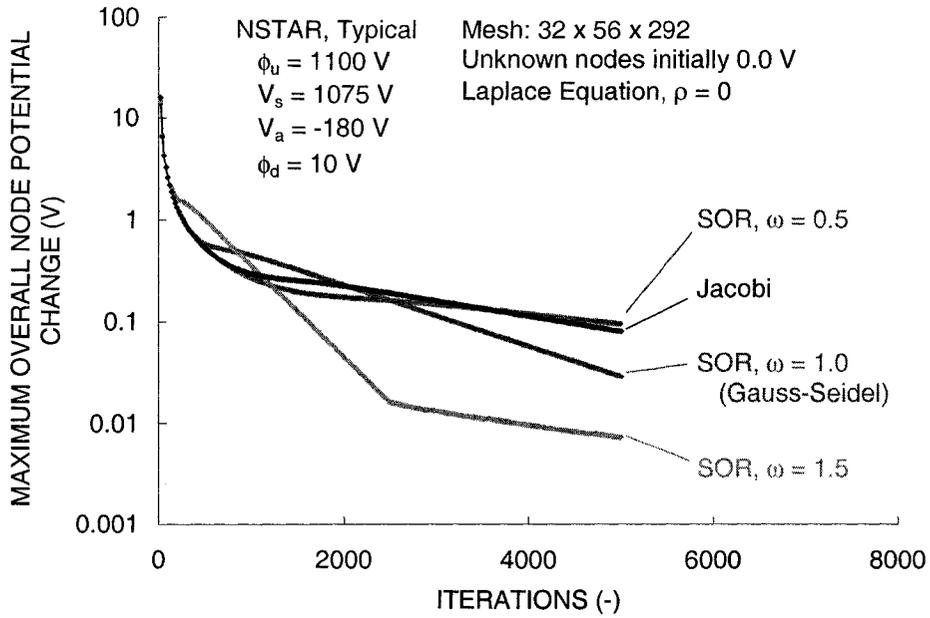


Figure 8.12 Comparison of the Jacobi and Successive Relaxation methods.

The ffx code normally uses the FAS multigrid method with three grid levels in a v-cycle pattern. Figure 8.13 compares the multigrid method using one, two, and three grid levels on the same NSTAR problem. Newton's method is used for the relaxation iterations in the FAS algorithm. It is easily seen that a significant increase in convergence is obtained when using multiple grids. With a simulation of this size in three dimensions, the extra programming complexity of the multigrid method is well worth the effort.

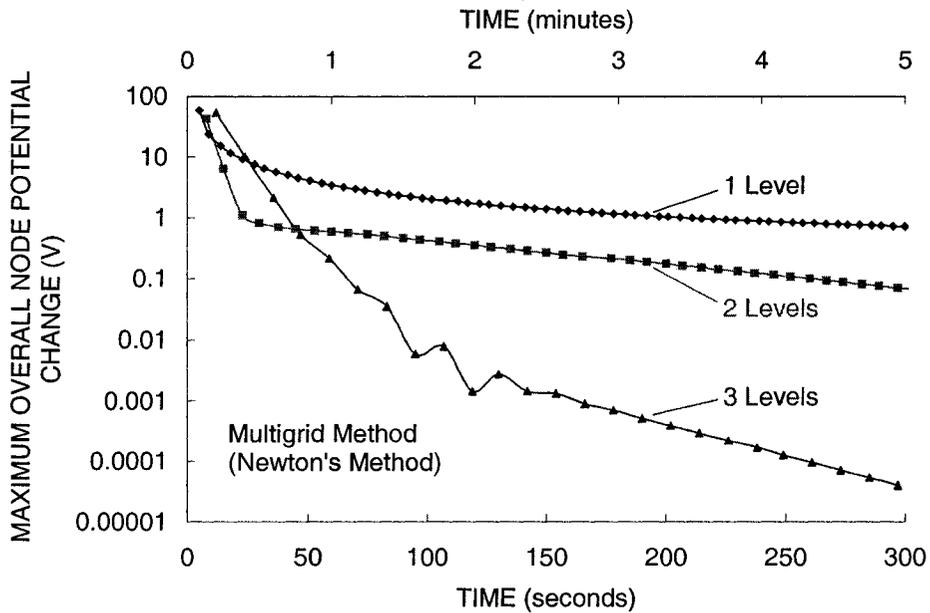


Figure 8.13 Comparison of the multigrid method using one, two, and three grids.

Ion density information is obtained while ion macro particles are tracked through the simulation volume. The Boltzmann electron equations are then used to solve for electron densities while solving for mesh potential values. Plots of potential, ion density, and electron density are shown in Figure 8.14 for an aperture of NSTAR grid dimensions operating at a beamlet current of 0.233 mA. Electron densities lower than $1 \cdot 10^{14}$ electrons/m³ are omitted to illustrate the absence of electrons in the intra-grid region.

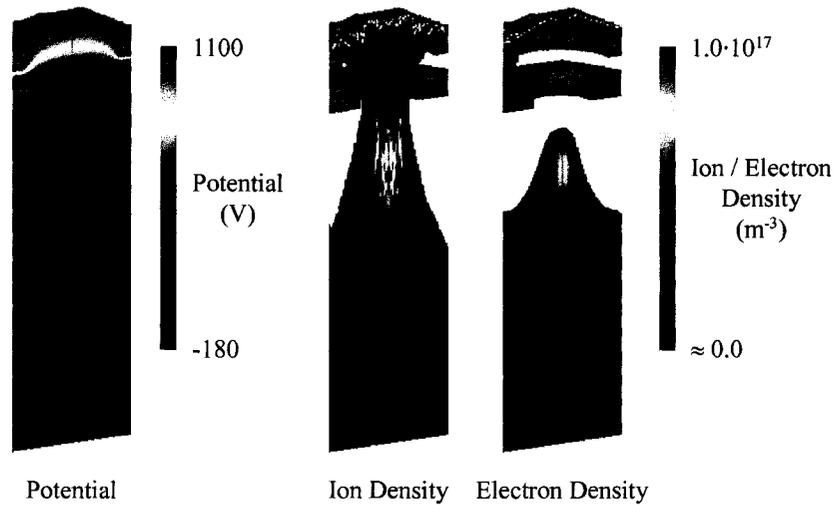


Figure 8.14 Electron density resulting from the combination of ion density and potential.

8.3 Symmetry

The upstream and downstream simulation surfaces are treated as Dirichlet boundaries, while the four remaining sides of the volume are treated as Neumann boundaries. A Dirichlet boundary condition is the case where the value of a function is known at the boundary, in this case the potentials of the upstream and downstream surfaces. The Neumann boundary condition describes a different case where the derivative of the function is known at the boundary. In this case, the derivative of potential through the four side boundaries is zero, resulting in potential symmetry.

Consider the Poisson equation. At one x boundary, where $i = 0$, the value of ϕ at $i = -1$ (ϕ_{-1}) is needed to compute the potential $\phi_{0,jk}$. However, ϕ_{-1} does not exist in the mesh. From the symmetry condition across this boundary, ϕ_{-1} is equal to ϕ_{+1} . Similarly, at the other x boundary, where $i = X$, the potential ϕ_{X+1} does not exist, and ϕ_{X-1} is used when needed instead.

When tracking an ion macro particle through the simulation volume, it may pass through symmetric boundaries. From the symmetry argument, when one particle exits the volume through a symmetric boundary, an identical particle is considered to be entering the volume from an adjacent region.

To simulate this, the particle is placed back into the simulation volume and its velocity and acceleration components are reversed depending on which boundary the particle passed through.

For example, a particle might end up with a negative x position, $x = -x_{i+1}$, following a particle push using the Leapfrog method. This indicates that the particle passed through the boundary at $x = 0$. Accordingly, the particle is placed back into the volume by setting its x position to $x = +x_{i+1}$. Also, its x direction velocity and acceleration components prior to the particle push are reversed, $v_{x_i} = -v_{x_i}$ and $a_{x_i} = -a_{x_i}$. The velocity and acceleration at time i then have the correct signs when computing the new particle velocity at time $i + 1$ using $v_{x_{i+1}} = v_{x_i} + \frac{1}{2}(a_{x_i} + a_{x_{i+1}})\Delta t$ in the Leapfrog method.

8.4 Distributions

Distributions are used in the ffx code primarily with regard to grid erosion and charge exchange ion production. In terms of grid erosion, a cosine distribution is used to describe how atoms are sputtered away from the grid surface. In simulating charge exchange ions, a Maxwell-Boltzmann speed distribution is used to obtain initial charge exchange ion speeds.

8.4.1 Cosine Distribution

Sputtering and gas-surface interactions are two areas in which the cosine distribution arises related to ion optics modeling.

The differential sputter yield of many materials over a range of energies is described reasonably well by a diffuse cosine law. The differential cosine distribution $y(\alpha)$, in units of atoms per ion per steradian (atoms/(ion·Ω)) is given in Eq. 8.40, where alpha (α) is the polar angle measured from the surface normal vector. Using this distribution, the differential sputter yield is not a function of the azimuthal angle theta (θ).

$$y(\alpha) = \frac{Y \cos(\alpha)}{\pi} \quad \text{Eq. 8.40}$$

The integral of the differential sputter yield over the hemisphere above a surface should yield the total sputter yield, Y , in units of atoms per ion (atoms/ion) as shown in Eq. 8.41. The differential solid angle is $d\omega$, given in Eq. 8.42. Combining the differential yield with the differential solid angle leads to Eq. 8.43, which can be integrated easily using the trigonometric identity in Eq. 8.44.

$$Y = \iint_S y(\alpha, \theta) d\omega \quad \text{Eq. 8.41}$$

$$d\omega = \sin(\alpha) d\alpha d\theta \quad \text{Eq. 8.42}$$

$$Y = \int_0^{\pi} \int_0^{2\pi} \frac{Y \cos(\alpha)}{\pi} \sin(\alpha) d\alpha d\theta \quad \text{Eq. 8.43}$$

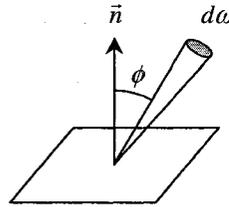
$$\cos(\alpha) \sin(\alpha) = \frac{1}{2} \sin(2\alpha) \quad \text{Eq. 8.44}$$

Studies done by Knudsen in the early 1900s on the flow of rarefied gases through tubes led to what is known as the Knudsen cosine law [Feres]. A rarefied gas is one in which the Knudsen number λ/r is large, where λ is the mean free path of the neutral gas molecules and r is a characteristic length of the system, such as the diameter of a tube. In the high Knudsen number regime, where λ/r is roughly greater than about 10, the gas molecules will essentially hit the walls or surfaces of the system much more often than other gas molecules. In Eq. 8.45, k is Boltzmann's constant, T_n is the temperature of the neutral molecules, p is the pressure, and d is the diameter of the molecules.

$$\frac{\lambda}{r} = \frac{1}{r} \frac{1}{\sqrt{2\pi} d^2 n} = \frac{1}{r} \frac{1}{\sqrt{2\pi} d^2} \frac{kT_n}{p} \quad \text{Eq. 8.45}$$

Knudsen proposed that the direction in which gas molecules reflect from a surface is independent of their impacting angles. In that case, the reflection of molecules was governed by a diffuse cosine law. The probability ds that the particle will reflect within the solid angle $d\omega$ was given by Eq. 8.46.

$$ds = \frac{d\omega}{\pi} \cos(\phi)$$



$$\text{Eq. 8.46}$$

Notice that the form of the differential cosine sputter yield $y(\alpha)$ and the gas-surface reflection probability ds is the same, where α and ϕ are measured identically. From this point on, the polar angle will be phi (ϕ) and the azimuthal angle will be theta (θ).

In the spherical coordinate system, the solid angle $d\omega$ is given by Eq. 8.47, which can be substituted into Eq. 8.46 to obtain Eq. 8.48. The function $f(\phi, \theta)$ will be defined as being part of this equation and is called a probability density function, or pdf, as shown in Eq. 8.49. In this case, the probability density function is the probability that a particle will leave the surface at the angle ϕ . One thing to notice about the pdf is that $f(\phi, \theta) \geq 0$ for the range of values over which it is valid, i.e., $0 \leq \phi \leq \frac{\pi}{2}$ and $0 \leq \theta \leq 2\pi$.

$$d\omega = d\phi \cdot \sin(\phi) d\theta \quad \text{Eq. 8.47}$$

$$\begin{aligned}
ds &= \frac{1}{\pi} \sin(\phi) \cos(\phi) d\phi d\theta \\
&= f(\phi, \theta) d\phi d\theta
\end{aligned}
\tag{Eq. 8.48}$$

$$f(\phi, \theta) = \frac{1}{\pi} \sin(\phi) \cos(\phi)
\tag{Eq. 8.49}$$

Because Eq. 8.49 does not depend on the angle θ , it can be integrated from $\theta = 0$ to $\theta = 2\pi$ to obtain the pdf as a function of ϕ alone, as given in Eq. 8.50. At this point, to conceptually keep the integrated angle θ separate from the angle ϕ , the uppercase letter theta, Θ , will be defined to be equal to 2π .

$$\begin{aligned}
f(\phi) &= \int_0^{2\pi} \left(\frac{1}{\pi} \sin(\phi) \cos(\phi) d\phi \right) d\theta \\
f(\phi) &= \frac{2\pi}{\pi} \sin(\phi) \cos(\phi) d\phi
\end{aligned}
\tag{Eq. 8.50}$$

$$\Theta = 2\pi \quad f(\phi) = \frac{\Theta}{\pi} \sin(\phi) \cos(\phi) d\phi$$

For a continuous distribution, the probability that the particle reflects at a single angle is zero, because the differential area $d\phi$ is zero for a single point, so an integral is done to describe the probability that the reflection angle is within a certain range of angles. In general, the cumulative distribution function, or cdf, that corresponds to a pdf is the integral from the lowest possible value of the pdf to an arbitrary value of the pdf. In this case, the cdf is the integral of the pdf from $\phi = 0$ to an arbitrary angle $\phi = \phi$, as in Eq. 8.51.

$$\begin{aligned}
F(\phi) &= \int_0^{\phi} f(\phi) d\phi \\
&= \int_0^{\phi} \frac{\Theta}{\pi} \sin(\phi) \cos(\phi) d\phi
\end{aligned}
\tag{Eq. 8.51}$$

Trigonometric identities can be used to perform this integral. In this case, the sum identity for the sine function will be used, Eq. 8.52, which results in Eq. 8.53.

$$\begin{aligned}
\sin(x \pm y) &= \sin(x) \cos(y) \pm \cos(x) \sin(y) \\
\sin(2x) &= 2 \sin(x) \cos(x)
\end{aligned}
\tag{Eq. 8.52}$$

$$\begin{aligned}
F(\phi) &= \int_0^{\phi} \frac{\Theta}{\pi} (\sin(\phi) \cos(\phi)) d\phi \\
&= \int_0^{\phi} \frac{\Theta}{\pi} \left(\frac{1}{2} \sin(2\phi) \right) d\phi \\
&= \frac{\Theta}{\pi} \frac{1}{4} (1 - \cos(2\phi))
\end{aligned}
\tag{Eq. 8.53}$$

Once again, a trig identity can be used to further simplify this expression. In this case, the sum formula for the cosine function will be used, given as Eq. 8.54.

$$\begin{aligned}\cos(x \pm y) &= \cos(x)\cos(y) \mp \sin(x)\sin(y) \\ \cos(2x) &= \cos^2(x) - \sin^2(x) \\ &= \cos^2(x) - (1 - \cos^2(x)) = 2\cos^2(x) - 1 \\ &= (1 - \sin^2(x)) - \sin^2(x) = 1 - 2\sin^2(x)\end{aligned}\tag{Eq. 8.54}$$

To eliminate any confusion, the commonly used notation used for the trig functions is $\cos^2(x) = \cos(x) \cdot \cos(x)$ and $\sin^2(x) = \sin(x) \cdot \sin(x)$. In substituting the trig identity from Eq. 8.54 into the integral for the cdf in Eq. 8.53, one obtains Eq. 8.55.

$$\begin{aligned}F(\phi) &= \frac{\Theta}{\pi} \frac{1}{4} (1 - (1 - 2\sin^2(\phi))) \\ &= \frac{\Theta}{\pi} \frac{1}{2} \sin^2(\phi)\end{aligned}\tag{Eq. 8.55}$$

Similar to the pdf, the cdf also has certain intrinsic properties. For instance, since the pdf $f(\phi, \theta) \geq 0$ for all values of ϕ and θ , $F(\phi)$ is monotone increasing. Also, for the lowest value of ϕ , $F(\phi = 0) = 0$, and for the greatest value of ϕ , $F(\phi = \pi/2) = 1$. The pdf and cdf for the cosine distribution are shown in Figure 8.15.

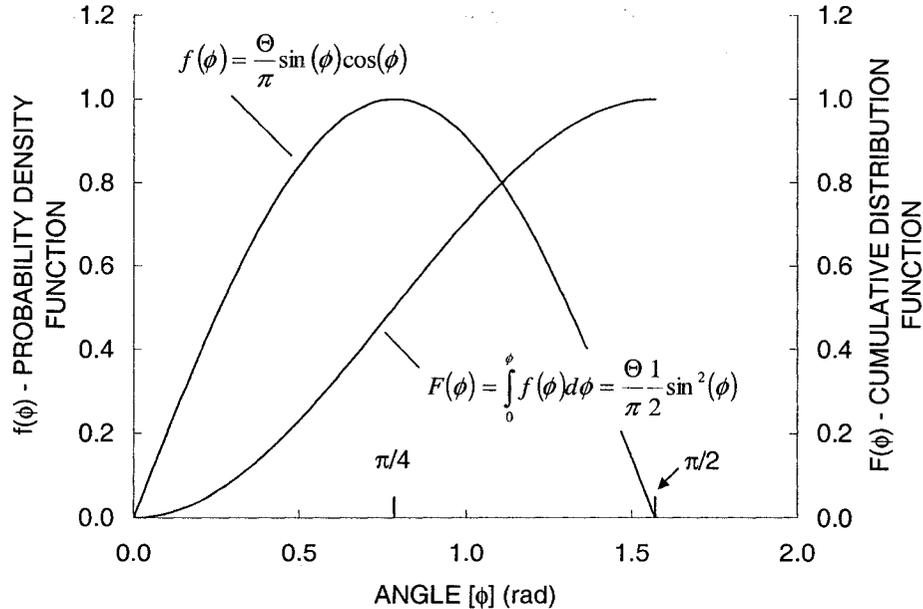


Figure 8.15 The probability density function (pdf) and cumulative distribution function (cdf) for the cosine distribution.

8.4.2 Distribution Sampling

Useful information is obtained from a distribution through sampling [Saucier]. A distribution is sampled in order to obtain an individual instance of the distribution. For example, in terms of sputtering, the cosine distribution is sampled to determine what direction individual atoms will travel away from an ion impact site. The Maxwell-Boltzmann distribution is sampled to determine what speed to give a newly created ion following a charge exchange reaction.

Two of the most common ways of sampling from a distribution will be discussed here. One is referred to as rejection sampling and the other is referred to as inverting a cdf. In both cases, the idea is to take random numbers from an available random number generator and use them to obtain samples of values from a distribution.

Random number generators usually generate roughly even values of numbers over the range of 0.0 to 1.0 (inclusive of 0.0 and 1.0). The idea behind both rejection sampling and inverting a cdf is to transform the distribution of numbers generated by the random number generator into the desired distribution, such as the cosine distribution.

In rejection sampling, two random values, R_1 and R_2 , obtained from the random number generator are used to test for possible values of ϕ that satisfy the probability density function (pdf).

First, a value M is found that is greater than any single probability over the range of ϕ values for which the distribution is valid. For the cosine distribution, this is given in Eq. 8.56.

$$f(\phi) \leq M \text{ for } 0 \leq \phi \leq \frac{\pi}{2} \quad \text{Eq. 8.56}$$

The first random value R_1 is used to generate a test angle ϕ_t over the range $0 \leq \phi \leq \frac{\pi}{2}$. This is done by linearly interpolating over the range of valid ϕ values, as in Eq. 8.57. In a general sense not relating to the specific problem at hand, if one were looking for test value X over the range from a to b , one would use $X = a + R_1(b - a)$.

$$\phi_t = 0 + R_1 \left(\frac{\pi}{2} - 0 \right) = R_1 \frac{\pi}{2} \quad \text{Eq. 8.57}$$

The second random value R_2 is used to generate a random value Y over the range $0 \leq Y \leq M$. Here again, linear interpolation is used, as in Eq. 8.58.

$$Y = 0 + R_2(M - 0) = R_2 M \quad \text{Eq. 8.58}$$

The test angle ϕ_t is accepted as a valid angle from the distribution if $Y \leq f(\phi_t)$. If $Y > f(\phi_t)$, then the test angle ϕ_t is rejected and the process starts over again.

It is observed that the most efficient value of M to use is to choose M such that it is exactly equal to the maximum value of the pdf. In this case, the maximum value of the pdf occurs when $\phi = \frac{\pi}{4}$, thus

$$M = f\left(\phi = \frac{\pi}{4}\right) = 1.$$

If the equations for ϕ_t and Y are combined with $M = 1$, one obtains a very simple test to accept or reject the value for ϕ_t . In simplest terms, the test value $\phi_t = R_1 \frac{\pi}{2}$ is accepted only if

$$R_2 \leq \frac{\Theta}{\pi} \sin\left(R_1 \frac{\pi}{2}\right) \cos\left(R_1 \frac{\pi}{2}\right).$$

Once a suitable angle ϕ_t is found, the angle θ_t can be found using yet another randomly chosen value R_3 . The angle θ_t is linearly interpolated over the range $0 \leq \theta \leq \Theta$ where $\Theta = 2\pi$ from before, as in Eq. 8.59.

$$\theta_t = 0 + R_3(\Theta - 0) = R_3(2\pi) \quad \text{Eq. 8.59}$$

In rejection sampling, Figure 8.16, the process of testing for possible angles continues until a valid angle is found. Because test angles can be rejected, the computer likely does excess computations before an acceptable test angle is found. The rate of acceptance is equal to the area under the curve $y = f(\phi)$ divided by the area under the curve $y = M$. By definition, the area under $y = f(\phi)$ is unity, i.e. $F\left(\phi = \frac{\pi}{2}\right) = 1$, while the area under $y = M$ is equal to $M \frac{\pi}{2}$. Thus, the rate of acceptance for sampling from the cosine distribution is given in Eq. 8.60, where M has been set to unity. It is observed that in this case, the acceptance rate decreases linearly with the choice of the value M .

$$\begin{aligned} \text{Acceptance Rate} &= \frac{\text{Area Under } y = f(\phi)}{\text{Area Under } y = M} \\ &= \frac{1}{M \frac{\pi}{2}} = \frac{2}{M\pi} = \frac{2}{(1)\pi} \approx 63.7\% \end{aligned} \quad \text{Eq. 8.60}$$

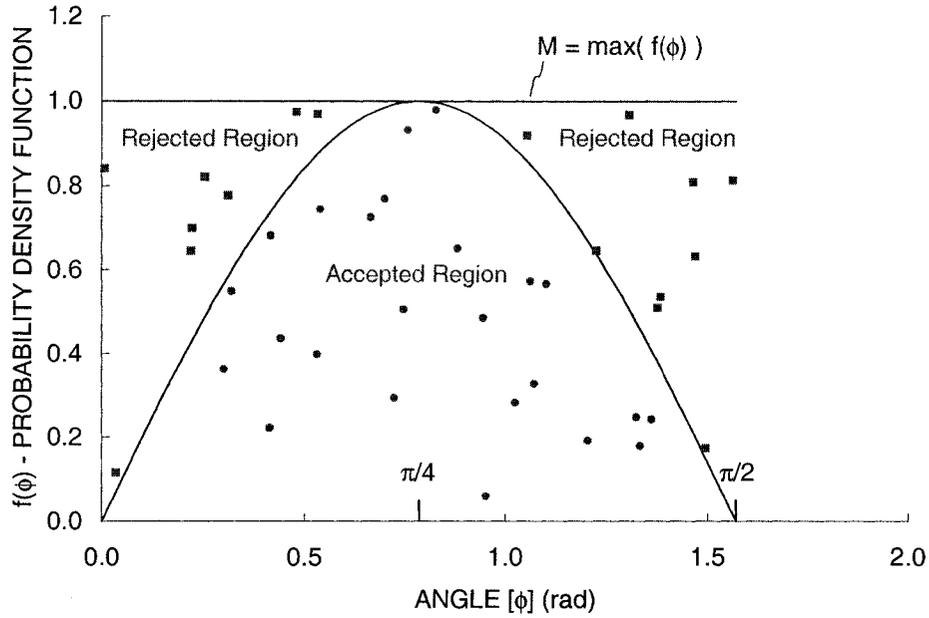


Figure 8.16 Rejection sampling with the cosine distribution.

Using the inverted cumulative distribution function (cdf) to sample the distribution is a more efficient alternative to using rejection sampling. This is because a value from the distribution is guaranteed to be found for every random number that is chosen. The only reason why inversion would not be used instead of rejection sampling is if it is not possible to solve for the cdf and/or invert the cdf easily.

Fortunately, the cdf for the cosine distribution can be found, and it is invertible. From before, the cdf for the cosine distribution is given in Eq. 8.61.

$$F(\phi) = \frac{\Theta}{\pi} \frac{1}{2} \sin^2(\phi) = \frac{2\pi}{\pi} \frac{1}{2} \sin^2(\phi) = \sin^2(\phi) \quad \text{Eq. 8.61}$$

To find suitable values for ϕ using the random number generator, the cdf is inverted to solve for ϕ in terms of the value of the cdf, $F(\phi)$. Because the cumulative distribution function varies over the range 0 to 1 and the random value R_1 also varies over the range 0 to 1, $F(\phi)$ is simply set equal to R_1 and the value of ϕ is calculated directly, as in Eq. 8.62.

$$\frac{\Theta}{\pi} \frac{1}{2} \sin^2(\phi) = F(\phi) = R_1 \quad \text{Eq. 8.62}$$

$$\phi = \sin^{-1} \left(\sqrt{\frac{\pi}{\Theta} 2R_1} \right) = \sin^{-1} \left(\sqrt{\frac{\pi}{2\pi} 2R_1} \right) = \sin^{-1}(\sqrt{R_1})$$

As when using the rejection sampling method, the value of θ is found using a separately generated random number, R_2 , by linearly interpolating over the possible range of θ values, as in Eq. 8.63.

$$\theta = 0 + R_2(\Theta - 0) = R_2(2\pi) \quad \text{Eq. 8.63}$$

8.4.3 Maxwell-Boltzmann Distribution

The distribution of neutral atom speeds is described by the Maxwell-Boltzmann speed distribution, given in Eq. 8.64. The speed distribution is only one form of the Maxwell-Boltzmann distribution. Other forms, for instance, give information about particle velocities and energies. To clarify, the speed in this case has been called v , where v is really the magnitude of the particle velocity vector.

$$f(v) = 4\pi \left(\frac{m}{2\pi kT} \right)^{\frac{3}{2}} v^2 \exp\left(\frac{-mv^2}{2kT} \right) \quad \text{Eq. 8.64}$$

$$\text{where } v = \sqrt{v_x^2 + v_y^2 + v_z^2}$$

Several useful quantities can be derived from the speed distribution, given in Eq. 8.65. One is called the most probable velocity, v_{mp} . The most number of neutral atoms will have the most probable velocity, and it is the maximum value of the Maxwell-Boltzmann speed distribution. A second value is the average velocity, \bar{v} . The average velocity is shifted to the right of the most probable velocity because the distribution is not symmetric. The distribution's tail at higher energies makes the average velocity greater than the most probable velocity. A third velocity is the root mean squared velocity, v_{rms} . It is found essentially by squaring the velocity of each particle, adding them together, and then taking the square root.

$$\begin{aligned} v_{mp} &= v \left(\frac{df(v)}{dv} = 0 \right) = \sqrt{\frac{2kT}{m}} \\ \bar{v} &= \int_0^{\infty} v f(v) dv = \sqrt{\frac{8kT}{\pi m}} \\ v_{rms} &= \sqrt{\int_0^{\infty} v^2 f(v) dv} = \sqrt{\frac{3kT}{m}} \end{aligned} \quad \text{Eq. 8.65}$$

Rejection sampling from the Maxwell-Boltzmann distribution is illustrated here, as the distribution is not easily invertible. Figure 8.17 shows the Maxwell-Boltzmann distribution for xenon at a temperature of 500 K.

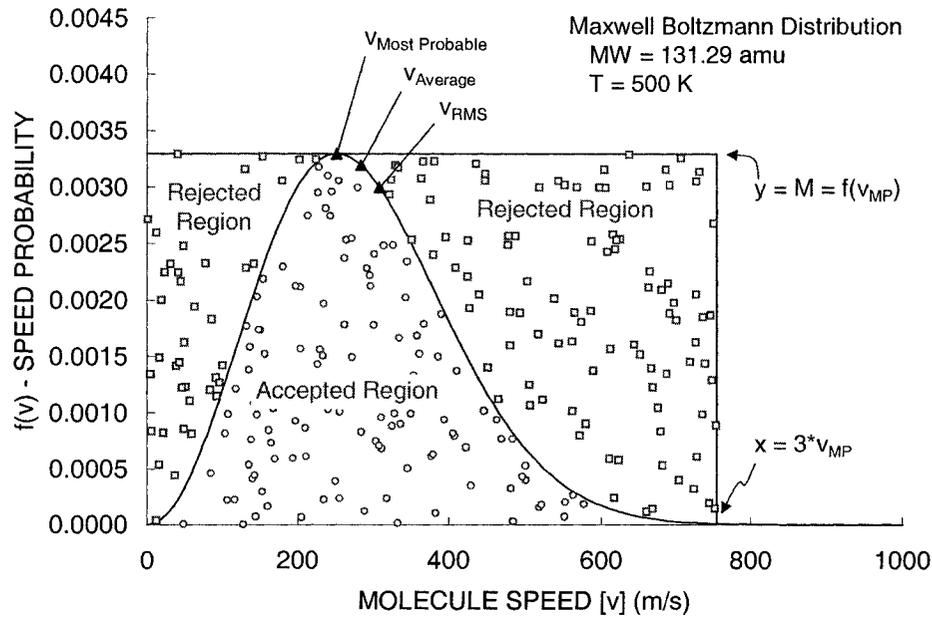


Figure 8.17 An example Maxwell-Boltzmann speed distribution.

The best choice for the value M is to set it equal to the speed probability at the most probable velocity. In mathematical terms, $M = f(v_{mp})$. This will result in the greatest acceptance rate of test particle velocities.

In this case, the Maxwell-Boltzmann distribution is valid for particle speeds (in m/s) from zero to infinity. In using rejection sampling, test particle speeds, v_t , must be chosen over a much smaller speed range to keep the acceptance rate at a reasonable level. Inspection of the Maxwell-Boltzmann distribution shows that roughly 99.95 percent of neutral particles have speeds less than $3v_{mp}$. In other words, it is very unlikely that many particles will have speeds greater than $3v_{mp}$. This simplification allows particle test velocities to be sampled from the range $0 \leq v_t \leq 3v_{mp}$, as in Eq. 8.66.

$$Y = \text{RandNum1} \cdot M = \text{RandNum1} \cdot f(v_{mp})$$

$$v_t = \text{RandNum2} \cdot 3v_{mp}$$

Eq. 8.66

The test particle speed, v_t , is accepted as a valid speed from the distribution if $Y \leq f(v_t)$. Otherwise, new random numbers are chosen until a valid particle speed is found.

The acceptance rate is equal to the area under the distribution divided by the total area under the curve $y = M$. In this case, the area under the distribution over the sampling range $0 \leq v_t \leq 3v_{mp}$ is not exactly equal to unity, as shown in Eq. 8.67.

$$\text{Acceptance Rate} = \frac{\int_0^{3v_{mp}} f(v)dv}{M \cdot 3v_{mp}} \cdot 100 \quad \text{Eq. 8.67}$$

$$\approx \frac{0.99956}{0.0033 \cdot 3 \cdot 251.7} \cdot 100 = 40.1 \%$$

8.5 Neutral Density

An analytical model is used to describe the neutral density variation near the grids. In addition to the geometry of the grids, there are five main values input either directly or indirectly into the ffx code that affect the neutral density calculations. These values are the overall beamlet current, J_b , the propellant atom weight, m_n , the propellant temperature, T_n , the double-to-single ion current ratio, J_b^{++}/J_b^+ , and the propellant utilization efficiency, η .

The double-to-single ion current ratio, J_b^{++}/J_b^+ , has a small effect on the neutral density calculations with regard to the propellant flow rate. The beamlet current, J_b , is assumed to be comprised of a singly charged ion current, J_b^+ , and a doubly charged ion current, J_b^{++} . The singly and doubly charged ion currents can be found from the overall beamlet current and the double-to-single current ratio, as in Eq. 8.68.

$$J_b = J_b^+ + J_b^{++} = J_b^+ \left(1 + \frac{J_b^{++}}{J_b^+} \right)$$

$$J_b^+ = \frac{J_b}{1 + \frac{J_b^{++}}{J_b^+}} \quad \text{Eq. 8.68}$$

$$J_b^{++} = J_b^+ \cdot \frac{J_b^{++}}{J_b^+}$$

The propellant utilization efficiency, η , is defined as the ratio of the number of propellant atoms that exit through the aperture pair as ions to the total number of propellant atoms flowing through the aperture pair. The propellant utilization efficiency can be calculated using the overall beamlet current, J_b , the double-to-single current ratio, J_b^{++}/J_b^+ , and the propellant flow rate, \dot{m}_{Aeq} , in Amps equivalent (Aeq.), as in Eq. 8.69.

$$\eta_u = \frac{\frac{J_b^+}{e} + \frac{J_b^{++}}{2e}}{\frac{\dot{m}}{m_n}} = \frac{J_b^+}{e} \left(1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+} \right) \frac{m_n}{\dot{m}}$$

$$= \frac{\left(1 + \frac{J_b^{++}}{J_b^+} \right) \left(1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+} \right)}{\frac{\dot{m}e}{m_n}}$$

Eq. 8.69

$$= \frac{\left(1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+} \right) J_b}{\left(1 + \frac{J_b^{++}}{J_b^+} \right) \frac{\dot{m}e}{m_n}} = \frac{\left(1 + \frac{1}{2} \frac{J_b^{++}}{J_b^+} \right) J_b}{\left(1 + \frac{J_b^{++}}{J_b^+} \right) \dot{m}_{Aeq}}$$

$$\dot{m}_{Aeq} = \frac{\dot{m}e}{m_n}$$

The average thermal velocity of the neutral propellant atoms is given in Eq. 8.70. In this equation, k_B is the Boltzmann constant (in J/K), T_n is the neutral atom gas temperature (in K) which is taken to be equal to the discharge chamber wall temperature, and m_n is the mass of the neutral propellant atom (in kg).

$$v_n = \sqrt{\frac{8k_B T_n}{\pi m_n}}$$

Eq. 8.70

As an example, the neutral atom speed is calculated for xenon atoms at a temperature of 500 K in Eq. 8.71.

$$k_B = 1.3807 \cdot 10^{-23} \left\{ \frac{J}{K} \right\}$$

$$T_n \approx 500 \{K\}$$

Eq. 8.71

$$m_n = m_{xenon} = 131.29 \cdot 1.6605 \cdot 10^{-27} = 2.18 \cdot 10^{-25} \{kg\}$$

$$v_{n \text{ xenon @ } 500K} = 284 \left\{ \frac{m}{s} \right\}$$

The neutral atom flow rate, \dot{N} , is constant through each grid hole. It is found by subtracting the singly charged and doubly charged ion flow rates from the total propellant atom flow rate, as in Eq. 8.72.

$$\dot{N} = \frac{\dot{m}}{m_n} - \frac{J_b^+}{e} - \frac{J_b^{++}}{2e}$$

$$= \frac{\dot{m}}{m_n} (1 - \eta_u)$$

Eq. 8.72

The flow rate of atoms in a particular direction at a certain location is equal to the density of atoms at that location multiplied by the area through which the atoms are passing multiplied by the velocity of the atoms in that direction. In this case, the directional velocity of the propellant atoms is 1/4th of their thermal velocity. At the downstream exit of the last grid, the flow rate of neutral atoms in the downstream direction is equal to the flow rate of neutral atoms through the aperture pair, as in Eq. 8.73.

$$n_e A_e \frac{1}{4} v_n = \dot{N} \text{ or } n_e = \frac{4\dot{N}}{v_n A_e} \quad \text{Eq. 8.73}$$

For the typical case where the apertures are circular in nature, the exit area used in the above equation is simply the circular open hole area, Eq. 8.74. Here, d_e is the diameter of the holes in the last grid in the system.

$$A_e = \frac{\pi d_e^2}{4} \quad \text{Eq. 8.74}$$

Clousing factors are used to describe the resistance to flow of the neutrals through the grids. A curve fit is used to obtain the Clousing factor for each grid as a function of each grid's thickness, t , and hole diameter, d , as in Eq. 8.75.

$$Fc(t, d) = 0.9968 - 0.9317 \left(\frac{t}{d} \right) + 0.6967 \left(\frac{t}{d} \right)^2 - 0.2997 \left(\frac{t}{d} \right)^3 + 0.0524 \left(\frac{t}{d} \right)^4 \quad \text{Eq. 8.75}$$

Through any grid in the system, the difference between the flow of neutrals in the downstream and upstream directions is equal to the net flow rate of neutrals, \dot{N} , through the aperture pair, as given in Eq. 8.76. Here, n_i is the neutral density at the upstream edge of the grid and n_{i+1} is the neutral density at the downstream edge of the grid. The area A_i is the open area through which the neutrals pass, equal to the total aperture area multiplied by the physical open area fraction.

$$\dot{N} = n_i A_i Fc_i \frac{1}{4} v_n - n_{i+1} A_i Fc_i \frac{1}{4} v_n$$

$$n_i = n_{i+1} + \frac{4\dot{N}}{v_n A_i Fc_i} \quad \text{Eq. 8.76}$$

The variation of the neutral density through each grid is taken to be linear, varying from n_i to n_{i+1} with the axial distance z . Also, the neutral density between each grid is taken to be constant.

As a special case, the backflow of neutrals downstream of the last grid in the system can be considered to be zero. In this case, the neutral density at the downstream edge of the last grid, n_e , is still calculated in the same way as before, but the neutral density at the upstream edge of the last grid, n_{e-1} , is calculated without using the downstream neutral density n_e , as in Eq. 8.77.

$$n_{e-1} = \frac{4\dot{N}}{v_n A_e F C_e} \quad \text{Eq. 8.77}$$

The method of characterizing the neutral density decay downstream of the grids is somewhat more arbitrary. The neutral density downstream of each aperture is complicated by the neutral flow from surrounding apertures and by the overall thruster diameter. The model used in the ffx code is based on an expansion using a characteristic length in relation to the grid dimensions. This length has been chosen through simulation in order to match experimental results of grid impingement current. The neutral density decay downstream of the last grid is given in Eq. 8.78, where z is the axial distance measured from the downstream side of the last grid in the system to the point of interest.

$$n(z) = \frac{n_e}{\left(1 + 2\frac{z}{l_{char}}\right)^2} \quad \text{where } l_{char} \approx 5 \cdot l_{cc} \quad \text{Eq. 8.78}$$

Frequently, special modeling of the neutral density near the downstream simulation plane is performed. In order to match experimentally observed grid impingement currents, the neutral density in the region near the downstream plane is adjusted automatically by the code during the simulation. This in turn causes the charge exchange ion production rates to be adjusted linearly according to the charge exchange ion production equation. Additional care is often taken to axially align the charge exchange ion injection vectors, more closely approximating charge exchange ions originating far downstream.

With respect to a full thruster, the neutral density within the discharge chamber can be calculated more directly using Eq. 8.79, where \dot{N} is calculated using the overall flow rate and propellant utilization efficiency, A_B is the total area over which ions are extracted (including both closed and open area), and ϕ is the overall transparency to ions. For a two grid system, the overall transparency is calculated using Eq. 8.80 where ϕ_s is the physical open area fraction of the screen grid for example. The pressure within the discharge chamber is then given by Eq. 8.81.

$$n_n = \frac{4\dot{N}}{v_n A_B \phi} = \frac{4\frac{\dot{m}}{m_n}(1-\eta_u)}{\sqrt{\frac{8k_B T_n}{\pi m_n}} A_B \phi} \quad \text{Eq. 8.79}$$

$$\frac{1}{\phi} = \frac{1}{\phi_s F C_s} + \frac{1}{\phi_a F C_a} \quad \text{or } \phi = \frac{\phi_s F C_s \phi_a F C_a}{\phi_s F C_s + \phi_a F C_a} \quad \text{Eq. 8.80a}$$

$$P = nk_B T_n \quad \text{Eq. 8.81}$$

8.6 Random Number Generators

Perhaps the most common kind of random number generator is one that produces random numbers uniformly distributed in the range from zero to one, inclusive of zero and one. Mathematical operations are often applied to this base kind of random number generator to provide random number generators that sample from other types of distributions.

Ideally, a random number generator should produce numbers that are truly independent of each other. Physical devices exist that produce random numbers based on processes known to be truly unpredictable, such as radio active decay or the electrical noise that comes from resistors and semiconductors.

More common than using a physical source of random numbers is to use a computer algorithm. Random number generators generally fall into two broad categories: pseudo-random number generators and quasi-random number generators. Pseudo-random number generators attempt to be as close to truly random as possible. Quasi-random number generators are more predictable in their randomness, and are used for specific applications.

In general, any random number generator has to pass several tests in order to be considered a “good” random number generator. When random numbers are generated using an algorithm, an initial state called the seed state is usually used to start the random number generator at a certain point in its sequence. The period of a generator describes its length, or the number of numbers that can be generated before the generator returns to the same state where it started. The period of the generator should ideally be much greater than the number of numbers to be obtained from the generator. The most important feature of a generator should be its uniformity of random number generation. Other nice properties to have in a generator are a short time to access each random number and a low memory requirement.

8.6.1 Pseudo-Random Number Generators

Many types of pseudo-random number generators have been devised. Some examples are linear congruential generators, lagged Fibonacci generators, linear feedback shift registers and generalized feedback shift registers [Press].

Already imbedded within the C++ programming language is the `rand()` function. Calling `rand()` produces integers from zero to `RAND_MAX`, inclusive. The value of `RAND_MAX` is defined within each specific C++ implementation, but it is at least 32,767 in all cases. One source suggests that only about 20,000 useful random numbers can be obtained if `RAND_MAX` is defined as 32,767.

The C++ random number generator is initialized with a call to `srand(seed)`, where `seed` is an integer that initializes the state of the random number generator. For instance, the seed value could be 0, 1, or 10000. Following the call to `srand`, subsequent calls to `rand()` produce integers in the range

$0 \leq rand() \leq RAND_MAX$. Dividing the integers obtained with `rand()` by `RAND_MAX` then produces random numbers in the range $0 \leq r \leq 1$.

Another pseudo-random number generator with increasing popularity is the Mersenne Twister, first developed in 1997 by Makoto Matsumoto and Takuji Nishimura [Matsumoto]. It is based on a twisted generalized linear feedback shift register and has a period of $2^{19937}-1$, which is longer than any other generator to date. It is thought to be at least as fast as the `rand()` function when sampling random numbers.

8.6.2 Quasi-Random Number Generators

Quasi-random number generators generate numbers based on low discrepancy sequences. In certain instances, one wishes to obtain numbers that are roughly evenly distributed over a certain range. Quasi-random sequences know how to “fill in” a certain range. There are many commonly used low discrepancy sequences, including Faure, Niederreiter, Sobol, Halton, Hammersley, Van der Corput, Centroidal Voronoi, and Tessellation [Bratley, Fox].

The Faure, Niederreiter, and Sobol sequences are compared here. In these sequences, one has to know the number of dimensions in which to generate points. For instance, if one is generating random points in a rectangle, the sequences are called specifically to generate two-dimensional points.

In the `ffx` code, roughly 10,000 random ion macro particle starting points are used on each code iteration. Figure 8.18 shows the first 10,000 randomly chosen two-dimensional points using the Mersenne Twister and Niederreiter random number generators. The C++ `rand()` function would generate similar points to the Mersenne Twister, and the Faure and Sobol sequences would generate similar points to the Niederreiter sequence. One can easily see in this plot how even the Niederreiter points are compared to the Mersenne Twister points.

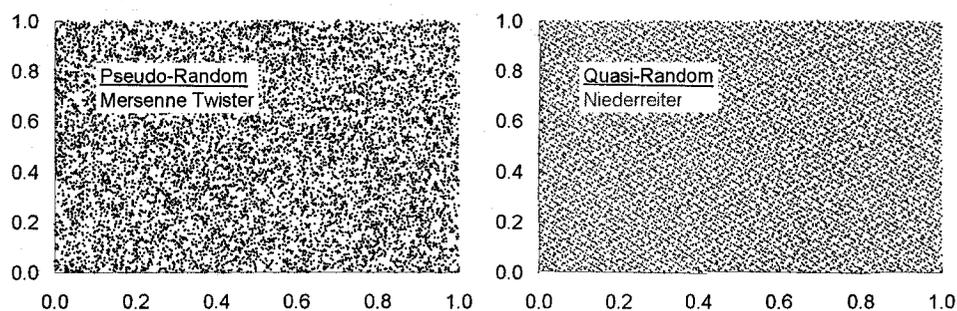


Figure 8.18 The first 10,000 Mersenne Twister and Niederreiter randomly generated two-dimensional points.

A second comparison between the pseudo- and quasi-random number generators is shown in Figure 8.19. Here, a histogram of the first 10,000 points generated with each number generator was created in ten increments. Ideally, there would be exactly 1000 points in each bin for a perfectly even distribution of points. Because the Faure, Niederreiter, and Sobol sequences are designed to uniformly fill in the range

from 0 to 1, there is very little variation from 1000 points in each bin. Because the C++ rand() function and the Mersenne Twister generators are supposed to be more random, there is more variation among the bins. Additionally, the rand() function is seen to be less uniform than the Mersenne Twister generator. There are only 902 points generated by the rand() function in the bin from 0.2 to 0.3 for instance.

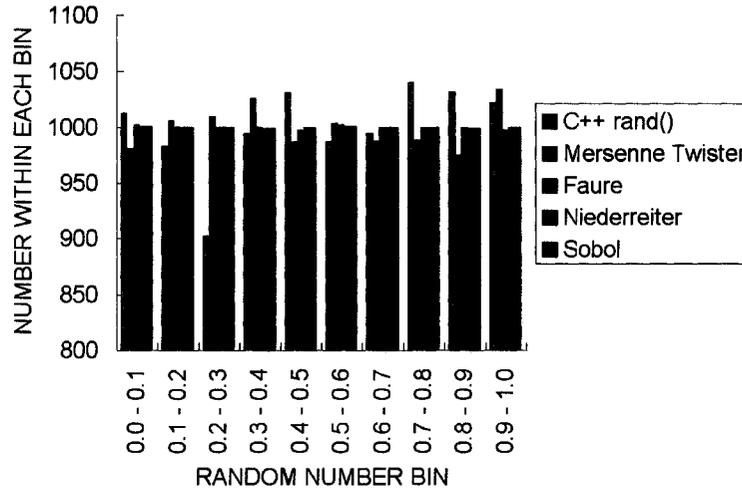


Figure 8.19 Histogram of the first 10,000 points generated with the pseudo- and quasi-random number generators.

Table 8.2 lists the first ten three dimensional points obtained with the Mersenne Twister and Niederreiter random number generators. Expectedly, one cannot easily see a pattern in the Mersenne Twister numbers. However, one can easily see that the Niederreiter points have a pattern, especially in the first two columns of numbers. Each successive call to the Niederreiter algorithm gives a new point that helps to fill in the domain evenly. The two dimensional Niederreiter sequence consists of the first two columns of numbers in the table.

Table 8.2 Comparison of three dimensional random points generated with the Mersenne Twister and Niederreiter random number generators.

Point (<i>i</i>)	Mersenne Twister (Pseudo-Random)			Niederreiter (Quasi-Random)		
1	0.549	0.593	0.715	0	0	0
2	0.844	0.603	0.858	0.5	0.5	0.75
3	0.545	0.847	0.424	0.75	0.25	0.3125
4	0.624	0.646	0.384	0.25	0.75	0.5625
5	0.438	0.298	0.892	0.375	0.375	0.875
6	0.057	0.964	0.273	0.875	0.875	0.125
7	0.383	0.478	0.792	0.625	0.125	0.6875
8	0.812	0.529	0.480	0.125	0.625	0.4375
9	0.568	0.393	0.926	0.1875	0.3125	0.51563
10	0.836	0.071	0.337	0.6875	0.8125	0.26563

The ffx code uses the Mersenne Twister and Niederreiter random number generators. The Niederreiter random number generator is mainly used to obtain a two dimensional sequence of numbers that serve as ion macro particle starting locations upon injection into the upstream simulation plane.

The Mersenne Twister random number generator is used in many instances within the ffx code. If the Niederreiter sequence is not used, the Mersenne Twister generator is used to find ion macro particle injection locations. Another instance of the Mersenne Twister generator is used to randomly determine if ions are singly or doubly charged. With respect to charge exchange ions, a Mersenne Twister generator is used to select random initial velocity vectors and to sample from the Maxwell-Boltzmann speed distribution for initial velocity magnitudes. With respect to sputtering, a generator is used to select grid material ejection angles from a cosine distribution. Mersenne Twister generators are also used in several facets of evolutionary algorithms.

9 References

Sheaths and Neutralization Surfaces

Boerner, J. J. and Boyd, I. D. "Numerical Simulation of Faraday Probe Measurements in a Multi-component Non-equilibrium Plasma." IEPC-2005-085, 29th International Electric Propulsion Conference, Princeton, NJ, October 2005.

Brown, Ian. *The physics and technology of ion sources*, John Wiley & Sons, New York, 1989.

Charge Exchange

J. Scott Miller, Steve H. Pullins, Dale J. Levendier, Yu-hui Chiu, and Rainer A. Dressler. "Xenon Charge Exchange Cross Sections for Electrostatic Thruster Models." *Journal of Applied Physics*, Vol. 91, No. 3, pages 984-991, Feb. 1 1992.

Donald Rapp and W. E. Francis. "Charge Exchange between Gaseous Ions and Atoms" *The Journal of Chemical Physics*, Volume 37, Number 11, Pages 2631-2645, Dec. 1 1962.

Shuji Sakabe and Yasukazu Izawa. "Cross Sections for Resonant Charge Transfer Between Atoms and Their Positive Ions: Collision Velocity ≤ 1 a.u." *Atomic Data and Nuclear Data Tables*, Volume 49, Issue 2, Pages 257-314, Nov. 1991.

Sputtering

R. P. Doerner, D. G. Whyte, and D. M. Goebel. "Sputtering Yield Measurements During Low Energy Xenon Plasma Bombardment." *Journal of Applied Physics*, Volume 93, Number 9, Pages 5816-5823, May 1 2003.

Robert D. Kolasinski "Oblique Angle Sputtering Yield Measurements for Ion Thruster Grid Materials" AIAA-2005-3526. 41st Joint Propulsion Conference and Exhibit, Tucson, AZ, July 2005.

Yamamura, Y., and Tawara, H. *Energy Dependence of Ion-Induced Sputtering Yields from Monatomic Solids at Normal Incidence*. *Atomic Data and Nuclear Data Tables* Vol. 62, No. 2, pp. 149-254, March 1996.

Zoerb, K. A., Williams, J. D., Williams, D. D., and Yalin, A. P. "Differential Sputtering Yields of Refractory Metals by Xenon, Krypton, and Argon Ion Bombardment at Normal and Oblique Incidences." 29th International Electric Propulsion Conference, Princeton, NJ. October 2005.

Grid Failure Mechanisms

Foster, John E. et al. "The High Power Electric Propulsion (HiPEP) Ion Thruster." AIAA-2004-3812. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL. July 2004.

Flatness Parameter

Soulas, G. C., Kamhawi, H. K., Patterson, M. J., Britton, M. A., and Frandina, M. M., "NEXT Ion Engine 2000 Hour Wear Test Results" AIAA-2004-3791. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL, July 2004.

Propellant Utilization Efficiency

Malone, S. P., and Hickman, T. A., "Neutral Flux Distribution Downstream of Ion Thruster Optics." Joint Army Navy NASA Air Force Conference. Monterey, CA, December 2005.

Offset Apertures

Okawa, Y., Hayakawa, Y., and Kitamura, S. "Experiments on Ion Beam Deflection Using Ion Optics with Slit Apertures." Japanese Journal of Applied Physics. Vol. 43, No. 3, 2004, pp. 1136-1143.

Fixed Emission Surface

Wilbur, P. J., Farnell, C. C., and Williams, J. D. "Ion Thruster Optical Performance Enhancement via Ion-emissive-surface Shaping." IEPC-2005-095. 29th International Electric Propulsion Conference. Princeton, NJ, October 2005.

Wilbur, P. J. "The Emissive Membrane Ion Thruster Concept." IEPC-2005-164. 29th International Electric Propulsion Conference. Princeton, NJ, October 2005.

Thruster Current Density

Reynolds, Thaine W. "Mathematical Representation of Current Density Profiles from Ion Thrusters." AIAA-1971-693. 1971.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition, ISBN 0-521-43108-5, Cambridge University Press, Cambridge, 1992.

Evolutionary Algorithms

Busetti, Franco. "Genetic Algorithms Overview."

Obitko, Marek and Slavík, Pavel. "Visualization of Genetic Algorithms in a Learning Environment." Spring Conference on Computer Graphics (SCCG), ISBN 80-223-1357-2, Comenius University, Bratislava, pp. 101-106, 1999.

Obitko, Marek. "An introduction to genetic algorithms with Java applets." "<http://cs.felk.cvut.cz/~xobitko/ga/>", 1998.

Pohlheim, Hartmut. "Evolutionary Algorithms: Principles, Methods, and Algorithms." Technical Report, Technical University Ilmenau, 1994-1997.

Srinivas, M. and Patnaik, L. M. "Adaptive Probabilities of Crossover and Mutation in Genetic Algorithms." IEEE Transactions on Systems, Man and Cybernetics, Vol. 24, No. 4, April 1994.

Whitley, Darrell. "A Genetic Algorithm Tutorial." *Statistics and Computing*, Volume 4, pp. 65-85, 1994.

Whitley, Darrell. "An Overview of Evolutionary Algorithms: Practical Issues and Common Pitfalls." *Information and Software Technology*, Volume 43, pp. 817-831, 2001.

Previous Genetic Algorithms

Nakayama, Y., and Wilbur, P. J. "The Feasibility of a Genetic-Algorithm-Based Ion Thruster Design Code." 37th Joint Propulsion Conference and Exhibit, AIAA 2001-3786, Salt Lake City, UT, July 2001.

Nakayama, Y., and Wilbur, P. J. "Numerical Simulation of Ion Beam Optics for Many-grid Systems." 37th Joint Propulsion conference and Exhibit, AIAA 2001-3782, Salt Lake City, UT, July 2001.

Nakayama, Y., and Wilbur, P. J. "Numerical Simulation of High Specific Impulse Thruster Optics." IEPC 2001-99, 27th International Electric Propulsion Conference, Pasadena, CA, October 2001.

Carbon Based Ion Optics, CBIO

Wang, J., Cao, Y., Kafafy, R., Martinez, R., and Williams, J. "Ion Impingement Limits of Sub-Scale Ion Optics: Comparison of Simulation and Experiment." 42nd Joint Propulsion Conference and Exhibit. Sacramento, CA, July 2006.

NSTAR

Anderson, J. R., Sengupta, A., Brophy, J. R. "Post-Test Analysis of the Deep Space One Spare Flight Thruster Ion Optics" AIAA-2004-3610. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL. July 2004.

Brophy, J. R. "NASA's Deep Space 1 ion engine" Review of Scientific Instruments, Volume 73, Number 2, American Institute of Physics, February 2002.

Diaz, E. M., Soulas, G. C. "Grid Gap Measurement for an NSTAR Ion Thruster" IEPC-2005-244, 29th International Electric Propulsion Conference. Princeton, NJ, October 2005.

Foster, J. E., Soulas, G. C., and Patterson, M. J. "Plume and Discharge Plasma Measurements of an NSTAR-type Ion Thruster" AIAA-2000-3812. 36th Joint Propulsion Conference and Exhibit. Huntsville, AL, July 2000.

Sengupta, A., Brophy, J. R., Anderson, J. R., Garner, C., de Groh, K., Karniotis, T., and Banks, B. "An Overview of the Results from the 30,3000 Hr Life Test of Deep Space 1 Flight Spare Ion Engine" AIAA-2004-3608. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL. July 2004.

Soulas, G. C. "Improving the Total Impulse Capability of the NSTAR Ion Thruster With Thick-Accelerator-Grid Ion Optics" IEPC-2001-081, 27th International Electric Propulsion Conference, Pasadena, CA, October 2001.

Soulas, G. C. "Performance Evaluation of Titanium Ion Optics for the NASA 30 cm Ion Thruster" IEPC-2001-092, 27th International Electric Propulsion Conference, Pasadena, CA, October 2001.

Life Demonstration Test

Polk, J. E., Anderson, J. R., Brophy, J. R., Rawlin, V. K., Patterson, M. J., Sovey, J., and Hamley, J. "An Overview of the Results from an 8200 Hour Wear Test of the NSTAR Ion Thruster" AIAA-1999-2446. 35th Joint Propulsion Conference, Los Angeles, CA, 1999

Polk, J. E., Anderson, J. R., Brophy, J. R., Rawlin, V. K., Patterson, M. J., and Sovey, J. "In Situ, Time-Resolved Accelerator Grid Erosion Measurements in the NSTAR 8000 Hour Ion Engine Wear Test" IEPC-1997-047, 25th International Electric Propulsion Conference, Cleveland, OH, 1997.

NEXT

Emhoff, J. W., Boyd, I. D., Farnell, C. C., and Williams, J. D. "Perveance and Beamlet Expansion Modeling of the NEXT Ion Engine." IEPC-2005-155, 29th International Electric Propulsion Conference, Princeton, NJ, October 2005.

Kafafy, R. and Wang, J. "Whole Subscale Ion Optics Simulation: Direct Ion Impingement and Electron Backstreaming." AIAA-2005-3691. 41st Joint Propulsion Conference and Exhibit, Tucson, AZ, July 2005.

Malone, S. P. "Investigation of NEXT Ion Optics Erosion Processes Using Computational Modeling." Joint Army Navy NASA Air Force Conference, Monterey, CA, December 2005.

Petillo, J. J., et al., "The MICHELLE Three-Dimensional Electron Gun and Collector Modeling Tool: Theory and Design." IEEE Transactions on Plasma Science, vol. 30, no. 3, pp. 1238-1264, June 2002.

Petillo, J. J., Nelson, E. M., Deford, J. F., Dionne, N., and Levush, B. "Recent Developments to the MICHELLE 2-D/3-D Electron Gun and Collector Modeling Code." IEEE Transactions on Electron Devices, vol. 52, no. 5, pp. 742-748, May 2005.

NEXT 2000 Hour and Long Duration Tests

Frantina, M. M., Arrington, L. A., Soulas, G. C., Hickman, T. A., and Patterson, M. J. "Status of the NEXT Ion Thruster Long Duration Test" AIAA-2005-4065. 41st Joint Propulsion Conference and Exhibit. Tucson, AZ, July 2005.

Kamhawi, H. K., Soulas, G. C., Patterson, M. J., and Frandina, M. M. "NEXT Ion Engine 2000 hour Wear Test Plume and Erosion Results" AIAA-2004-3792. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL, July 2004.

Soulas, G. C., Kamhawi, H. K., Patterson, M. J., Britton, M. A., and Frandina, M. M., "NEXT Ion Engine 2000 Hour Wear Test Results" AIAA-2004-3791. 40th Joint Propulsion Conference and Exhibit. Fort Lauderdale, FL, July 2004.

Soulas, G. C., Domonkos, M. T., Kamhawi, H. K., Patterson, M. J., and Gardner, M. M. "Status of the NEXT Ion Engine Wear Test" AIAA-2003-4863. 39th Joint Propulsion Conference and Exhibit. Huntsville, AL, July 2003.

Ion Optics Simulation

Brown, Ian. *The physics and technology of ion sources*, John Wiley & Sons, New York, 1989.

Chapra, S. C., and Canale, R. P. *Numerical methods for engineers*, 3d ed., McGraw-Hill, New York, 1998.

Piet Hut, Jun Makino, and Steve McMillan. "Building a Better Leapfrog" *The Astrophysical Journal (Letters)*, Volume 443, Pages L93-L96, Feb. 1995.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition, ISBN 0-521-43108-5, Cambridge University Press, Cambridge, 1992.

Distributions

R. Feres and G. Yablonsky. "Knudsen's cosine law and random billiards." *Chemical Engineering Science*, Volume 59, Pages 1541-1556, 2004.

Richard Saucier. "Computer Generation of Statistical Distributions." Army Research Laboratory, ARL-TR-2168, March 2000.

Random Number Generators

Paul Bratley, Bennett Fox, Harald Niederreiter. "Algorithm 738: Programs to Generate Niederreiter's Low-Discrepancy Sequences" ACM Transactions on Mathematical Software, Volume 20, Number 4, pages 494-495, 1994.

Paul Bratley and Bennett Fox. "Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator" ACM Transactions on Mathematical Software, Volume 14, Number 1, pages 88-100, 1988.

Paul Bratley, Bennett Fox, Harald Niederreiter. "Implementation and Tests of Low Discrepancy Sequences" ACM Transactions on Modeling and Computer Simulation, Volume 2, Number 3, pages 195-213, 1992.

Bennett Fox. "Algorithm 647: Implementation and Relative Efficiency of Quasirandom Sequence Generators" ACM Transactions on Mathematical Software, Volume 12, Number 4, pages 362-376, 1986.

M. Matsumoto and T. Nishimura, "Mersenne Twister: A 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator", ACM Trans. on Modeling and Computer Simulation Vol. 8, No. 1, January pp.3-30 (1998)

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, Second Edition, ISBN 0-521-43108-5, Cambridge University Press, Cambridge, 1992.