New engineering design, instrument modeling, and data analysis techniques for spaceborne mass spectrometers

by

Daniel J. Gershman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Space and Planetary Physics)
in The University of Michigan
2012

Doctoral Committee:

Professor Thomas H. Zurbuchen, Chair
Professor George Gloeckler
Professor Timothy A. McKay
Professor Gregory Tarlé
Engineer in Research Lead Bruce P. Block

For Mom

ACKNOWLEDGEMENTS

I would like to thank a lot of people - pretty much everyone I have interacted with over the past 5 years. Perhaps a short list is in order, keeping in mind it is nowhere near complete: Family Gershman, friends from AOSS, EEB, 826Michigan, A2 YMCA Judo, A2Ultimate, SAGE, Ahmed, Mizzou, Jacob, Trantham, Susan, Tami, Shintaro, George, Gina, Alicia, Molly, Cat, Amy, Amanda, and others; SHRG, AOSS, and collaborators, i.e., Jim Raines, Jon Thomas, Len Fisk, Bob Lundgren, Debbie Eddy, Mark Stakhiv, Aleida Young, Micah Weberg, Justin Edmonson, Sue Lepri, Liang Zhao, Martin Rubin, Mike Liemohn, Mark Moldwin, Frank Marsik, Amanda Brecht, Ed Stone, Chuck Smith, Jason Mclain, Thom Orlando, and Jason Gilbert, who gets an extra shoutout for putting together this glorious LaTeX template; SPRL, i.e. Dave Boprie, Mike Long, Rob Gillespie, Chuck Navarre, John Eder, and others; AOSS admin and tech, i.e., Margaret Reid, Rick Baker, Mary Nehls-Frumkin, Tamas Gombosi, Jim Slavin, Bryan White, and others; GSFC, i.e. Paul Mahaffy, Mehdi Benna, Eric Raaen, Mike Barceniak, Heather Franz, John Westberg; and the MESSENGER team, i.e. Sean Solomon, Ralph McNutt Jr., Brian Anderson, Haje Korth, Keri Hallau, to name a few.

I want to especially thank the members of my dissertation committee: Thomas Zurbuchen, Bruce Block, George Gloeckler, Tim McKay, and Greg Tarlé, to whom I sincerely apologize for how long this thesis is.

This work was mostly supported by the NASA Graduate Student Research Program grant NNX09AL50H and the NASA PIDDP grant NNX08AO05G. In addition, the MESSENGER project is supported by the NASA Discovery Program under contracts NAS5-97271 to The Johns Hopkins University Applied Physics Laboratory and NASW-00002 to the Carnegie Institution of Washington. It's pretty awesome that you actually paid me to do this. Thanks.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

LIST OF APPENDICES

Appendix

ABSTRACT

New engineering design, instrument modeling, and data analysis techniques for spaceborne mass spectrometers

by

Daniel J. Gershman

Chair: Thomas H. Zurbuchen

This work describes technological innovations that can be used to improve upon spaceborne mass spectrometers (MS), enabling breakthrough science and the development of the next-generation of sensors. Emphasis is placed on the two classes of MS with the strongest spaceflight heritage: quadrupole mass spectrometers (QMS) and time-of-flight mass spectrometers (TOF-MS). For the QMS, higher order auxiliary excitation techniques are modeled and implemented for the first time for both commercial and spaceflight-like sensors. These techniques, through modest modification of instrument electronics, are shown to significantly improve upon the maximum attainable mass resolution, sensitivity, ion rejection efficiency, and stability of measured mass spectra. For the TOF-MS, a complete analysis of instrument noise sources is conducted, and a mathematical framework for instrument measurements is developed. Such a framework results in an end-to-end forward modeling of instrument noise, dataset signal-to-noise estimation, and noise event removal algorithms. The developed noise processing techniques are applied to the Fast Imaging Plasma Spectrometer (FIPS) instrument on the MErcury Surface, Space ENvironment, GEochemistry and Ranging

(MESSENGER) spacecraft to enable the first ever mapping of the spatial distribution of heavy ions at Mercury, the first in situ measurements of solar wind heavy ion non-thermal properties in the inner heliosphere, as well as the first in situ measurements made inside of Earth's orbit of ionized helium originating from interstellar space.

CHAPTER I

Introduction

Spaceborne mass spectrometers (MS) are vital to the understanding and characterization of some of the most unique and interesting environments in the solar system. MS instruments on a variety of spacecraft, probes, or rovers make in situ measurements, probing particles - both ion and neutral - originating from planets, moons, the Sun, or even interplanetary space. The composition, abundance, and dynamic properties of these particles provide insights into the dynamics of planetary atmospheres, magnetospheres and ionospheres, the origins and evolution of planetary bodies, and even the inner workings of the Sun and its extended atmosphere.

Neutral particles of interest can be gas or dust in interplanetary space (*Gloeckler and Geiss* 1998a), planetary (*Niemann et al.* 1979; *Mahaffy et al.* 2000) or lunar atmospheres (*Hoffman et al.* 1973), samples of rocks and minerals (*Mahaffy et al.* 2012), or atoms from the interstellar medium (*Blum and Fahr* 1970). These particles can range from simple, low mass and high abundance atoms and molecules such as H, H_2, (*Mahaffy et al.* 2000) or CO_2 (*Niemann et al.* 1979), to complex volatile and organic molecules: building blocks of life that are only found in abundances of a few parts per billion (*Belton* 2003; *Lunine* 2001). Most neutrals, even if part of a hot atmosphere (*Kasprzak et al.* 1993), have energies of only a few eV, though some can be accelerated through charge exchange reactions with ions to have energies in the

keV-MeV range (*Gruntman* 1997).

Ions of all different types are measured in the space environment, with a host of masses, charge states, energies, and abundances. Ions from the Sun are accelerated in the corona, nominally forming the solar wind (*Parker* 1965), a ∼1keV plasma with high charge state ions ranging from hydrogen to iron (*von Steiger et al.* 2000). Interstellar neutrals (He, Ne, H (*Gloeckler and Geiss* 1998a)) can become photoionized and are convected out with the solar wind as singly charged pickup ions (*Kallenbach et al.* 2000). Ionospheric (*Waite et al.* 2004) and magnetospheric ions (*Zurbuchen et al.* 2008, 2011) are typically atoms or molecules from a planetary atmosphere that become singly charged, and depending on their energetization processes, can range from a few eV to several keV. Finally, high energy (MeV - GeV) cosmic rays, mostly H and He (*Simpson* 1983), enter the solar system from interstellar space, and have fluxes into the heliosphere that are modulated by changes in the solar cycle (*Forbush* 1954; *Burlaga et al.* 1993).

With such a wide variety of particles to measure, it would be challenging at best to design a single MS to effectively analyze them all. There are a number of different types of MS instruments, each best suited for a particular type of measurement. In particular, two important instrument types that will be discussed in detail in this work are the quadrupole mass spectrometer (QMS), which has been widely used to study neutral particles in planetary atmospheres, and the time-of-flight (TOF) mass spectrometer, commonly used to study ion composition in both planetary and interplanetary environments. Missions that have employed either QMS or TOF sensors span practically every planetary body explored to date.

1.1 Quadrupole Mass Spectrometers

Until the mid 1900s, mass spectrometers relied on magnetic deflection of ionized particles, using the Lorentz force to separate them by their mass-per-charge (m/q)

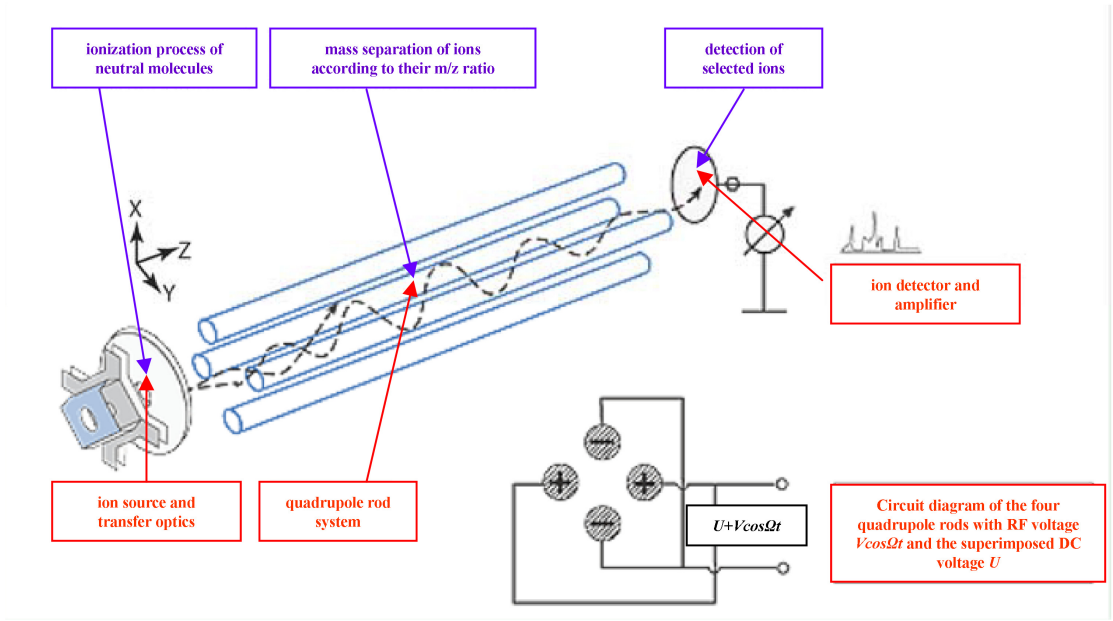

Figure 1.1: Overview of quadrupole mass spectrometer sensor operation. Ionized neutrals are filtered through a set of conducting rods. Only ions with a particular m/q will successfully traverse the rods, reaching a detector. Image adapted from: *in-process.com*

ratios. Such sensors required large, cumbersome magnets to enable mass separation of species within a reasonably sized instrument. These analyzers were electro- and/or magneto-static, using constant electric and magnetic fields to guide particle trajectories. However, in 1953, Paul and Steinwedel first demonstrated mass separation using an electrodynamic quadrupole field (*Paul and Steinwedel* 1953), an accomplishment that would later lead to a Nobel Prize in Physics for Paul in 1989 (*Dawson* 1997). As illustrated in Figure 1.1, a sample of neutral gas is ionized and accelerated through an ion optics system forming a beam of low energy ions. The beam traverses four conducting rods of hyperbolic or circular cross section arranged so that they create a quadrupole electric field with applied potentials. The applied potential has both a DC and an RF component. With proper selection of amplitude and frequencies, ions with a particular m/q ratio will successfully traverse the rod set, reaching a particle detector, with others striking the rods.

Such an instrument boasted faster scanning speed and lighter weight when compared with previous instruments, with no need for the inclusion of a large, bulky magnet (*Dawson* (1997)). In 1958, Paul and his team were able to demonstrate mass resolutions on the order of 16,000 using 6m long rods (*Paul et al.* 1958). This technique was scalable, and therefore could be miniaturized and used for atmospheric and space applications. By the 1960s-1970s, QMS were being developed for sounding rocket flights and satellite payloads, such as the OGO-6 neutral atmosphere composition experiment (*Carignan and Pinkus* 1968) and the Atmospheric Explorer and Dynamics Explorer missions, intended to investigate the thermosphere of Earth (*Pelz et al.* 1973).

The first QMS flight instrument for a planetary science application was the Orbiter Neutral Gas Spectrometer (ONMS) (*Niemann et al.* 1980) on the Pioneer Venus Orbiter that launched in 1979. ONMS made the first in situ measurements in the thermosphere of Venus, confirming the presence of CO_2, CO, N_2, O, and He that was suspected from remote sensing observations (*Niemann et al.* 1979). Since that time, quadrupoles have become the instrument of choice for neutral gas measurements in space, as illustrated in Figure 1.2. The Galileo Probe Mass Spectrometer (GPMS) (*Niemann et al.* 1992), launched in 1989 measured the in situ D/H and ^{3}He/^{4}He ratios in the atmosphere of Jupiter (*Mahaffy et al.* 1998). Two QMS instruments flew as part of the Cassini-Huygens mission, launched in 1997, with the Ion Neutral Mass Spectrometer (INMS) (*Waite et al.* 2004) on the Cassini orbiter measuring space plasmas around Saturn, and the Gas Chromatograph Mass Spectrometer (GCMS) (*Niemann et al.* 1997) on the Huygens probe, measuring the hydrocarbons on Titan (*Niemann et al.* 2005). The Neutral Mass Spectrometer (NMS) (*Niemann et al.* 1998) on the Planet-B spacecraft, (renamed 'Nozomi'), and the (NGIMS) (*Mahaffy* 2002) on the CONTOUR spacecraft were intended to measure neutral particles around Mars and several comets, respectively, but problems with the spacecraft prevented

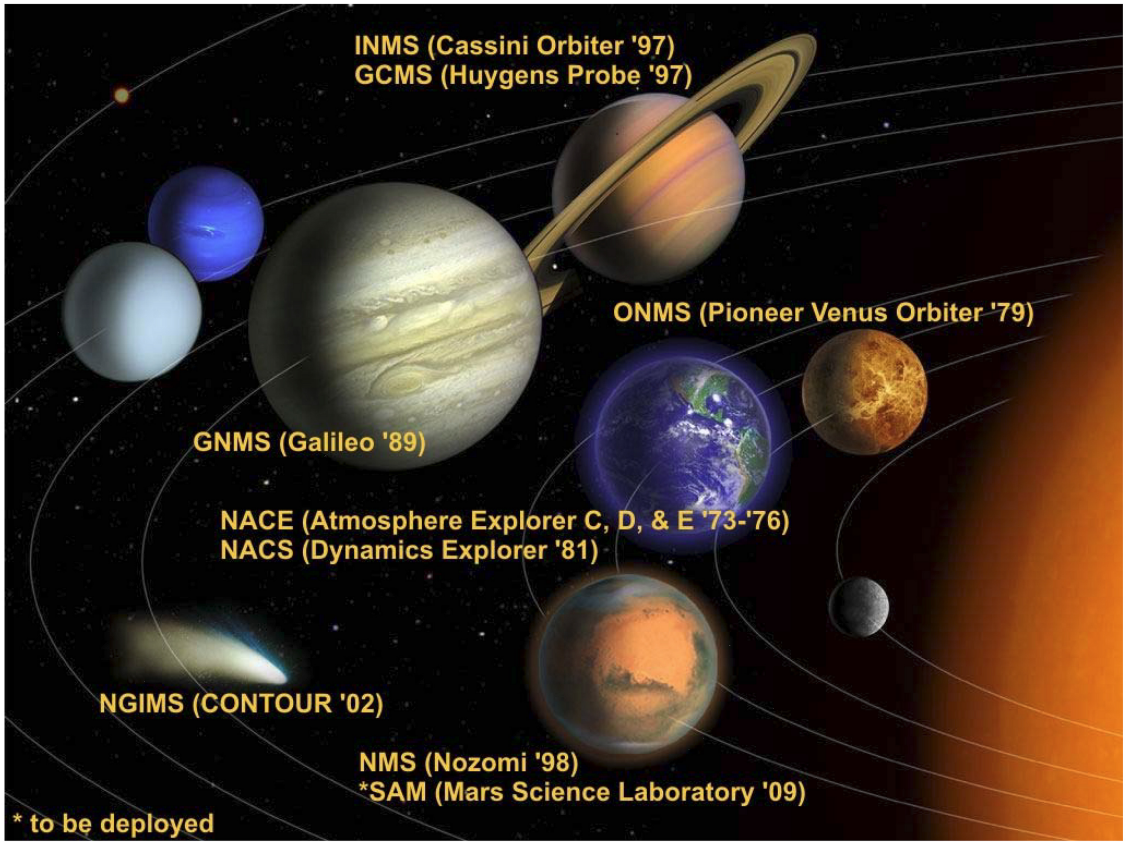

Figure 1.2: Illustration of spaceborne QMS instruments and their planetary body of interest. QMS sensors have been deployed to almost every explored planetary body in the solar system. MSL was eventually launched on November 26, 2011. Image credit: Martin Rubin

the successful collection of instrument data (*Kerr* 2003; *Harvey et al.* 2010). Most recently, the Sample Analysis at Mars instrument (SAM) (*Mahaffy et al.* 2012) on the Mars Science Laboratory (MSL) was launched on November 26, 2011. It is an example of the latest in spaceborne quadrupole mass spectrometry technology.

1.2 Time-of-flight Mass Spectrometers

For space applications, TOF-MS are typically used to measure higher energy ions, where QMS cannot easily be applied. However, the technology was original intended for to measure newly ionized neutrals. The idea of a TOF-MS was first proposed by

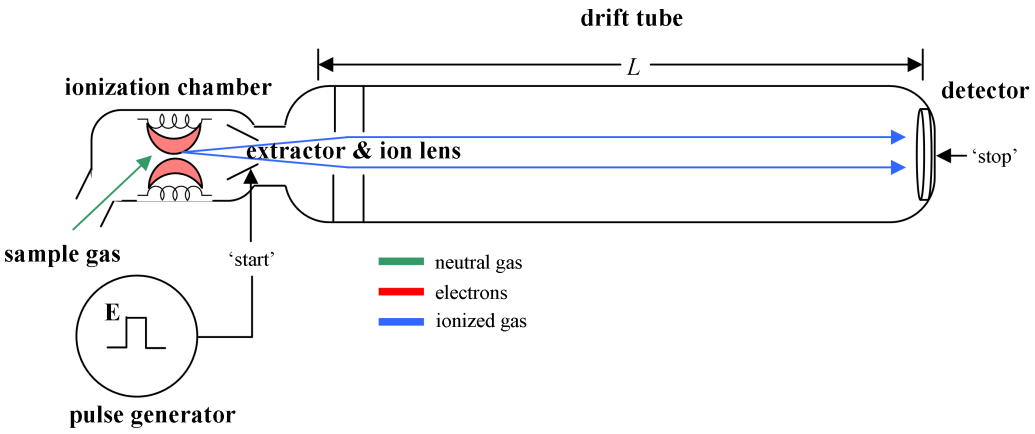

Figure 1.3: Schematic of the first TOF-MS, the 'velocitron,' following *Cameron and Eggers* (1948). A sample gas is ionized through electron-impact ionization and pulse through a a drift tube with length L. An extraction pulse accelerates ions to energy E and serves as the 'start' signal. Extracted ions strike the detector at the end of the drift tube, creating a 'stop' signal, enabling a TOF measurement.

W. E. Stephens in a meeting of the American Physical Society in 1946 (*Campana* 1987). The first published work was by *Cameron and Eggers* (1948) where they used a device called a 'velocitron' to measure ion species with m/q up to 200 amu/e. As shown in Figure 1.3, in a basic TOF-MS, neutrals are ionized, typically via electron-impact collisions. An extraction voltage on the ion source is pulsed, accelerating a beam of ions down a drift tube. The timing of this pulse serves as a 'start' signal. After traversing the length of the drift tube, the particle hits a detector, generating a 'stop' signal. With a known ion energy-per-charge ratio (E/q) from the source, drift tube length, and the ion TOF calculated between start and stop signals. The measured ion m/q can be determined.

The above instrument cannot easily measure already existing high energy ions, however, as high voltages cannot be pulsed fast enough to create a sufficient start-signal. This problem was originally addressed by those studying ions produced from nuclear reactors (*Schneider* (1970); *Pfeffer et al.* (1973)). In these cases, a thin

carbon foil was placed at the entrance of the TOF chamber. These foils were several atoms thick (10-20 $\mu g/cm^2$), such that ions with sufficient energies (keV - MeV) could penetrate them. As the ion passed through, it became neutralized, and also liberated secondary electrons. These electrons could be accelerated using electric fields and, due to their small mass when compared to even the lightest of ions, move extremely fast, suitable for their incidence on a detector to be considered the 'start', and the slower moving ion incident on a detector considered the 'stop'. The use of electrons as start signals in these devices was dubbed a 'zero-time detector'.

The energies required to penetrate these foils (MeV), however, were too large for most space plasmas, most notably the solar wind, even with an acceleration region inserted before the TOF chamber entrance. Furthermore, the large detectors required for particle measurements made their incorporation into a space instrument extremely challenging. *Gloeckler and Hsieh* (1979) solved this issue by building an apparatus that used a much thinner foil ($\sim 2.5 \mu g/cm^2$) and used smaller microchannel plates as both the start and stop detectors. This work paved the way for the first space-borne TOF instruments, the Suprathermal Energy Ionic Charge Analyzer (SULEICA) (*Möbius et al.* 1985b) and the CHarge-Energy-Mass Spectrometer (CHEMS) (*Gloeckler et al.* 1985) on the AMPTE/IRM satellite. CHEMS, as shown in Figure 1.4, had an electrostatic analyzer (ESA) to filter ions by their E/q using shaped electrodes, a post-acceleration region to accelerate ions to sufficient energies to penetrate a thin carbon foil, and a TOF chamber with start and stop detectors. The energy of the incoming ion (E) was also measured using a solid state detector. From these measurements, the instrument can derive the energy, mass, and charge state of incident ions.

Since the successful operation of CHEMS, most ion measurements in space have been done by TOF-MS such as the Solar Wind Ion Composition Spectrometers (SWICS) on Ulysses, ACE, and Wind (*Gloeckler et al.* 1992, 1995b, 1998), the

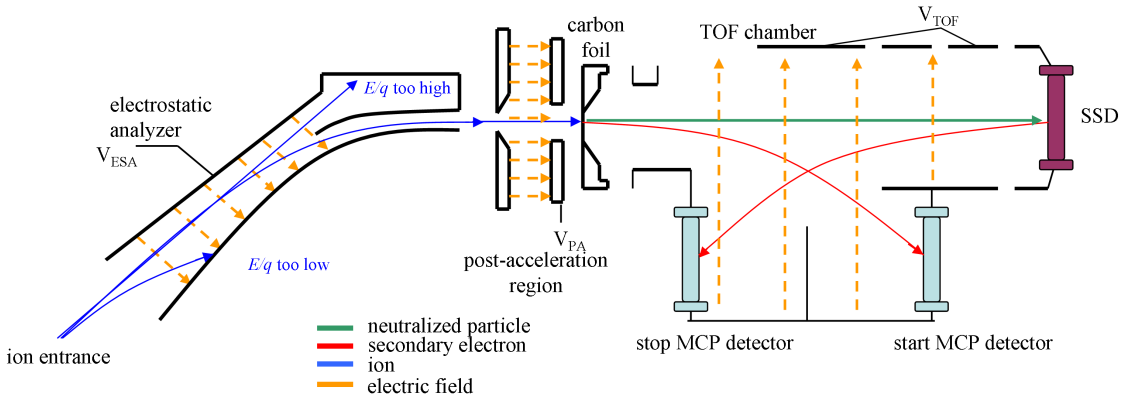

Figure 1.4: Schematic of spaceborne TOF-MS, following CHEMS and SWICS. Ions are filtered by their E/q by an electrostatic analyzer with applied voltage V_{ESA}. Filtered ions are accelerated by a post-acceleration voltage (V_{PA}) such that they have sufficient energy to penetrate a thin carbon foil. The ion becomes neutralized and hits an SSD that records its energy) (E). Secondary electrons from the carbon foil and SSD serve as 'start' and 'stop' signals, respectively, providing a measurement of ion TOF.

PLASma and SupraThermal Ion Composition (PLASTIC) experiment on STEREO (*Galvin et al.* 2008), the CAsinni Plasma Spectrometer (CAPS) (*Young et al.* 2004), and the Fast Imaging Plasma Spectrometer (FIPS) on MESSENGER. Each of these sensors has provided groundbreaking measurements of space plasmas.

1.3 Next-generation sensors

All instrument measurements will always have some finite signal to noise ratio that ultimately will determine the usability of a dataset. A neutral gas spectrometer such as a QMS typically only measures in one dimension: mass-per-charge. In a mass-per-charge spectrum from a QMS, it can be relatively simple by eye to determine the detector background rates and to distinguish between signal and noise, with the dominant source of noise typically being a background signal on the particle detector creating an effective particle flux detection threshold. Improving the quality of QMS measured datasets therefore becomes an issue of enhancing the basic

characteristics of the instrument: sensitivity, mass resolution, and stability. Without a fundamental shift in instrument topology, such improvements can only be accomplished through modifications to instrument hardware such as the ion source, rod system, ion detectors, or driving electronics.

Ion mass spectrometers, on the other hand, have a larger range of kinetic properties to recover in addition to mass-per-charge, namely the entire velocity distribution of a measured plasma, as well as ion charge states. In fact, spaceborne TOF-MS technology is currently extremely versatile, with flight instruments capable of making high resolution and/or three-dimensional velocity measurements over wide energy ranges. However, the goal of space missions is always to do more with less, where it is desired to combine these various features of TOF-MS into a smaller, lighter, and more power efficient package. However, these consolidations can result in unique and unexpected noise sources. In a multidimensional dataset such as one produced from a TOF-MS, it can be difficult to distinguish between signal and noise given an arbitrary accumulation of data. Even a TOF-MS with perfect angular, energy, and mass resolution can have a low measurement signal to noise ratio without it being immediately apparent. The difficulties in instrument operation therefore transition from the taking of measurements to the distinction between signal and noise. As the capabilities and complexity of future TOF-MS grow with the development of the next-generation of flight sensors, so too must the data processing and analysis tools available for instrument scientists and engineers.

Engineering design for space is also itself challenging, with special consideration required for the increased thermal, mechanical, and radiation stresses produced from the launch process and space environment, and the restrictions on instrument mass and power. The high cost of launch and operation necessitates a certain added caution and risk assessment of new sensors, leading to the concept of 'heritage', where technology that has proven to be successful in space is more likely to fly again. Her-

itage designs have been successful in the past, but can quickly become outdated and insufficient to make the latest desired measurements, creating a double-edged sword: proven technology with reduced performance, or newer technology at higher risk?

This work addresses these issues with two approaches to improving upon space instrumentation to enable breakthrough science: (1) add technological innovation to heritage designs to obtain improved performance or (2) reduce the risk of a new design with little heritage by developing innovative data analysis and recovery tools in the event that unforeseen issues may arise. Chapters 2,3, and 4 focus on the former, where the performance of existing QMS sensors is improved through modest modification of sensor electronics. For the latter, Chapters 5, 6, and 7 introduces simulation and data processing techniques that enable a complete forward modeling, and in some cases removal, of noise in TOF-MS instruments. These techniques are applied to the FIPS sensor on the MESSENGER spacecraft, a novel, low heritage, TOF-MS whose design enables three-dimensional ion measurements in a compact, low-power package. Although its small form factor led to unexpected noise sources in-flight, Chapters 8,9, and 10 describe heliospheric and magnetospheric science results enabled after application of the presented techniques, measurements that were not previously possible with a TOF-MS of comparable scale. The chapters in this dissertation are each derived from a published or soon to be submitted journal article, with appropriate references indicated. For those chapters on which I am not first author, my contribution is summarized before the main article text.

CHAPTER II

Higher order parametric excitation modes for spaceborne quadrupole mass spectrometers

This chapter is taken from: Gershman, D. J., B. P. Block, M. Rubin, M. Benna, P. R. Mahaffy, and T. H. Zurbuchen (2011), Higher order parametric excitation modes for spaceborne quadrupole mass spectrometers, *Rev. Sci. Instrum., 82*, 125109.

Abstract

This paper describes a technique to significantly improve upon the mass peak shape and mass resolution of spaceborne quadrupole mass spectrometers (QMSs) through higher order auxiliary excitation of the quadrupole field. Using a novel multiresonant tank circuit, additional frequency components can be used to drive modulating voltages on the quadrupole rods in a practical manner, suitable for both improved commercial applications and spaceflight instruments. Auxiliary excitation at frequencies near twice that of the fundamental quadrupole RF frequency provides the advantages of previously studied parametric excitation techniques, but with the added benefit of increased sensed excitation amplitude dynamic range and the ability to operate voltage scan lines through the center of upper stability islands. Using a field programmable gate array, the amplitudes and frequencies of all QMS signals are digitally generated and managed, providing a robust and stable voltage control

system. These techniques are experimentally verified through an interface with a commercial Pfeiffer QMG422 quadrupole rod system. When operating through the center of a stability island formed from higher order auxiliary excitation, approximately 50% and 400% improvements in 1% mass resolution and peak stability were measured, respectively, when compared with traditional QMS operation. Although tested with a circular rod system, the presented techniques have the potential to improve the performance of both circular and hyperbolic rod geometry QMS sensors.

2.1 Introduction

This paper describes a technique to significantly improve the performance of the spaceborne quadrupole mass spectrometer (QMS) through higher order auxiliary excitation of the quadrupole field. Through the invention of the quadrupole mass spectrometer, *Paul and Steinwedel* (1953) unlocked the means to obtain unprecedented measurements of neutral gas composition and isotopic ratios. Since then, QMSs have revolutionized chemical analysis, by simultaneously reducing instrument mass, size, and cost, and increasing instrument scan speed when compared with traditional magnetic sector mass spectrometers (*Dawson* 1997). Such advancements positioned QMS technology as the standard choice for low-energy spaceborne mass spectrometry applications. In fact, QMSs have flown on almost every mission to a planetary body in the solar system (*Waite et al.* 2004; *Niemann et al.* 1992, 1980; *Lunine* 2001). They have provided a set of ion and neutral composition data that has led to significant discoveries about planetary physics, chemistry, and composition, and have transformed the understanding of the origin of planetary atmospheres. (See, for example, *Lunine* (2001), and references therein.)

The specifications of an ideal measurement with a spaceborne QMS are three-

fold. (1) The ability to distinguish between adjacent ion masses is set by the mass resolution, which must be large enough to achieve readable mass spectra. (2) Repeatable and predictable measured mass peak shapes are vital for the determination of precise isotopic ratios and relative abundance of different species. (3) Sensitivity, i.e. the successful transmission of desired particles through the instrument, must be high in order to measure species whose relative abundances span multiple orders of magnitude. Obtaining sufficient performance in these areas is especially challenging for space applications due to the power and mass constraints imposed on instrument design. However, improvements in all aspects of QMS operation are vital for the continued use of these highly successful instruments for the continued advancement of space and planetary science.

Over the last several years, the scientific objectives of solar system exploration have evolved, and with them, the requirements on space instrumentation. The emphasis of many previous missions has been on the measurement of low mass atoms and molecules (less than 100 amu) such NH_3 and CH_4, and isotopic ratios, such as D/H or $^3He/^4He$, which provide insight into the origin of planetary bodies (*Geiss and Gloeckler* 2003; *Mahaffy et al.* 1998, 2000). After having spent decades concentrating on solar system formation, the planetary community has shifted much of its focus to a new puzzle: what is the origin of life? The answer to this question requires the study of complex organics and volatiles in primitive bodies and planetary environments (*Belton* 2003). These molecules are both high in mass (100-1000 amu) and low in abundance, making them difficult to detect with current spaceborne QMS technology. In addition, noble gases and isotopic ratios in simple molecules are still of fundamental importance and have not yet been measured well in comets or on Venus. New QMS sensors must maintain the ability to make these measurements and at the same time, be scalable to meet new science requirements. An order of magnitude improvement in both mass resolution (targeting $M/\Delta M \sim 10^3$) and abundance sensitivity (targeting

the ppb range) will be required to detect and study these trace molecules, which may hold the key to understanding the potential for life in the solar system.

Similar QMS performance is desired in the commercial realm, which represents a much higher number of deployed sensors. Applications of this technology range from medical and biological analysis to forensic science (*March and Todd* 2005). These commercial spectrometers are typically lower-cost than their space-flight counterparts, and rely on the use of less expensive manufacturing and development processes that may lead to reductions in attainable instrument performance (*Dawson* 1997). Any enhancement in the quality of QMS measurements has direct impact on both spaceborne and commercial QMS instruments, especially if this enhancement is achievable with relatively minor modifications. In fact, the improvements discussed in this paper are expected to be more profound for such commercial systems.

Parametric excitation of the quadrupole field is technique that was originally implemented to mitigate the low mass side peak tail (*Devant et al.* 1989) that can occur as a consequence of using cicular geometry, or from ions not spending sufficient time traversing the rods in order to be ejected. However, it has since been shown to provide improvement in maximum attainable mass resolution, improved quality of mass peak shapes, and, for higher resolution modes, increased instrument sensitivity (*Konenkov et al.* 2001, 2005; *Zhao et al.* 2009; *Sudakov et al.* 2000; *Zhao et al.* 2010). In this approach, a second periodic signal drives the QMS rods at one of an ion trajectory's normal modes in order to hasten its ejection. However, parametric excitation techniques have previously (*Konenkov et al.* 2001; *Zhao et al.* 2009) relied on methods that were rather inefficient in driving the quadrupole rods with auxiliary signals, limiting the maximum excitation amplitude, and consequently making them impractical for use in flight QMS sensors and unsuitable for many commercial applications. This paper describes a highly efficient technique that enables parametric excitation for spaceborne applications. Section 2.2 reviews parametric excitation techniques and

discusses a new operating mode that uses an auxiliary frequency nearly twice that of the fundamental radio frequency (RF) in order to increase the dynamic range of the sensed auxiliary signal amplitude and to provide increased peak stability through operation of voltage scan lines through the center of upper stability islands. Section 2.3 introduces a multiresonant tank circuit that enables efficient driving of the quadrupole rods with these larger excitation amplitudes, as well as a field programmable gate array (FPGA)-based voltage control system that is used to manage and drive all QMS frequency components. Finally, Section 2.4 presents experimental verification of the proposed techniques and describes their implications for future space missions.

2.2 Parametric Excitation

A typical spaceborne QMS sensor is composed of four parallel conducting rods of hyperbolic cross section spaced by distance r_o from the center. DC (amplitude U) and RF (amplitude V and frequency Ω) voltages are applied to opposing pairs of rods with opposite polarities. Low-energy ions, which may be recently ionized, are injected through the rods and deflected by the electric field such that only ions with a particular mass-to-charge (m/Z) ratio have stable trajectories, passing through the volume between the rods. All other ions are ejected out of that volume and are lost. Less expensive circular rods are sometimes used in commercial instruments that can produce a slightly distorted quadrupole electric field, resulting in reduced performance (*Dawson* 1997). Although proper sizing and alignment of circular rods can help to minimize these distortions, (*Douglas and Konenkov* 2002) all previous spaceflight instruments have utilized hyperbolic geometry to ensure the best possible instrument performance. Figure 2.1 illustrates the application of combined DC and RF potentials to opposite rod pairs.

A similarity transformation can be applied to the equations of ion motion in an ideal quadrupole field to produce a set of Mathieu differential equations (*Dawson*

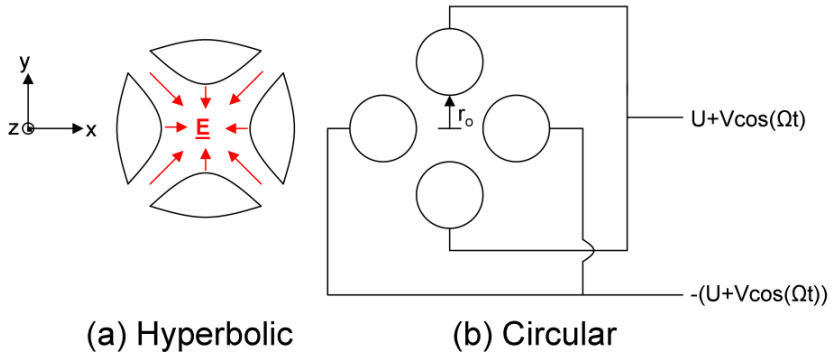

Figure 2.1: Geometry of quadrupole rods for (a) hyperbolic and (b) circular rod sensors. The potential applied to opposing pairs of quadrupole rods, $\pm(U + V cos(\Omega t))$ is shown for the circular rods, though identical connections exist for hyperbolic geometry. Ions are injected into the rods along the $\hat{z}$ direction, and their trajectories are modified by the quadrupole electric field, $\mathbf{E}$, in the $\hat{x}$ and $\hat{y}$ directions. The electric field for circular rod geometry mimicks that formed from hyperbolic geometry, but becomes increasingly distorted with increasing distance from the center of the rod system.

1997) that are written in terms of dimensionless voltages a and q and dimensionless time $\xi = \frac{\Omega t}{2}$, with geometry (x,y,r_o) defined in Figure 2.1:

$$\frac{d^2x}{d\xi^2} + (a - 2qcos(2\xi))x = 0 \text{ and} \tag{2.1}$$

$$\frac{d^2y}{d\xi^2} - (a - 2qcos(2\xi))y = 0, \tag{2.2}$$

where

$$a = \frac{8ZeU}{mr_o^2\Omega^2} \text{ and } q = \frac{4ZeV}{mr_o^2\Omega^2}. \tag{2.3}$$

Here, e is the magnitude of the charge of an electron and Z is the ion charge number. m is the ion mass in units of *kilograms*, x,y, and r_o are in units of *meters*, U and V are in units of *Volts*, and Ω is in units of *rad/s*. These equations have several classes of solutions that depend on the combinations of scaled RF and DC

voltage amplitudes q and a, respectively. Stable ion trajectories are periodic and bounded, allowing ions to travel through the rods unhindered. While most unstable ion trajectories will result in ions hitting the rods, never reaching the detector, some may not be ejected by the time they traverse the rod set, resulting in imperfect mass peak shapes. For constant rod geometry and operating frequencies, q and a vary only with ion m/Z, creating a set of voltages that only allow a particular range of m/Z ratios to reach the detector, i.e., in normal quadrupole operation, a controlled scan line of voltages that passes through a stable region yields a series of mass peaks with increasing m/Z. The slope of this scan line (λ) is frequently defined as

$$\lambda = \frac{U}{V} = \frac{a}{2q}. \tag{2.4}$$

The resolution (R_{w}) and transmission (T_{w}) of these mass peaks must be defined in terms of a reference value (w) that define the peak boundaries. For example, for $w = 0.10$, $R_{0.1}$ is defined as M/ΔM where M is the the mass of interest and ΔM defined as the width of the mass peak with transmissions greater than 10% of the maximum transmission value. $T_{0.10}$ is defined as the average transmission of the mass peak within that same ΔM range. Results of previous work have been in terms of $w = 0.50$ (*Zhao et al.* 2009, 2010) and $w = 0.10$ (*Konenkov et al.* 2005, 2001). However, for cases where the dynamic range of measured mass peaks is of vital importance, such as for space applications, $w \leq 0.01$ are of interest.

The region of stable ion trajectories is ordered by lines of constant β_{x} and β_{y}, parameters that correspond to the frequency components of ion trajectories in the x and y coordinates, respectively. For typical stable trajectories (*Dawson* 1997), β_{x} and β_{y} are both between 0 and 1 as shown in Figure 2.2a. Driving the quadrupole rods at one of these frequencies with voltage amplitude V_{ex} excites parametric resonances.

These resonances create bands of instability in the stable region along lines of constant β as shown in Figure 2.2b. The original stability region breaks into smaller

sections, including an upper stability island that was originally discussed in a patent by *Miseki* (1993). Since that time, the upper stability island has been the source of more in depth analytical and experimental study (*Konenkov et al.* 2001; *Zhao et al.* 2009; *Konenkov et al.* 2005; *Zhao et al.* 2010). This splitting is analogous to the normal modes of a dynamical system (*Alfred et al.* 1993), where an ion trajectory is modeled as a superposition of harmonic oscillators with secular frequencies related to β_{x} and β_{y}. Excitation at a given frequency can create a resonant ion trajectory, just as driving a system at one of its normal modes can result in large amplitude oscillations. The amplitude of the ion trajectory will quickly grow such that the ion strikes the QMS rods rather than reaching a detector, even if it is of the appropriate m/z in nominal QMS operation. It has been shown through modeling that operating a QMS in the upper stability island can result in improved peak shapes and mass resolution even for ideal hyperbolic rod sensors (*Konenkov et al.* 2005). This technique can also be used to correct for nonlinear resonances that arise naturally as a consequence of using circular rod sensors or from imperfections in the rod alignment or manufacturing processes (*Zhao et al.* 2009, 2010).

Sudakov et al. (2000) have derived a general resonance condition that relates the parameter β and quadrupole operating frequencies. This condition can be written in terms of an excitation frequency parameter, f , the auxiliary frequency, ω_{aux}, and the fundamental RF frequency, Ω_{RF}, such that

$$f \equiv \frac{\omega_{\mathrm{aux}}}{\Omega_{\mathrm{RF}}}, \text{ and } Kf = |n + \beta|. \tag{2.5}$$

Here, K is the order of resonance i.e., $1, 2, 3, ...$, and n is the resonant mode number i.e., $\pm 0, \pm 1, \pm 2, ...$

This auxiliary signal will modify the ion equations of motion with:

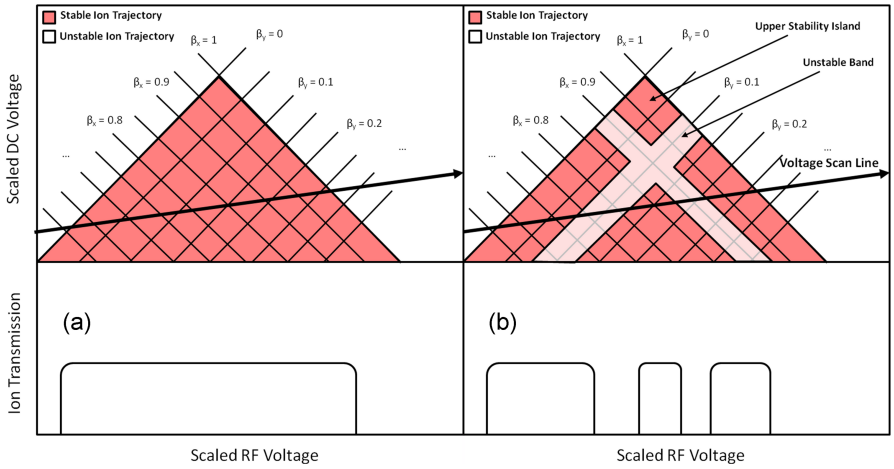

Figure 2.2: (a) **Nominal QMS operation:** Combinations of RF and DC voltage amplitudes result in either stable or unstable ion trajectories. A controlled scan line of voltages that passes through a stable region yields a mass peak. The stability region is ordered by lines of constant β_x and β_y that are parameters that correspond to frequency components of ion trajectories. (b) **Parametric excitation mode:** The insertion of additional frequency components excites parametric resonances along iso-β lines, creating bands of instability and an upper stability island.

$$\frac{d^2x}{d\xi^2} + (a - 2qcos(2\xi) - 2q'cos(2\xi f))x = 0 \text{ and} \tag{2.6}$$

$$\frac{d^2y}{d\xi^2} - (a - 2qcos(2\xi) - 2q'cos(2\xi f))y = 0, \tag{2.7}$$

where

$$q' \equiv q\frac{V_{ex}}{V}. \tag{2.8}$$

As an example, consider a first-order resonance, $K = 1$. A single auxiliary frequency, $f = 0.9$, can yield resonances at both $\beta = 0.1$ and $\beta = 0.9$ for low mode numbers $|n| = 0, 1$. As shown in Figure 2.2a, at the tip of the stability region, β_x approaches 1 and β_y approaches 0. Therefore, for parametric excitation of $f = 0.9$, the resonances along $\beta_x = 0.9$ and $\beta_y = 0.1$ will form an upper stability island, as

shown in Figure 2.2b.

2.2.1 Stability Island Operation

There are multiple ways to adjust the mass resolution when operating in an upper stability island. For a fixed island size, i.e., fixed frequency f and excitation amplitude q', the slope of the scan line λ can be adjusted such that it passes through the upper or lower tips of the island. Large ranges of mass resolution can be achieved through formation of relatively large stability islands, requiring only small excitation amplitudes to achieve the desired island splitting. This method has been the preferred mode of operation of previous work (*Zhao et al.* 2009, 2010).

Alternatively, the size of the stability island itself can be adjusted, i.e. variation of either frequency and/or excitation amplitude. The slope of the scan line is adjusted such that it passes through the center of the island, forming a mass peak having the maximum width of the island. Such operation has two advantages over operation at the island tips. First, for high resolution modes, transmission is maximized (*Konenkov et al.* 2005; *Zhao et al.* 2010). Second, mass peak variability is minimized. Consider small perturbations of the slope of the scan line due to electronics or control system errors. Near the tip of a stable region, both the transmission and peak width change significantly with small changes in slope, creating potentially wildly varying peak shapes. The center of the stability island, however, is much less sensitive to errors in scan line slopes, as it is more symmetric than the island tips with respect to both changes in peak width and transmission. Operation through the center of the stability island has not previously been applied to laboratory work due to the prohibitively large excitation amplitudes required to form sufficiently small islands.

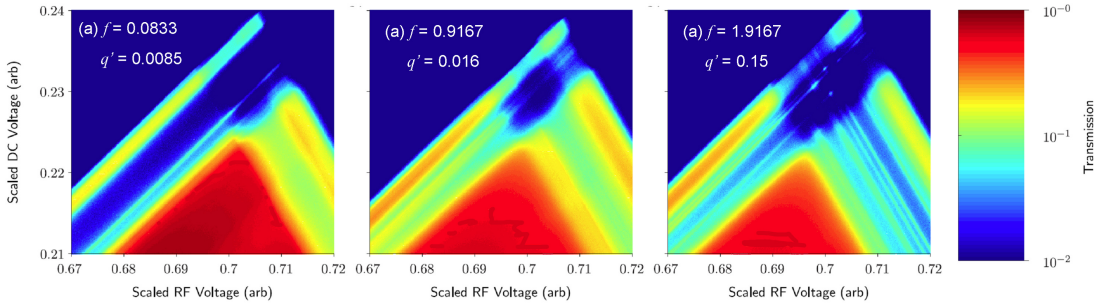

Figure 2.3: Numerically simulated quadrupole stability diagrams for (a) $f = 0.08333, q' = 0.0085$, (b) $f = 0.9167, q' = 0.016$, and (c) $f = 1.9167, q' = 0.15$. Ions spent 40 RF cycles traversing the quadrupole rods with an aperture of $0.232r_o$. While the stability regions for each excitation mode exhibits slightly different features, the location and size of the upper stability island are similar for each.

2.2.2 Higher Order Resonances

Consider resonances defined by equation (2.5) of a given order, i.e., fixed K. As the excitation frequency parameter f increases, the absolute value of n must increase in order to have a resonance at a particular $0 < \beta < 1$, i.e., higher frequency excitations can only cause resonances of higher order modes. The energy available for each mode drops off quickly (*Alfred et al.* 1993) with n, so significantly more excitation energy must be added before a resonance of sufficient strength creates adequate island formation at higher frequencies. As an example, Figure 2.3 illustrates the required excitation to obtain similar island splitting for operating frequencies corresponding to $f = 0.0833$, $f = 0.9167$, and $f = 1.9167$, using a numerical simulation of equations (2.6). Trajectories for 40,000 total random ions were calculated at each of 500x500 (q,a) points using a 4th order Runge-Kutta numerical scheme. Ions were axially initialized within an aperture of $0.232r_o$ ($r_o = 6.7mm$) with one of 20 different initial phases. Simulated ions spent approximately 40 RF cycles in the rods to match the experimental setup used later in this paper. Fringing-field and ion energy spread effects were not taken into account.

From Figure 2.3, is is clear that for higher auxiliary frequencies f, higher excitation amplitudes are required. Adding more excitation energy to the system has negative connotations, as it implies increased power consumption of the QMS. However, the efficiency at which the energy is inserted into the system must be considered, as will be discussed. Futhermore, for increased q', the dynamic range of the sensed auxiliary signal amplitude is higher. The sensed amplitude is the potential difference between the quadrupole rods created by the auxiliary signal. Electronics that measure this signal as part of a closed-loop feedback system require less precision to sense larger amplitudes. Therefore, from both control system and electronics development perspectives, operation near $f = 2$ is favored.

While auxiliary excitation techniques near $f = 0$ and $f = 1$ have been experimentally verified (*Zhao et al.* 2009, 2010), a practical space-based application has not yet been achieved due to power and other technical implications of the proposed solutions. As mentioned previously, these studies operate scan lines through the upper and lower tips of stability islands. This work, however, chooses to scan voltages such that they pass through the center of formed stability islands, simultaneously maximizing both ion transmission and mass peak shape stability, vital metrics for a spaceborne sensor, where scan integration time is held at a premium. While the island splitting from Figure 2.3 is similar for the $f \approx 0, 1$, and $f \approx 2$ operating frequencies, the subtle differences in island shapes and locations due to the relative strengths of resonant modes can lead to differences in instrument performance. In comparing $f \approx 0$ and $f \approx 1$ operating points, *Zhao et al.* (2010) found that $f \approx 0$ yields the highest attainable mass resolution for a system, though $f \approx 1$ gave higher average transmission and still achieved relatively high resolutions. Given the limitations of electronics used for this work, we are unable to experimentally compare $f \approx 0, 1$, and $f \approx 2$ at high resolutions. However, $f \approx 2$ operation is expected to produce similar performance to the $f \approx 1$ modes tested due to their similar sets of resonant β_x and

β_y values from equation (2.5).

High resolution measurements will require the formation of smaller islands, i.e. increased excitation signal amplitude. This problem is seemingly exacerbated through the use of higher order resonant excitation, which requires significantly stronger driving amplitudes. However, as will be shown, a new multiresonant tank circuits permits the use of such amplitudes without significant increases in required driving current. The nature of this circuit necessitates operation near $f \approx 2$, which maximizes the sensed excitation amplitude and enables these technologies to be applied to the next generation of QMS space instruments.

2.3 Implementation

As mentioned previously, existing quadrupoles have already been modified to take advantage of parametric excitation near $f = 0$ and $f = 1$. These modifications have involved the direct insertion of an auxiliary signal into the quadrupole tank circuit independent of the fundamental RF waveform. The relative phases and amplitudes of each frequency component were not controlled in a precise way and are therefore currently unsuitable for integration into spaceborne QMS sensors. This section introduces a novel multiresonant tank circuit and an FPGA-based voltage control system to create a more precise implementation of auxiliary excitation techniques, and to enable the use of $f \approx 2$ higher order parametric excitations for space applications.

An operating frequency of $f = 1.9167$ is used as an example throughout this paper, since for a relatively large excitation amplitude ($q' \approx 0.15$), a sufficiently small stability island will be formed that should yield the maximum mass resolution $O(10^2)$ within the limitations of the developed electronics. Consequently, the increased transmission predicted for high resolution modes (*Konenkov et al.* 2005) when using auxiliary excitation will not be experimentally verified in this work.

2.3.1 Typical Quadrupole Tank Circuit

The maximum peak-to-peak voltage amplitude applied across the quadrupole rods is dependent on the desired maximum mass-to-charge ratio $(m/Z)_{\max}$ in *amu/charge number*, operating frequency Ω_{RF} in *rad/s*, and rod spacing r_{o} in *meters*, such that $V_{\max}$ in *Volts* becomes (*Dawson* 1997),

$$V_{\max} = \frac{(\Omega_{\mathrm{RF}}/2\pi)^2 r_{\mathrm{o}}^2}{7 \times 10^6} \left(\frac{m}{Z}\right)_{\max}. \tag{2.9}$$

For practical space instruments, amplitudes of several hundred volts are required to distinguish between neighboring species with large mass-to-charge ratios. In order to produce AC voltages of these magnitudes in a power efficient way, resonant circuit techniques are employed. The quadrupole rods themselves present an almost ideal capacitance C_{o}. When placed in parallel with an inductance L_{o} (typically the secondary windings of a step-up transformer), the rods form a parallel resonant LC tank circuit with resonant frequency,

$$\Omega_{\mathrm{RF}}^2 = \frac{1}{L_{\mathrm{o}} C_{\mathrm{o}}}. \tag{2.10}$$

The operating frequency of the QMS is therefore selected to be Ω_{RF}. Using the complex representation of the L_{o} and C_{o} components, the input impedance of this circuit (see Figure 2.4) can be written as,

$$Z_{\mathrm{IN,o}}(\omega) = \frac{-j\omega L_{\mathrm{o}} \Omega_{\mathrm{RF}}^2}{\omega^2 - \Omega_{\mathrm{RF}}^2}, \tag{2.11}$$

where there is a single pole at $\omega = \Omega_{\mathrm{RF}}$. In practice, $Z_{\mathrm{IN,o}}$ will remain finite as $\omega \to \Omega_{\mathrm{RF}}$ due to losses of inductors and capacitors used in the resonant circuit, as discussed in Appendix A.1. The sharpness of the actual resonant peak can be written in terms of the quality-factor (Q-factor) of the circuit,

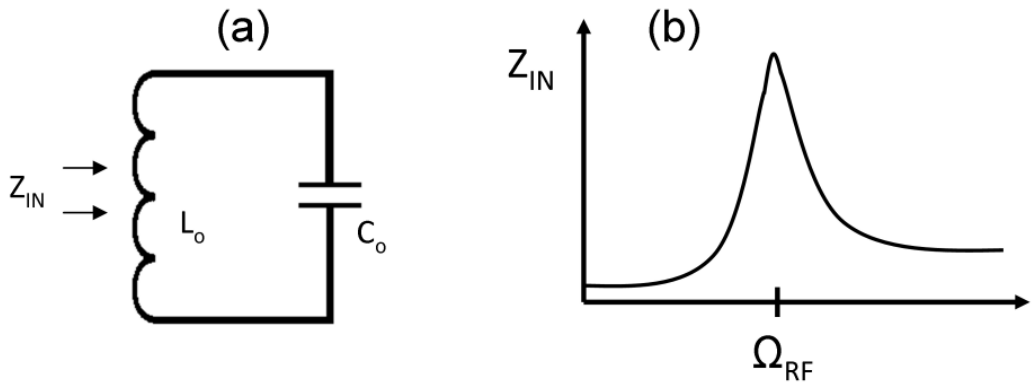

Figure 2.4: Typical single pole resonant LC tank circuit. (a) Schematic and (b) corresponding input impedance.

$$Q \equiv \frac{\omega}{\Delta\omega}, \tag{2.12}$$

where ω is the resonant frequency and $\Delta\omega$ is the $-3dB$ bandwidth of the peak. For a high-Q, low-loss, resonant LC circuit, the impedance quickly drops as one moves away from the pole in frequency. Resonant circuits for quadrupole applications typically require $Q \approx 100$.

Operation at the resonant frequency has a number of advantages. First, the driving current required to create a given potential difference across the quadrupole rods is a minimum since the current required to create a particular voltage drop across the rods varies inversely proportional to Z_{IN}. Second, the driving point impedance of the LC circuit is purely real, relaxing requirements on driving circuitry and causing zero phase shift of the RF waveform.

For a single-pole resonant circuit with quality factor Q, the required driving current magnitude I (normalized to the required current of an unmodulated QMS I_{o}) for an auxiliary signal with excitation strength q' at frequency $\omega_{\mathrm{aux}} = f\Omega_{\mathrm{RF}}$ can be written as (see Appendix A.1.e),

$$\left|\frac{I}{I_o}\right| = 1 + q'\sqrt{1 + Q^2(1 - \frac{1}{f^2})^2}. \tag{2.13}$$

Here, the fundamental frequency is driven at resonance, but the auxiliary signal is not, leading to a reduction of driving efficiency at the excitation frequency, as well as a significant phase advance of the auxiliary signal.

2.3.2 Multiresonant Tank Circuit

Unlike the traditional single pole tank circuit, a tank circuit that is resonant at more than one frequency enables resonances at all QMS operating frequencies. Like a single pole tank circuit, a multiresonant tank circuit should still have the quadrupole capacitance in parallel with the secondary windings of a step-up transformer. An impedance network synthesized by Foster's second form expansion (*Foster* 1924) adds elements in parallel to the traditional LC circuit, creating an additional pole and zero. This second pole is introduced at a frequency $\omega_{\text{aux}} \approx 2\Omega_{\text{RF}}$, resulting in an impedance network that has resonant peaks at both driving frequencies. This expansion can be generalized for any number of added poles, just by adding more Foster-derived networks to the circuit. Figure 2.5 shows a schematic for a tank circuit with an added Foster-derived network of L_1 and C_1. Here the inductance L_1 has been split up into 2 equal halves, $L_{1,\text{a}}$ and $L_{1,\text{b}}$, so that the circuit is balanced with respect to ground.

2.3.2.1 Component Selection

Given the schematic in Figure 2.5, the following frequencies are defined:

$$\omega_o^2 \equiv \frac{1}{L_o C_o} \text{ and } \omega_1^2 \equiv \frac{1}{L_1 C_1}. \tag{2.14}$$

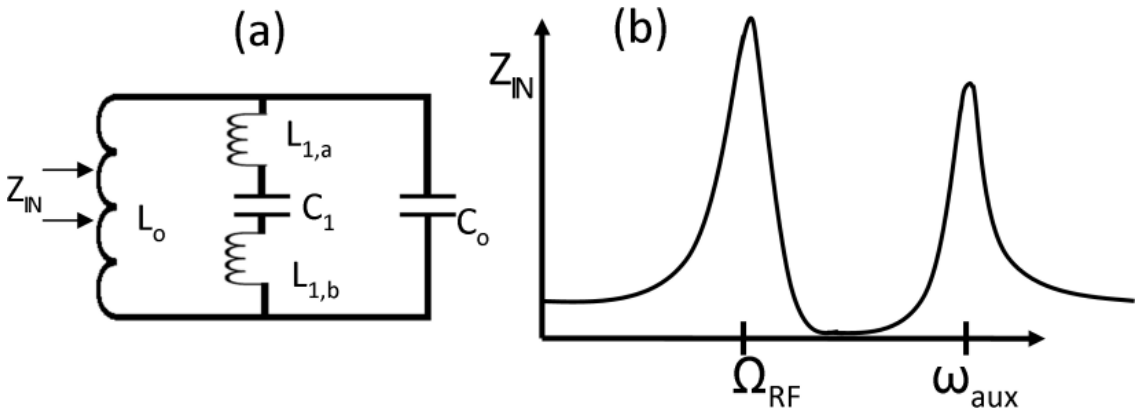

Figure 2.5: Multiresonant tank circuit. (a) Schematic with added Foster-derived impedance network and (b) corresponding circuit input impedance as a function of frequency.

Using the complex representation of the $L_{\rm o}, L_1, C_{\rm o}$, and C_1 components, the input impedance of this circuit, $Z_{\rm IN}$ can be written in terms of these frequencies and component values as,

$$Z_{\rm IN}(\omega) = \frac{-j\omega L_{\rm o}(\omega^2 - \omega_1^2)\omega_{\rm o}^2}{\omega^4 - \omega^2(\omega_1^2 + \omega_{\rm o}^2[1 + \frac{L_{\rm o}}{L_1}]) + \omega_{\rm o}^2\omega_1^2}. \tag{2.15}$$

This circuit has one zero at $\omega = \omega_1$ and two poles (i.e., the operating frequencies) on either side of that zero, $\Omega_{\rm RF}$ and $\omega_{\rm aux}$. These frequencies are related by parameters f and g defined such that f is the ratio between the two operating frequencies (equation (2.5)) and g is the location of the zero frequency relative to the two poles,

$$g \equiv \frac{\omega_{\rm aux} - \omega_1}{\omega_{\rm aux} - \Omega_{\rm RF}}. \tag{2.16}$$

Following the derivation in Appendix A.1.a, values of components $L_{\rm o}$, L_1, and C_1 can be determined for a desired operating point. Parameter values of $f \approx 2$ and $g \approx 0.5$ are used, yielding

$$L_{\rm o} \approx \frac{0.56}{\Omega_{\rm RF}^2 C_{\rm o}}, \tag{2.17}$$

$$C_1 \approx 0.43 C_{\mathrm{o}}, \tag{2.18}$$

and

$$L_1 \approx \frac{1.03}{\Omega_{\mathrm{RF}}^2 C_{\mathrm{o}}}. \tag{2.19}$$

From equations (2.17)-(2.19), the inductance values of L_1 (split into two equal parts $L_{1\mathrm{a}}$ and $L_{1\mathrm{b}}$) and L_{o} are of similar magnitudes, as are the capacitance values of C_1 and C_{o}. Therefore, in general, the design and component requirements of the single-resonant tank circuit apply to the capacitors and inductors used for a multiresonant tank circuit.

2.3.2.2 Driving Current Requirements

Following Appendix A.1.d, when driving the QMS at two resonant frequencies with excitation strength q', the new driving current I, normalized by I_{o}, can be written as

$$\left|\frac{I}{I_{\mathrm{o}}}\right| \approx \frac{Z_{\mathrm{INo}}(\Omega_{\mathrm{RF}})}{Z_{\mathrm{IN}}(\Omega_{\mathrm{RF}})}\left[1 + q' \times \frac{Z_{\mathrm{IN}}(\Omega_{\mathrm{RF}})}{Z_{\mathrm{IN}}(\omega_{\mathrm{aux}})}\right]. \tag{2.20}$$

The ratios $Z_{\mathrm{INo}}(\Omega_{\mathrm{RF}})/Z_{\mathrm{IN}}(\Omega_{\mathrm{RF}})$ and $Z_{\mathrm{IN}}(\Omega_{\mathrm{RF}})/Z_{\mathrm{IN}}(\omega_{\mathrm{aux}})$ are functions of the tank circuit parameters f and g. For example, for an operating point of $f = 1.9167$, $g = 0.5$, and Q of 200, the required driving current becomes

$$\left|\frac{I}{I_{\mathrm{o}}}\right| \approx 1.37(1 + 2.68q'). \tag{2.21}$$

This current can be compared to an operating point of $f = 0.9167$ for a single-pole system using equation (2.13) to yield

Table 2.1: Parameters for a high Q quadrupole single-pole-resonant tank circuit

Component(s)	Value(s)
Pole(s)	$1MHz$
Inductors	$L_0 = 154\mu H$
Losses (Q = 200 @ 1MHz)	$R_\mathrm{p} = 250k\Omega, R_\mathrm{s} = 1\Omega$
Capacitors	$C_0 = 165pF$
Transformer Coupling	$k = 0.5$
Transformer Primary	$L_\mathrm{p} = 500nH$

Table 2.2: Parameters for a high Q quadrupole multiresonant tank circuit

Component(s)	Value(s)
Pole(s)	$1MHz$, $1.9167MHz$
Inductors	$L_0 = 154\mu H$, $L_{1,\mathrm{a}} = L_{1,\mathrm{b}} = 163\mu H$
Losses (Q = 200 @ 1MHz)	$R_\mathrm{p} = 250k\Omega$,$R_\mathrm{s} = 1\Omega$
Capacitors	$C_0 = 95pF$, $C_1 = 37pF$
Transformer Coupling	$k = 0.5$
Transformer Primary	$L_\mathrm{p} = 500nH$

$$\left|\frac{I}{I_\mathrm{o}}\right| \approx 1 + 38q'. \tag{2.22}$$

For example, consider two high Q tank circuits with values as described in Tables 2.1 and 2.2. As a simplification, the self-capacitance of the inductors has been neglected.

The driving point impedance (as seen through a transformer primary, as modeled in Appendix A.1.d) for each circuit as a function of frequency has been calculated and is shown in Figure 2.6. The reduction in the resonant peak amplitude at the fundamental frequency is apparent, as well as the enhancement in the resonant peak at the chosen auxiliary frequency. Previously published operating points from *Konenkov et al.* (2001) and *Zhao et al.* (2009, 2010) are marked on Figure 2.6 as well as a proposed new operating point of $f = 1.9167$. For high Q resonant peaks, there is a significant decrease in required driving current in the new operating point and an increase in the dynamic range of the applied excitation amplitude. For lower Q inductors, the relative reduction in driving current will be less significant, but there

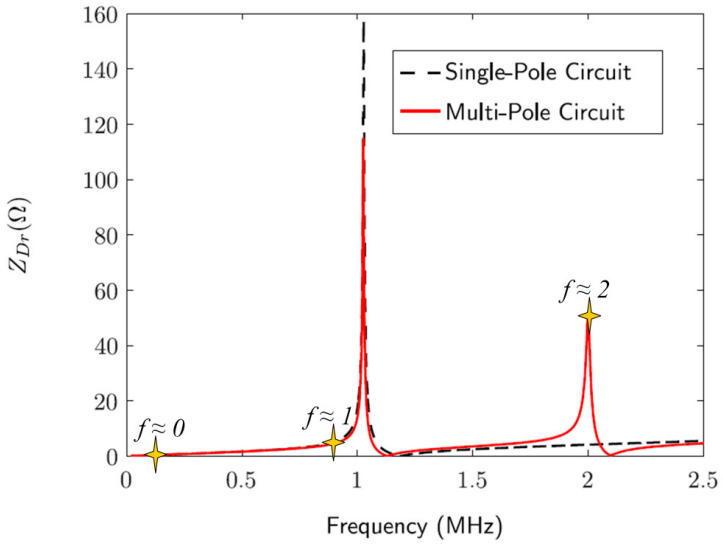

Figure 2.6: Calculated driving point impedance versus frequency for sample single-pole and multiresonant tank circuits. Experimentally verified operating points corresponding to $f \approx 0$ from *Zhao et al.* (2009, 2010), $f \approx 1$ from *Konenkov et al.* (2001), and $f \approx 2$ (this work), are indicated.

is still an increase in the dynamic range of the sensed excitation signal amplitude.

Consider the the island splitting illustrated in Figure 2.3 that results in $R_{0.10}$ on the order of 10^2. The relative driving current for each operating point is shown in Table 2.3, using the approximate calculations of equations (2.21) and (2.22), and also the simulated impedance curves shown in Figure 2.6. While it can be difficult to achieve both a maximum in the dynamic range of the auxiliary signal amplitude and a minimum in required driving current, an order of magnitude increase in the dynamic range is easily accessible for only a modest increase in required driving current over single-pole systems. From Table 2.3 it can be seen that equations (2.21) and (2.22) provide reasonable estimations for the required driving current for $f \approx 1$ and $f \approx 2$ operating points. These equations do not apply for $f \approx 0$ operating points since their derivation rely on the frequencies of interest being near circuit resonances.

With an appropriate selection of auxiliary frequency, higher resolution modes are accessible for approximately the same relative driving current as those that provide

Table 2.3: Relative driving current requirements for simulated upper stability island QMS operating points $R_{0.10}$ ≈100. The voltage scan line for each operating point passes through the center of the formed island.

Tank Circuit	f	q'	I/I_o [1]	I/I_o [2]
Single-Pole	0.0833	0.0085	-	5.98
Single-Pole	0.9167	0.016	1.61	1.38
Multi-Pole	1.9167	0.15	1.92	1.85

[1] Calculated using equations (2.21) and (2.22)

[2] Calculated from simulated driving point impedances and equation (2.20)

lower resolution measurements. Due to inherent limitations in the quality of developed electronics for this work, this paper focuses on lower mass resolution modes. However, with adequate precision of electronics, higher resolution modes should be accessible using the approach and techniques discussed here. For example, numerical simulations show that an operating point of $f = 1.95, q' = 0.15$ can yield a $R_{0.10}$ on the order of 10^3.

2.3.3 FPGA Voltage Control System

In order to drive a quadrupole, there must be circuitry that generates and manages all operating frequencies. In a typical single-frequency quadrupole, Ω_{RF} is synthesized through an analog oscillator and amplified to the appropriate amplitude. For multiple-frequency applications, however, the frequencies and amplitudes of multiple signals must be controlled. This complexity supports a transition from analog-based signal generation and management to a digital solution, namely a voltage control system housed in an FPGA, an integrated circuit that is composed of customizable digital logic blocks. FPGA-based control of quadrupoles has been previously studied and successfully implemented in the laboratory (*Jiang et al.* 2008; *Schaefer et al.* 2008). FPGA technology has also been successfully flown on spacecraft (*Mabry et al.* 1993), making these devices a suitable medium for development. While not a tremendous problem for terrestrial applications, the stability of the quadrupole RF source is an

issue for space sensors due to the mass and power limitations of the instrument, as well as the large variations in ambient temperature that is the instrument is likely to encounter. Since spaceborne quadrupole RF sources have traditionally been implemented as LC self-oscillators, their stability is a function of the stability of the employed inductances and capacitances, which must be matched to those of the resonant tank circuit. Digitally generated signals from an FPGA are more stable than their analog counterparts both in terms of frequency and amplitude drift, and can be dynamically adjusted to match the resonant frequencies of a tank circuit (*Vankka and Halonen* 2001).

2.3.3.1 System Overview

Figure 2.7 shows a block diagram of an FPGA-based voltage control system. An FPGA-based system allows for the use of direct digital synthesis (DDS) techniques to generate all frequency components of interest. These DDS blocks generate single-frequency sinusoidal signals of the appropriate amplitude for a signal with each clock cycle. Internal to the FPGA, each frequency component can be scaled and summed together to create the final quadrupole driving waveform that is output using a digital-to-analog converter (DAC). A second DAC controls the voltage sources that provides the DC component to the quadrupole rods. Analog-to-digital converters (ADC) read a mass-select line from the quadrupole control system to determine the output voltage amplitude, as well as a feedback signal from the quadrupole rods through a capacitive divider, which contributes to the value of C_0. This signal is used to detect the amplitudes of all of the frequency components and then, along with the desired operating voltage, provides input to control loops that create a closed-loop system ensuring that the correct amplitudes for each component are properly produced. Such a feedback mechanism employing matched ceramic capacitors is commonly used in flight instruments.

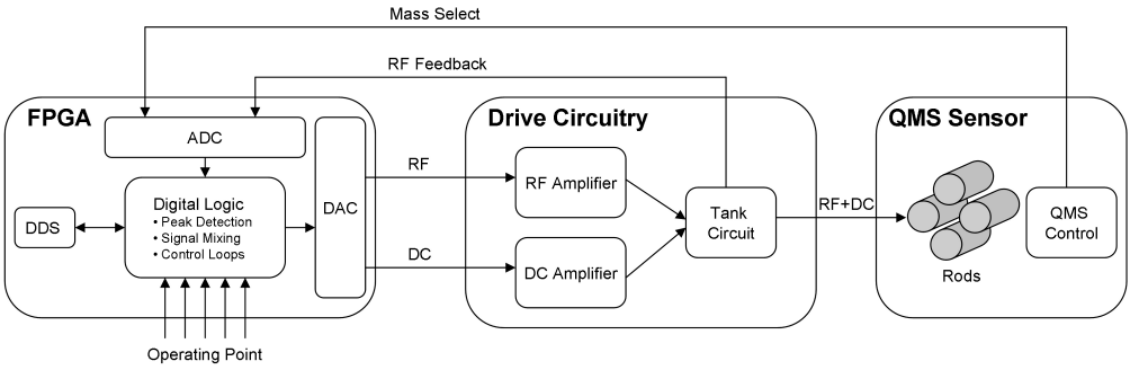

Figure 2.7: Block diagram of an FPGA-based voltage control system interfaced with QMS sensor and auxiliary electronics.

2.4 Experimental Verification and Discussion

The techniques presented can be applied to the development of completely new QMS systems, but they can also be applied to existing sensors. To experimentally verify auxiliary excitation near $f = 2$, a multiresonant tank circuit and an FPGA control system were designed to function with an existing Pfeiffer QMG422, a commercial quadrupole system. The Pfeiffer QMS has a set of 30-cm-long circular rods with 8-mm radii and an r_o spacing of 6.7-mm. As mentioned previously, most spaceborne QMS sensors use hyperbolic rods. However, for the purposes of this work, a circular rod sensor still allows one to the demonstrate the proposed techniques. The tuned tank circuit (see A.1.6) yielded resonant frequencies of 1.04 MHz and 1.99 MHz, corresponding to an operating point of $f = 1.9167$, and $g \approx 0.5$. The rods are biased at a negative potential with respect to the entrance aperture to control ion energy. The rod bias was set to inject ions of approximately 10 eV. Due to limitations of the ion source used, lowering the rod bias to inject ions below ≈10eV results in significantly reduced transmission without an improvement in performance.

For the FPGA voltage control system, a Xilinx XtremeDSP Virtex 4 development kit with 14-bit ADC/DAC capabilities was interfaced to a PC with MATLAB via a PCI bus. Since only two DAC channels were available for use with this FPGA kit,

Table 2.4: Resolution and stability of measurements in Figure 2.8. Our operating mode ($q' = 0.15$) eliminates the low-mass peak tail and increases the peak stability measured at all levels by approximately a factor of ≈ 4 with small impact on other performance characteristics.

	$R_{0.50}$	$R_{0.10}$	$R_{0.01}$	$T_{0.50}$	$T_{0.10}$	$T_{0.01}$
$q' = 0.00$[1]	288 ± 152	129 ± 29	52 ± 13	0.026 ± 0.017	0.018 ± 0.015	0.010 ± 0.012
$q' = 0.10$[2]	274 ± 105	129 ± 21	75 ± 8	0.021 ± 0.013	0.014 ± 0.010	0.010 ± 0.008
$q' = 0.15$[3]	207 ± 99	107 ± 9	81 ± 8	0.017 ± 0.004	0.012 ± 0.003	0.009 ± 0.002

[1] Scan line through the tip of the stability region, i.e. traditional operation.
[2] Scan line operation through the tip of a stability island. *Konenkov et al.* (2001); *Zhao et al.* (2009, 2010)
[3] Scan line operation through the center of a stability island.

a National Instrument USB-6215 device takes the output from the FPGA DAC and converts it to a DC+ and DC- control line using a LabVIEW VI. The mass select line from the Pfeiffer QMS electronics was sampled by one of the FPGA ADC channels and the Pfeiffer QMS RF control system was disabled. Auxiliary analog electronics, such as a simple class-C-like RF amplifier (see Appendix A.1.f) and high voltage DC amplifiers, were implemented separately from any Pfeiffer QMS electronics. It should be noted that although the Pfeiffer QMG422 is specified to have unity mass resolution up to a mass of 340 amu, the tank circuit/FPGA control system were designed for masses less than 100 amu, since they were only built for experimental verification of higher order auxiliary excitation. However, extension of this system to a higher mass range could be readily accomplished through the use of higher voltage rated components.

Argon gas was ionized and injected into the quadrupole and measured using the newly created RF/DC supply for three operating points: (1) traditional operation without auxiliary excitation, (2) scanning through the tip of a large stability island formed from 10% auxiliary excitation, and (3) scanning through the center of a small stability island formed from 15% auxiliary excitation. At 10 eV, the Ar^+ ions spend

approximately 40 RF cycles in the quadrupole rods. Stability diagrams created from sets of scan lines for each operating point are shown in Figures 2.8a, 2.8b, and 2.8c. The accumulation time for each scan line was 16ms per step, with 64 steps per amu, allowing sufficient settling time for each point. Note that this large accumulation time was a consequence of the unoptimized feedback control algorithm (see A.2.c), which could in principle be significantly accelerated. Mass peaks from 10 successive scans are shown for each operating point in Figures 2.8d,2.8e, and 2.8f, with the mean and standard deviation of measured 50%, 10%, and 1% mass resolutions and transmission values for the 10 scans indicated for each operating point in Table 2.4. The slope of each scan line was selected to produce the same average transmission for each operating point.

When operating with 10% auxiliary excitation scanning through the tip of the stability island (Figure 2.8e), the low mass peak tail appearing in the traditional mode of operation (Figure 2.8d) is eliminated, resulting in an almost 50% increase in 1% mass resolution. Some of this peak tailing may be attributed to ions not spending sufficient time between the rods due to their high axial energies, but can also be a consequence of using circular rod geometry. Nevertheless, the excitation peak shapes clearly do not exhibit this feature, demonstrating an increased ejection efficiency of unwanted ions when operating with excitation, consistent with expectations. However, the variability of transmissions for each of the 10 scans is relatively unchanged from that of the non-excitation case. Following the discussion in Section 2.2.1, mass peaks formed from operation at the tip of any stability region will be very sensitive to errors in the scan line slope, including those formed from the tip of an upper stability island.

For operation with 15% auxiliary excitation through the center of a stability island (Figure 2.8f), the same reduction in the low mass peak tail and corresponding 50% increase in 1% mass resolution are observed. However, as discussed in Section 2.2.1,

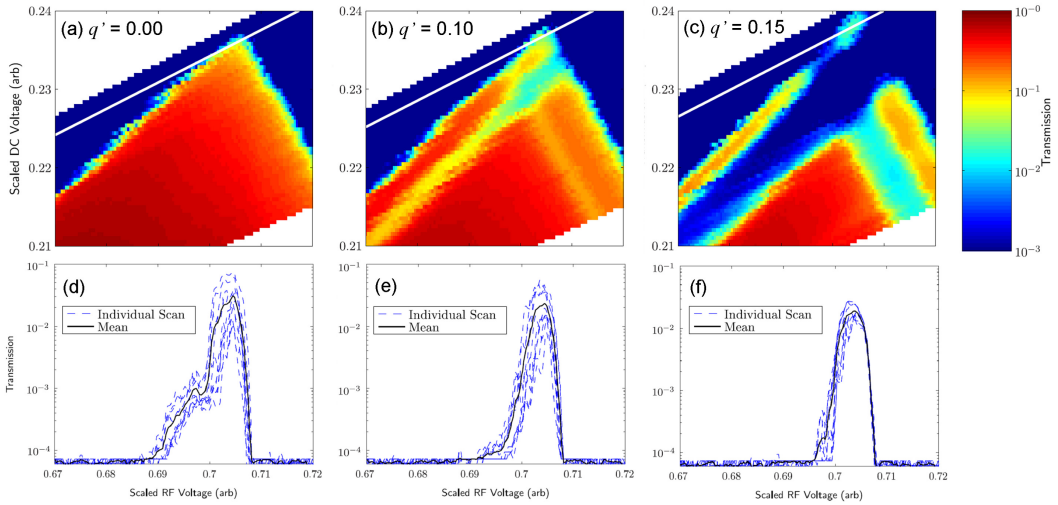

Figure 2.8: Laboratory results comparing higher order auxiliary excitation with traditional QMS operation. Measurements were performed with 10eV Ar^+ ions for (a) no auxiliary excitation ($q' = 0.00$) i.e., traditional operation, (b) $f = 1.9167$,$q' = 0.10$ auxiliary excitation analagous to previous work *Konenkov et al.* (2001); *Zhao et al.* (2009, 2010) where the scan line passes through the tip of the upper stability island, and (c) $f = 1.9167$,$q' = 0.15$ newly enabled auxiliary excitation where the scan lines passes through the center of the stability island. To measure the stability of each technique, (d),(e), and (f) show 10 successively measured mass peaks for each operating mode as shown as dashed lines, with the averages appearing as solid lines. Table 2.4 quantifies the resolution and stability of these peaks, showing that our proposed mode provides superior mass peaks.

mass peaks created from scan lines passing through the center of an island are less sensitive to errors in scan line slope. The result is a four-fold reduction in peak variability, or conversely stated, a four-fold improvement in peak stability over the other operating modes, as calculated by the reduction in the standard deviation of measured transmissions in Table 2.4. This repeatability enables a more accurate computation of relative abundances of molecular species, which is vital to achieve science goals of modern planetary and cometary missions employing QMS sensors. For the same count statistical error as the unmodulated QMS, this repeatability also allows for a decreased measurement time or signal accumulation, crucial for flybys or other transient effects that require high temporal resolution measurements.

Finally, it should be noted that the measured stability regions are in good agreement with the simulations from Section 2.2.2 in terms of the location of the expected upper stability island. The absolute transmission of ions through the QMS are necessarily different due the limitations of the simulation in terms of fringing field effects, ion entrance velocity distributions, and imperfections of the rods. The measurements do, however, show the improved peak shapes and mass resolution from operation in an upper stability island, experimentally verifying the presented techniques.

2.5 Conclusion

The presented techniques and technology are well suited for QMS sensors for space applications as they relax constraints on both the precision and power consumption of the drive electronics. Auxiliary excitation formation of an upper stability island enables QMS operating modes that boast higher mass resolution and improved peak shapes, as expected from previously discussed studies. However, implementation of these excitation techniques is now practical. Operation near $f = 2$ results in an increase in the dynamic range of the sensed auxiliary signal amplitude, and the use of the multiresonant tank circuit enables this increase, as well as the ability to operate

scan lines through the center of formed stability islands without a significant increase in driving current and power. Finally, digital synthesis of all signals allows for a more stable RF source whose operating frequencies can be dynamically adjusted in flight in response to any changes in the ambient conditions of the instrument. Such improvements enable the miniaturization of existing QMS sensors without sacrificing resolution, which could be important for various applications, such as a lander/rover based QMS.

Furthermore, multi-resonant tank circuit techniques can be generalized to any system that drives a capacitance with multiple frequencies. Therefore in addition to the discussed applications to QMS sensors, the approach introduced in this work has the potential to lead to improved performance of quadrupole ion traps and multipole arrays. The discussed FPGA-based voltage control system can also serve to increase the stability and decrease the analog circuit complexity of such sensors.

Previous experimental work primarily used auxiliary excitation techniques to reduce peak-tailing and counteract the effect of non-linear resonances, but these two effects are more prominent in commercial circular rod sensors rather than spaceflight instruments. Nevertheless, operation in the upper stability island should still have significant advantages even for ideal hyperbolic sensors. Stable peak shapes, for example, should be a significant improvement observed for both rod geometries. For flight sensors, the use of hyperbolic rods is still necessary in order to obtain the best possible performance. A switch to circular rod geometry would lead to significant cost reduction in both the production and calibration of such instruments, if similar performance to hyperbolic rod sets could be achieved.

We have developed a thorough analysis and novel design that allows operation of QMS technology using improved peak shapes and mass resolutions with minimal impact on total power of the overall QMS design. In particular, we have shown experimentally, that the QMS operation through the center of an upper stability island

formed from higher order parametric excitation provides improved peak shapes, an over 50% increase in instrument 1% mass resolution, and a 400% improvement of peak stability when compared with traditional sensor operation. While not demonstrated in this work due to limitations, auxiliary excitation techniques are also expected to provide increased transmission at high resolutions when compared to a traditional ideal quadrupole. Higher frequency auxiliary excitation enables all the advantages of previously studied lower order resonant excitation, but also results in an order of magnitude increase in the dynamic range of the sensed excitation signal amplitude, and in some cases, a net savings in required driving current for similar performance. Precision control of the excitation signal and large required driving currents for $f = 0$ and $f = 1$ operating points have been prohibitive factors for feasible implementation of parametric excitation techniques in quadrupole sensors. With the introduction of $f = 2$ operating points, this paper has successfully developed and experimentally verified a method designed for the practical implementation of parametric excitation for spaceborne, and even commercial QMS applications. Using a multiresonant tank circuit and FPGA voltage control system, this method enables breakthrough science for future planetary and cometary missions using some of the oldest high-heritage measurement techniques.

CHAPTER III

Enabling the next generation of spaceborne quadrupole mass spectrometers

This chapter is taken from: Gershman, D. J., M. Rubin, B. P. Block, M. Benna, P. R. Mahaffy, and T. H. Zurbuchen (2012), Enabling the next generation of spaceborne quadrupole mass spectrometers, in *Proceedings of Earth and Space 2012: Engineering for Extreme Environments*, American Society of Civil Engineers, Pasadena, CA.

Abstract

The quadrupole mass spectrometer (QMS) has over 30 years of spaceflight heritage in making important neutral gas and low energy ion observations. Given their geometrical constraints, these instruments are currently operated at the extreme limit of their capabilities. However, a technique called higher order auxiliary excitation provides a set of novel, robust, electronics-based solutions for improving the performance of these sensors. By driving the quadrupole rods with an additional frequency nearly twice that of the normal RF operating frequency, substantially increased abundance sensitivity, maximum attainable mass resolution, and peak stability can be achieved through operation of voltage scan lines through the center of formed upper stability islands. Such improvements are modeled using numerical simulations of ion trajectories in a quadrupole field with and without applied higher order auxiliary excitation.

When compared to a traditional QMS with a mass range up to 500Da, sensors can be designed with the same precision electronics to have expected mass ranges beyond 1500Da with a power increase of less than twice that of its heritage implementations.

3.1 Introduction

The quadrupole mass spectrometer (QMS) is the standard instrument for space-borne neutral gas composition measurements, with over 30 years of spaceflight heritage. The main science objectives of these missions have been to study solar system formation and evolution, which required precise isotopic measurements of relatively low mass species, such as ^{3}He/^{4}He (*Mahaffy et al.* 1998). These observations were typically enabled with a range of $< 150Da$. However, over the past decade, the planetary community has shifted its focus towards astrobiology: the exploration of the building blocks of life beyond Earth. Such a search is enabled by accurate measurements of higher masses ($> 1000Da$), organic compounds and volatiles with trace abundances (ppb level) (*Belton* 2003), requiring a new generation of QMS sensors to accurately characterize and resolve them.

In a typical QMS (*Dawson* 1997), low energy ions/recently ionized neutrals are focused into a beam and directed through a set of four conducting rods of hyperbolic cross section that form a quadrupole electric field with applied voltage. This applied potential has RF and DC components that modify the trajectories of incident ions. Proper tuning of these voltage amplitudes allows ions of a desired mass-per-charge ratio to successfully traverse the rod set and reach a particle detector, while other species are ejected, as illustrated in Figure 3.1.

Higher order auxiliary excitation of the quadrupole field is a robust, electronics-based solution that can compensate for distorted mass peak shapes produced from

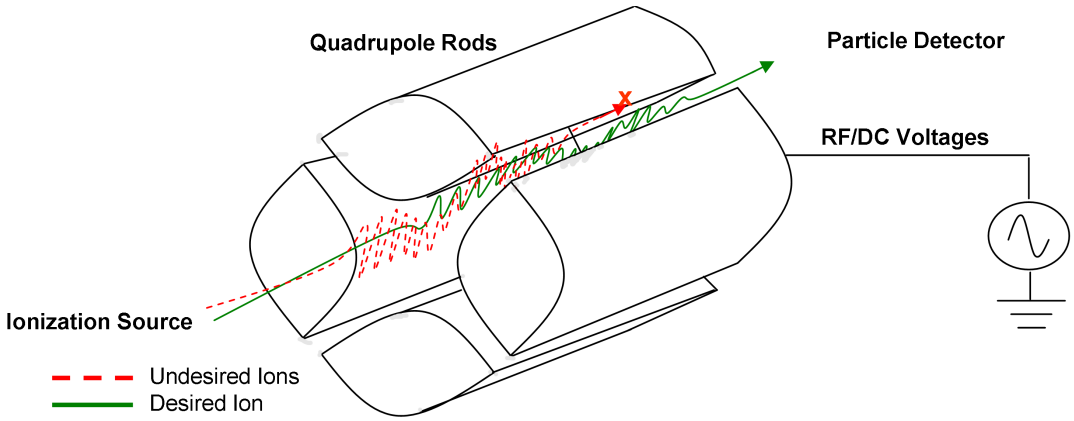

Figure 3.1: Overview of a spaceflight quadrupole mass spectrometer. RF and DC voltages are applied to a set of hyperbolic cross section rods to allow ions of a desired mass-to-charge ratio to successfully traverse the rods and reach a particle detector.

rod imperfections or misalignment caused by the stresses on the instrument imposed during the duration of the mission (see *Gershman et al.* (2011) and the references therein). This technique requires the driving of the quadrupole rods with an additional frequency component nearly twice the frequency of the nominal RF to enhance the resonant ejection of ions. Section 3.2 provides an overview of auxiliary excitation and its effects on mass peaks. In Section 3.3, numerical simulations of ion trajectories in a quadrupole field with and without applied excitation are used to map out possible instrument operating modes. Finally, Section 3.4 places these operating modes in context with the requirements of a next-generation high performance spaceflight QMS sensor.

3.2 Stability Island Formation

For a given mass-to-charge ratio, only specific combinations of RF (amplitude V and frequency Ω_o) and DC (amplitude U) voltages correspond to stable (periodic and bounded) ion trajectories in the rod set. These (U,V) voltage pairs form a stability region as shown in Figure 3.2a. In order to obtain a mass peak with a particular

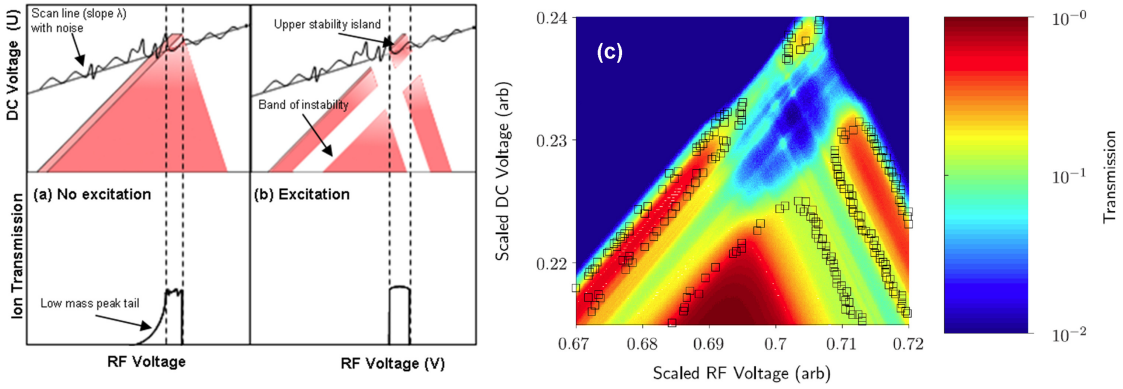

Figure 3.2: Stability region with a voltage scan line and corresponding ion mass peak shape for (a) traditional QMS with no applied excitation and (b) QMS with auxiliary excitation. Numerically modeled stability region splitting adapted from *Gershman et al.* (2011) for $f = 1.9167$, $q' = 0.15$ with corresponding laboratory measured boundaries indicated with squares, showing excellent agreement between the two.

resolution, the voltages are increased while keeping their ratio, $\lambda = U/V$, equal to a constant, creating a voltage scan line with slope λ that passes through a particular width of the stability region (*Dawson* 1997). The width of the mass peaks, i.e. the mass resolution, is set by the slope of this scan line, with higher resolution attained with higher scan line slopes. Near the tip of the stability region, transmission sharply decreases. Therefore, perturbations in the scan line due to errors in control electronics can result in significant changes in ion transmission, and consequently highly variable mass peak shapes.

As a consequence of rod imperfections or ion trajectories with too few RF cycles within the rods, the edge of the stability region on the low mass side tends to be diffuse and/or distorted, resulting in peak tailing, as shown in Figure 3.2a. This peak tail, which reduces the maximum attainable mass resolution of the instrument, can be mitigated through auxiliary excitation techniques, whereby an additional frequency component (frequency $f \cdot \Omega_o$ and amplitude $q' \cdot V$) drives the system. This excitation sharpens boundaries by enhancing the ejection efficiency of undesired ions, and creates

additional bands of instability within the original stability region, forming an isolated upper stability island, as shown in Figure 3.2b. (*Konenkov et al.* 2001). Operating scan lines through the center of these islands has been experimentally shown to result in increased peak stability when compared with operation near the tips of stability regions, the expected reason for this being that electronics errors in scan line slope do not result in significant changes to peak shape (*Gershman et al.* 2011). The formation of upper stability island formation has been studied extensively with both numerical modeling and experimental work with good agreement between the two, as shown in Figure 3.2c. However, there has not yet been a set of systematic simulations that map out the locations of these islands for a given set of excitation parameters.

3.3 Numerical Simulations

A 4th order Runge-Kutta numerical scheme as described by *Konenkov et al.* (2001) and discussed in detail by *Gershman et al.* (2011) is used to simulate the equations of motion of ions with an ideal quadrupole field with and without higher order auxiliary excitation. For each scaled voltage setting, 40,000 ions were randomly initialized about a circular aperture spanning 10% of the spacing between the quadrupole rods. The initial energy of each ion was set such that it spends 100 RF cycles traversing the rods, experiencing one of 20 uniformly distributed initial RF phases between 0 and 2π. Such numerical models provide excellent agreement with analagous laboratory measurements in terms of stability boundary locations, as shown in Figure 3.2b, and will be used to map stability island locations.

For a given relative excitation frequency, f, and excitation amplitude, q', a series of scan lines were swept through an upper stability island, producing a set of mass peaks. The mass resolution at 10% of the peak height ($R_{0.1}$) and the transmission ($T_{0.1}$) values were calculated for each, following the definitions in *Gershman et al.* (2011). While mass resolutions at levels below 10% are of interest to some sensors requiring very

high dynamic range, the $R_{0.1}$ and $T_{0.1}$ values are sufficient and well-suited to map stability island locations as well as general trends of instrument performance. Such a mapping will yield relationships between q',f, and λ, such as $d\lambda/dq'$.

In addition to $R_{0.1}$ and $T_{0.1}$, the required scan line precision to obtain a particular peak shape is also calculated. This precision, $(|d\lambda|/\lambda)$, is defined as the change of the scan line slope required to significantly change the mass peak shape. Consider a mass peak created from scan line slope λ and a mass peak created from scan line slope $\lambda + d\lambda$ for the same set of excitation parameters. The relative pointwise error between these two peaks can be readily computed. $|d\lambda|$ is set by the minimum change in slope that produces a relative mass peak difference greater than 10%.

For small changes in q', the location of a stability island shifts slightly, requiring a scan line of slope $\lambda + d\lambda$ to cross its new center location. The required excitation amplitude precision, $|dq'|/q'$, can therefore be estimated by dividing the calculated $|d\lambda|/\lambda$ values by the corresponding $q'd\lambda/dq'$ values obtained from simulations. Scan line slopes within the range $\lambda \pm |d\lambda|$ or excitation amplitudes within the range $q' \pm |dq'|$ will produce similar mass peak shapes, with almost identical $R_{0.1}$ and $T_{0.1}$ values. Therefore, high $|d\lambda|/\lambda$ or $|dq'|/q'$ values translate to less required precision of control electronics in order to obtain a stable measurement.

For excitation frequencies $f = 1.91$, 1.93, and 1.95, the excitation amplitude q' was increased in increments of 5%. This range of excitation frequencies was selected to create a wide range of upper stability island sizes such that $R_{0.1}$ from 100-1000 could be created for small excitation values ($q' < 0.20$). Data points corresponding to excitation amplitudes that were not sufficiently high to produce an upper stability island at a given excitation frequency are not considered.

Figures 3.3a and 3.3b show dependence of stability island size and location on excitation parameters f and q'. For a given excitation amplitude q' (Figure

The transmission for relative auxiliary frequencies $f = 1.91$, 1.93, 1.95 are shown

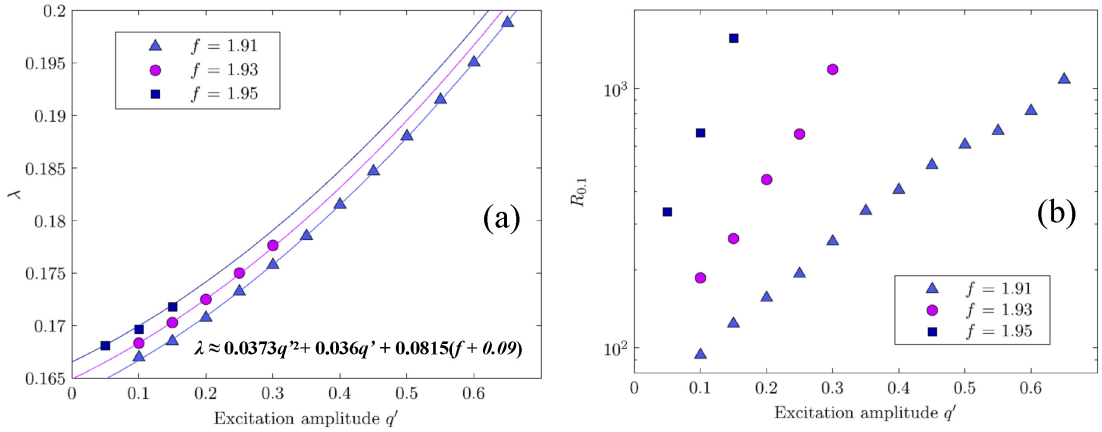

Figure 3.3: Dependence of the (a) scan line slope λ required to pass through the center of an upper stability island, and (b) corresponding mass resolution $R_{0.1}$ on the excitation amplitude q' for relative excitation frequencies $f = 1.91, 1.93$, and 1.95. An empirical relationship between q', f, λ is also included in (a).

in Figure 3.4a, along with the corresponding simulations for an ideal quadrupole without added excitation. For low resolutions, the non-excitation mass peaks yield higher transmissions, but this enhancement quickly drops, with a crossover with the excitation curves occurring at approximately $R_{0.1} = 300$. The maximum attained resolution for the non-excitation case was about 500, consistent with the expectations of ions spending 100 RF cycles in an ideal quadrupole (*Dawson* 1997). However, with applied excitation, $R_{0.1} > 1000$ values are readily achievable. The transmissions for mass peaks obtained with auxiliary excitation do not show strong dependence on frequency parameter f, suggesting that achievable instrument performance may be relatively independent of specific operating point selection.

Finally, Figure 3.4b shows the permitted variability of the scan line slope as a function of mass resolution. For all resolutions, auxiliary excitation operation requires about 3-4 times less scan line precision when compared with a traditional quadrupole, i.e. $|d\lambda|/\lambda$ is 3-4 times greater than that of the traditional QMS. This

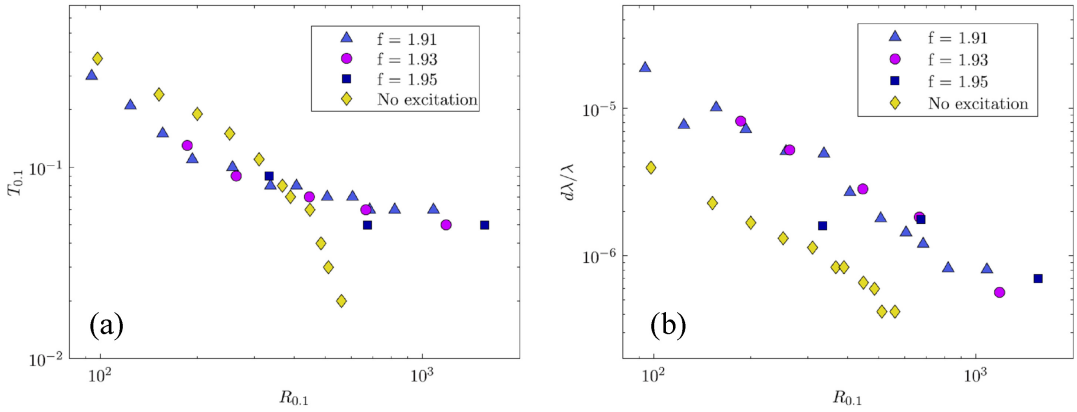

Figure 3.4: Transmission as a function of resolution for relative excitation frequencies $f = 1.91, 1.93$, and 1.95, as well as for the traditional non-excitation case. (b) Permitted scan line slope variability $|d\lambda|\lambda$ to maintain mass peak shape within 10% as a function of resolution for the same operating modes as in (a).

factor is consistent with the laboratory experiments from *Gershman et al.* (2011), where auxiliary excitation provided increased peak stability by a factor of 4. As with Figure 3.4a, there is no strong dependence of the required scan line precision on excitation frequency. Using the $d\lambda/dq'$ calculated from the data in Figure 3.3a, the quantity dq'/q' will be over an order of magnitude larger than its corresponding $|d\lambda|/\lambda$ values for all $q' < 1$. Therefore, the precision of the auxiliary electronics will likely not be the driving factor for determining maximum instrument performance.

3.4 Design Considerations

A previous laboratory implementation of higher order auxiliary excitation utilized an FPGA-based voltage control system that digitally synthesizes all frequency components of interest and drives a multi-resonant tank circuit attached to the quadrupole rods. The tank circuit is tuned to have two resonances, one at each operating frequency. This method is discussed in detail by *Gershman et al.* (2011), and is used in

this paper as a baseline solution for implementation. For such a system, specifications for a QMS sensor that utilizes higher order auxiliary excitation can be developed and then compared with those of a traditional sensor.

3.4.1 Tank Circuit Tuning

The tuning of a multi-resonant circuit may change with time due to stresses on instrument components from the space environment or the spacecraft launch process. While some amount of compensation can be built into the electronics to limit the scope of this detuning, the robustness of the instrument will ultimately depend on the flexibility of operating point selection for a given flight configuration. Digitization of the voltage waveforms with an FPGA in principle allows for arbitrary frequency selection. However, in practice, operation is limited to frequencies within the bandwidth of the system resonances. The -3dB bandwidth of each resonance ($\Delta\Omega_o$) is typically defined in terms of aQ-factor, such that, $Q = \Omega_o/\Delta\Omega_o$ (*Gershman et al.* 2011).

For a system with two resonant frequencies tuned to frequencies Ω_o and $f_o\dot{\Omega}_o$, the range of possible f values can be shown to be approximately $f_o \pm \Delta f/2$, with $\Delta f \approx 3/Q$. Therefore, for finite Q resonances, the system has some built-in tolerance for detuning. A smaller Q-factor will widen the bandwidth of the resonances at the cost of increased driving current, i.e. power. A conservative value of $Q = 100$ is assumed, noting the difficulties of implementing stable, high-Q inductors for use in the tank circuit given the large currents required to produce thousands of volts across the quadrupole rods. This bandwidth enables a relatively large set of possible operating frequencies, with $\Delta f = 0.03$. Following the results of Section 3.3, and enforcing a maximum $q' = 0.2$ to conserve power, a nominally tuned f value of 1.92 enables center-island operation for $R_{0.1}$ between 100 and 500. A nominally tuned f value of 1.94 enables center-island operation for $R_{0.1}$ between 200 and 1500 .

3.4.2 Power Consumption

The high voltages applied to the quadrupole rods will be the primary driver of instrument power consumption. The transfer impedance of the multiresonant tank circuit at each resonant frequency will dictate the required driving current necessary to produce such voltages. From *Gershman et al.* (2011), an estimate of the relative driving current required for auxiliary excitation can be obtained. For a tank circuit with $Q = 100$, the required driving current will be approximately 1.5 times that which is required for traditional operation, even with no applied excitation. This increase is solely a consequence of using a multiresonant tank circuit and can only be reduced by using higher Qinductors. For a $q'_{\max} = 0.20$, the required driving current will be approximately twice that of a traditional system.

3.4.3 Electronics

Using the chosen operating frequencies, an upper stability island cannot be formed that provides $R_{0.1} < 100$ for center-island operation. However, operation of scan lines through the center of a stability island only becomes vital for higher resolution modes, when errors in the control system electronics correspond to operating point changes on the order of the size of the stability island. From Section 3.3, center-island scan line operation allows for a mass resolution of 500 to be achieved with electronics that are approximately half the scan line precision as a traditional quadrupole, or with the same scan line precision, resolutions on the order of 1500 can be achieved. Lower resolution modes do not require scan line operation through the center of a stability island since sufficient mass peak shapes, peak stability and transmission can be obtained through operation of scan lines through the tips of a large island, well within the limitations of electronics precision. In fact, as shown in Section 3.3, higher transmission (i.e. higher sensitivity) measurements can be achieved at lower resolutions without the use of auxiliary excitation.

It should be mentioned that these precision requirements are based on the assumption that the operating frequencies are stable. This assumption should be valid for digitally synthesized frequency components, which are unlikely to significantly drift even with changing ambient conditions.

3.4.4 Scanning Time

The mass scanning speed of a QMS will depend on the settling time of the electronics associated with each change in voltage and the required accumulation time at each voltage step to obtain a reliable signal. The settling time will be dominated by the Q of the multiresonant tank circuit. The settling time varies as $2\pi Q/\Omega_o$ (*Gershman et al.* 2011), and should be independent of any auxiliary excitation implementation. However, there will be other increases in settling time associated with auxiliary excitation. First, following the simple sensing algorithm from *Gershman et al.* (2011), a minimum of $1/(2-f)$ periods of the RF must be measured to sense the envelope modulation of the quadrupole waveform. This acquisition time will be further increased if more modulation periods are measured for improved statistical accuracy. Fortunately, as the RF frequency is typically in the MHz range, an increased acquisition time may ultimately be negligible.

For a closed-loop amplitude control feedback system, a QMS utilizing auxiliary excitation will require approximately twice the number of time steps as a traditional sensor, since the excitation amplitude will necessarily converge after that of the main RF waveform. However, if the ion transmission increases as a result of auxiliary excitation, this increased settling time may be compensated by a reduction in the required accumulation time of the ion detector signal.

Table 3.1: Specifications for three QMS sensors

	Traditional (island tip)	Electronics-Driven (upper island operation)	Mass-Range Driven (upper island operation)
Resolution and Transmission			
$R_{0.1}$	0-500	100-500	200-1500
$T_{0.1}(R_{\max})$	3%	6%	5%
Electronics			
f	N/A	1.905-1.935	1.925-1.955
Power Consumption	1	1.5-2	1.5-2
Scan Line Precision	0.004%	0.01%	0.004%
Settling Time			
Tank Circuit	1	1	1
Waveform Acquisition	1	11-16	14-23
Feedback Control	1	2	2

3.4.5 New QMS Instrument Designs

To integrate the discussed design considerations, Table 3.1 shows specifications for three different QMS sensor designs. The first sensor is part of a class of traditional QMS designs with maximum $R_{0.1} = 500$. The other sensors are operated with higher order excitation, one electronics-driven, and the other mass-range driven. The electronics-driven QMS was designed to match the mass range of the traditional sensor but with relaxed electronics precision requirements. The mass-range driven QMS has the same electronics-precision requirements as the traditional sensor, but with an extended mass range. These designs assume a multiresonant tank circuit with $Q = 100$, that ions spend 100 RF cycles traversing the rods within an entrance aperture 10% of the rod spacing, and the excitation amplitude q' is less than 0.20.

Both of the auxiliary excitation sensors boast increased ion transmission at their highest mass resolution when compared to a traditional QMS. This increase translates directly into increased abundance sensitivity for given ion detection capabilities. The

main disadvantage of implementing a higher order auxiliary excitation technique is the power consumption of the instrument. Approximately a 50%-100% increase in power is required, regardless of any excitation, to achieve improved performance. However, from the designs in Table 3.1, it is clear that through implementation of higher order auxiliary excitation techniques, significantly improved performance can be achieved with currently available technology.

3.5 Conclusions

Using numerical simulations of ion trajectories in an ideal QMS, higher order auxiliary excitation techniques have been shown to provide a two-fold increase in abundance sensitivity, and a more than three fold improvement in both maximum attainable mass resolution and peak stability. Improving maximum attainable resolution for an ideal quadrupole without excitation requires an increase in the number of RF cycles that ions spend traversing the rods. This increase can be accomplished either though imposing additional restrictions on the instruments ion source performance, or through use of longer rods. Higher order auxiliary excitation, however, provides a flexible, electronics-based solution, with the caveat being an increase in instrument power consumption and complexity. The performance enhancements obtained through center-island scan line operation are relatively independent of operating frequencies or excitation amplitudes. Therefore, the engineering capabilities of the instrument can be used to directly inform selection of an appropriate operating point, making higher order auxiliary excitation a robust technique suitable for applications to the next-generation of spaceflight quadrupole mass spectrometers.

CHAPTER IV

Comparing the performance of hyperbolic and circular rod quadrupole mass spectrometers with applied higher order parametric excitation

This chapter is taken from: Gershman, D. J., B. P. Block, M. Rubin, M. Benna, P. R. Mahaffy, and T. H. Zurbuchen (2012), Comparing the performance of hyperbolic and circular rod quadrupole mass spectrometers with applied higher order parametric excitation, *Int. J. Mass. Spectrom.*, in press.

Abstract

This work applies higher order auxiliary excitation techniques to two types of quadrupole mass spectrometers (QMSs): commercial systems and spaceborne instruments. The operational settings of a circular rod geometry commercial system and an engineering test-bed for a hyperbolic rod geometry spaceborne instrument were matched, with the relative performance of each sensor characterized with and without applied excitation using isotopic measurements of Kr+. Each instrument was operated at the limit of the test electronics to determine the effect of auxiliary excitation on extending instrument capabilities. For the circular rod sensor, with applied excitation, a doubling of the mass resolution at 1% of peak transmission resulted from

the elimination of the low-mass side peak tail typical of such rod geometries. The mass peak stability and ion rejection efficiency were also increased by factors of 2 and 10, respectively, with voltage scan lines passing through the center of stability islands formed from auxiliary excitation. Auxiliary excitation also resulted in factors of 6 and 2 in peak stability and ion rejection efficiency, respectively, for the hyperbolic rod sensor. These results not only have significant implications for the use of circular rod quadrupoles with applied excitation as a suitable replacement for traditional hyperbolic rod sensors, but also for extending the capabilities of existing hyperbolic rod QMSs for the next generation of spaceborne instruments and low-mass commercial systems.

4.1 Introduction

For the past 50 years, the quadrupole mass spectrometer (QMS) has been the instrument of choice for high quality neutral gas composition measurements in both the commercial realm (*March and Todd* 2005) and for spaceborne applications (*Waite et al.* 2004; *Niemann et al.* 1992, 1980, 1997). Commercial sensors typically employ more easily manufactured circular rods, but can suffer from degraded peak shapes and difficulties in attaining high resolution measurements (*Dawson* 1997), though proper sizing of circular rods can help to minimize these effects (*Douglas and Konenkov* 2002).

Spaceborne instruments, on the other hand, have relied on hyperbolic geometry, as these rods provide a near-ideal quadrupole electric field. However, it can be extremely challenging and expensive to manufacture and align such rod sets. In their traditional mode of operation, these sensors are perhaps at the limits of their capabilities, as newly developed ion sources with optimal beam design virtually eliminate affects of

fringing fields, yet the highest attained mass range for modern space instruments extends only to 500 Da (*Mahaffy et al.* 2012). However, for mass spectrometers that are part of future planetary and cometary missions, mass ranges beyond 1000 Da will be required in order to measure complex organics and volatiles that are currently at the center of scientific interest, as they relate to the investigation of sources and processes of life in space (*Belton* 2003). In order to use well-proven QMS technology for the next generation of spaceborne mass spectrometers, new techniques that improve overall instrument performance must be developed and tested. Auxiliary excitation techniques will be shown here to accomplish such an improvement.

Stable ion trajectories in a QMS are created by applying two voltages to the quadrupole rods: an RF waveform at a frequency typically in the MHz range, and a DC potential. A subset of the possible RF and DC amplitudes results in stable trajectories for a given species of ion. Mass peaks corresponding to increasing ion mass-to-charge (m/z) ratios can be produced by increasing the RF/DC voltage amplitudes while keeping their ratio constant, forming a voltage scan line. The slope of this scan line will determine the width, i.e. the resolution, of the mass peak (*Dawson* 1997). In order to maximize the sensitivity of the instrument, quadrupoles are often tuned to unity mass resolution such that the scan line slope is set to increase with increasing m/z ratios (*Marchand and Marmet* 1964). However, for a measurement confined to a small mass range, a constant scan line slope is appropriate.

Auxiliary excitation techniques apply an addition RF component to the rods, exciting resonances in the system that split the original solid RF/DC stability region into small islands of stability. Operation with voltage scan lines through the tips of these islands has been shown through experimental work (*Devant et al.* 1989; *Miseki* 1993; *Konenkov et al.* 2001; *Zhao et al.* 2009, 2010) to significantly improve the performance of circular rod quadrupole mass spectrometers by improving peak shapes, abundance sensitivity, and mass resolution. Auxiliary frequencies are typically set to

be very low (*Zhao et al.* 2009, 2010), or near to, but not identical to that of the fundamental RF signal (*Konenkov et al.* 2001). These techniques cannot easily be applied to space instrumentation due to the high degree of precision required to drive and control such a small auxiliary signal amplitude. Furthermore, since the quadrupole rods typically act as part of a resonant tank circuit that enables more power efficient operation, inserting a signal off-resonance, even at small amplitudes, requires large amounts of driving current, a key resource on spacecraft (*Gershman et al.* 2011). Even for terrestrial applications, large required driving currents can prohibit the creation of very small islands of stability such that it has been easier to create larger islands of stability and operate scan lines through their tips.

Higher order auxiliary excitation techniques (*Gershman et al.* 2011, 2012c) drive the quadrupole rods at a frequency nearly twice that of the fundamental frequency. These higher order techniques provide the same advantages of previously demonstrated auxiliary excitation driven at lower frequencies, but allow improved dynamic range of the sensed auxiliary signal amplitude. When used in conjunction with a multi-resonant tank circuit (*Gershman et al.* 2011), the quadrupole rods can be driven efficiently at both the fundamental RF and auxiliary frequencies, enabling auxiliary excitation for power-limited applications such as space-flight. Furthermore, larger excitation amplitudes permitted by this circuit enables operation of voltage scan lines through the center of small upper stability islands, providing more stable peak shapes than voltage scan lines passing through the island tips (*Gershman et al.* 2011), an improvement important for space and commercial applications alike.

The advantages of higher order auxiliary excitation techniques when applied to circular rod QMSs at lower resolution (100) have recently been experimentally verified using measurements of Ar+ (*Gershman et al.* 2011). Numerical simulations have mapped out the set of practical instrument operating points (*Gershman et al.* 2012c) as well as predicted that these techniques should improve the performance of sensors

with ideal hyperbolic rod geometry (*Gershman et al.* 2012c; *Konenkov et al.* 2005) but these advantages have yet to be experimentally verified. This work will serve as the first experimental demonstration of higher order auxiliary excitation techniques in a hyperbolic geometry QMS. Here, higher order auxiliary excitation is applied to two sets of rod sets (Section 4.2) with circular and hyperbolic rod geometry respectively, whose operating characteristics have been well matched. Isotopic measurements of Kr+ (Section 4.3) are used as a baseline for gauging sensor performance.

4.2 Experimental setup

A Field Programmable Gate Array (FPGA)-based QMS voltage control system and multiresonant tank circuit described in detail by *Gershman et al.* (2011) were modularized and interfaced with two existing quadrupoles as drop-in replacements for their respective drive electronics. The first, a commercial Pfeiffer QMG422 circular rod system, was previously used to demonstrate the improved performance using high order auxiliary excitation with measurements of Ar+ *Gershman et al.* (2011). For this sensor, with a rod radius to spacing ratio of 1.20, a significant low-mass peak tail is expected (*Douglas and Konenkov* 2002) and in fact observed in nominal sensor operation (*Gershman et al.* 2011). The second quadrupole was a space-flight quality hyperbolic rod system, the engineering test bed for the Sample Analysis at Mars (SAM) instrument (*Mahaffy et al.* 2012) on the Mars Science Laboratory (MSL) mission. The physical characteristics as well as timing and operating parameters for each sensor are summarized in Table 4.1 with details on their selection described throughout this section.

In order to obtain a meaningful comparison between different QMS sensors, the operating characteristics of each must be matched as well as possible, though there will inevitably be some discrepancy between their observed performances due to differences in test setups. Nevertheless, matching the important non-dimensional num-

Table 4.1: Characteristics of QMS sensors used in this study

	Hyperbolic	**Circular**
Physical characteristics		
Rod length (l) (mm)	150	300
Rod spacing (r_o) (mm)	5.08	6.7
Rod radius (mm)	N/A	8
Operating point		
Fundamental frequency $(\Omega/2\pi)(MHz)$	1.40	1.06
Excitation frequency $(\omega/2\pi)(MHz)$	2.6834	2.0317
RF amplitude per m/z $(V/(m/z))$ (Vp/(amu/e))	3.68	3.67
Nominal rod bias potential (ϕ_b) (V)	-2	+70
Nominal source exit potential (ϕ_s) (V)	0	0
Nominal ion energy from source ($_{i,s}$) (V)	1±0.5	80±5
Accumulations		
Steps per m/z	20	64
Step dwell time (ms)	100	0
Step accumulation time (ms)	50	16

bers across systems ensures maximum consistency. The dimensionless numbers of a quadrupole sensor, a, q, and ξ, arise from the equations of motion of an ion in an ideal quadrupole field with added auxiliary excitation (*Konenkov et al.* 2005),

$$\frac{d^2x}{d\xi^2} + (a - 2qcos(2\xi))x - 2q'cos(2\xi\frac{\omega}{\Omega})x = 0 \text{ and} \tag{4.1}$$

$$\frac{d^2y}{d\xi^2} - (a - 2qcos(2\xi))y + 2q'cos(2\xi\frac{\omega}{\Omega})y = 0, \tag{4.2}$$

where

$$a = \frac{8eU}{(m/z)r_o^2\Omega^2} \text{ , } q = \frac{4eV}{(m/z)r_o^2\Omega^2} \text{ and } \xi = \frac{\Omega t}{2}. \tag{4.3}$$

Here, x and y are the transverse position coordinates of an ion in a 2D quadrupole field, U and V are the DC and RF voltage amplitudes pole to ground, respectively, r_o is the rod spacing, Ω is the fundamental operating frequency, ω is the auxiliary frequency, e is the unit of an electric charge, m/z is the mass-to-charge ratio, and t

is time. q' is the relative strength of the excitation defined by $q' = qV_{\text{ex}}/V$ where V_{ex} is the excitation voltage amplitude.

The matching of a and q will be accomplished through appropriate frequency selection and the use of identical drive electronics for each sensor. ξ will be matched across sensors through proper selection of ion axial energies. Finally, the dynamic range of the detected ion signal will be adjusted to ensure similar counting statistics for each QMS.

The experimental setup here has the same characteristics and limitations as those described in *Gershman et al.* (2011). Consequently, we will focus on low-resolution ($(m/z)/\Delta(m/z)$ $\sim$100) measurements. However, these measurements will be sufficient to show good agreement with previous numerical simulations and illustrate the advantages of auxiliary excitation for both rod geometries. For each sensor, the multiresonant tank circuit was tuned to obtain a ratio of auxiliary to fundamental frequencies (ω/Ω) of 1.9167. Such precise frequency selection is enabled through use of digitally synthesized signals by an FPGA. This excitation frequency will create an upper stability island that will produce a mass peak with resolution on the order of 100 when a voltage scan line, whose required precision is at the limit of the developed electronics, is swept through the island center. Only by examining the operation at this limit can the effectiveness of auxiliary excitation techniques on extending the capabilities of QMS sensors be determined.

4.2.1 Matching of a and q

For a typical high-resolution measurement, operation of scan lines through the tip of the stability region requires voltages such that a $\approx$0.237 and q $\approx$0.706. While it might make intuitive sense to operate each quadrupole at identical frequencies, the differing physical dimensions of each sensor will result in different required voltages to achieve stable trajectories for a particular ion. The driving electronics will control

the accuracy and precision of the instrument operating point, and consequently, the trajectories of analyzed ions. Following equation (4.3), if the $U/(m/z)$ and $V/(m/z)$ ratios are matched across rod sets, noise in the voltage-control system will have similar effects on ion trajectories for each sensor, and differences in performance should be dominated by differences in the physical properties of the rods, rather than electronic effects since these voltages scale as $r_o^2\Omega^2$.

Since all frequency components are digitally generated using an FPGA, a change in operating frequency is readily accomplished. Therefore, given the field radius, r_o, of a rod set, the operating frequencies can be adjusted to obtain the same $V/(m/z)$ ratios. For $\Omega/2\pi$ = 1.06MHz and r_o = 6.7mm for the circular rod set, and $\Omega/2\pi$ = 1.4MHz and r_o = 5.08 mm for the hyperbolic rod set, the $V/(m/z)$ values are well matched at 3.67 $V/(amu/e)$ and 3.68 $V/(amu/e)$ respectively, as shown in Table 4.1. Although tuning adjustments were made to the multiresonant tank circuit to accommodate each sensors rod capacitance, the overall drive electronics were identical for each quadrupole.

4.2.2 Matching of the number of RF cycles in the field

Assuming constant ion axial energy E, ξ from Section 4.2 can be written as a function of the rod length, ion energy, and ion m/z ratio, as,

$$\xi = \frac{\Omega l}{2\sqrt{\frac{2eE}{(m/z)}}}. \tag{4.4}$$

The number of RF cycles (n) experienced by an ion traversing the length of the quadrupole rods is commonly used instead of ξ to describe instrument behavior (*Dawson* 1997). The two quantities are related to one another by $\xi = \pi n$. The ideal quadrupole, which is infinitely long, has an infinite number of RF cycles to filter out unwanted ions, creating perfectly sharp stability boundaries. This filtering process

results in high-resolution mass peaks obtained with scan lines passing near the tip of the stability region. However, all practical instruments have finite length rods, and consequently, only a finite number of RF cycles to reject unwanted ions. If n is too low, a significant number of unwanted ions remain unfiltered, contaminating adjacent mass peaks. When the resolution is limited by the number of RF cycles, rather than rod quality or other factors, the maximum attainable mass resolution of a QMS scales as n^2 (*Dawson* 1997).

In order to compare the performance of different rod sets, the n-values of ions should be well matched. For a given quadrupole, l is fixed, and the operating frequency Ω is set as described in Section 4.2.1. Therefore, from equation (4.4) the ion axial energy (E) must be adjusted in order to obtain a given n-value. The available range of ion axial energies is typically set by the design of the ion source. However, the quadrupole rods themselves can be biased with a potential with respect to the ion source exit aperture. The final ion axial energy (E) is approximated by the potential difference between the rods (ϕ_{b}) and the exit aperture of the source (ϕ_{s}) subtracted from the ion energy produced by the source, ($E_{\mathrm{i,s}}$), such that,

$$E \approx E_{\mathrm{i,s}} - (\phi_{\mathrm{b}} - \phi_{\mathrm{s}}). \tag{4.5}$$

The approximation of equation (4.5) neglects fringing field and field penetration effects between the rods and the exit aperture of the source, as well as the spread of produced ion energies.

For the circular rod set, ions were created with energies approximately 80±5 eV. A +70V to +40V rod bias range is used to obtain ions between n ≈35 and n ≈70 RF cycles. For the hyperbolic rod set, ions were created at approximately 1±0.5eV. Therefore, a rod bias between -2V and -16V is used to obtain ions between n ≈35 and n ≈80 RF cycles. The best results are expected for energies that yield the maximum available n, leading to nominal rod biases of +70V and -2V for the circular and

hyperbolic QMSs respectively, as indicated in Table 4.1.

4.2.3 Matching of dynamic range

Although not directly related to any of the dimensionless quantities, a, q, or ξ, the instrument sensitivity is an important characteristic of QMS sensors that determines the effective dynamic range of the measured mass spectra signal. Determination of the total instrument sensitivity is a complex task as it will be a function of detector type, efficiency, accumulation time, ion source design, and the partial pressures of measured gases (*Dawson* 1997). Therefore, although the measurement timing (Table 4.1) for each sensor differs slightly due to the various requirements of each instruments interface software, the pressure of the krypton gas inside the ion source was adjusted to obtain similar counting statistics. For the circular rod QMS, the source pressure was set to 1×10^{-5} mbar. For the hyperbolic rod sensor, the source pressure was set to 1×10^{-7}mbar. These pressures resulted in measurements of Kr+ isotopes ranging over approximately 4 orders of magnitude in detected particle flux for both instruments.

4.3 Results and discussion

With these well-matched systems, isotopes of Kr+ in the (m/z) range 78 to 86 amu/e were measured by each sensor with and without applied higher order auxiliary excitation. These measurements were analyzed in terms of mass resolution, peak variability, and variation with ion energy, i.e., n. Figures 4.1 and 4.2 show the stability regions formed from using q'=0.00 (i.e., no excitation), $q' \approx$0.10, and $q' \approx$0.15 for the hyperbolic and circular rod sensors, respectively with corresponding voltage scan lines. The voltage scan line of the non-excitation measurements was set such that it passes through the tip of the stability region for a given ion. The voltage scan line of the excitation measurements was set such that it passes through the center of a formed upper stability island. Similar island splitting is shown for both sensors. The

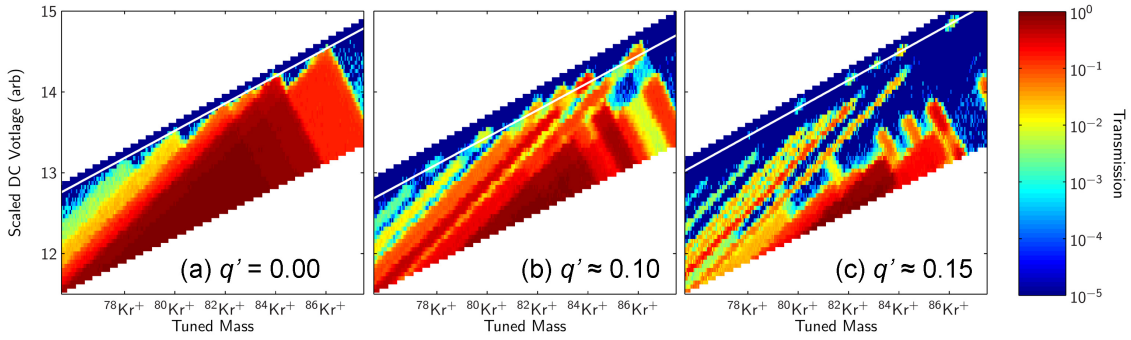

Figure 4.1: Stability island formation for the hyperbolic QMS (n ≈80) with (a) no added excitation (q'=0.00), (b) q' ≈0.10, and (c) q' ≈0.15. The scan line used to obtain high resolution measurements is shown on each. For the non-excitation case, i.e. nominal QMS operation, the scan line passes through the tip of the stability region. For the excitation cases, the scan line slope is tuned such that it passes through the center of the formed stability islands.

hyperbolic rod measurements without excitation create sharper stability boundaries than those of the circular rod sensor, as expected.

The scalar conversion from RF voltage to m/z will change depending on instrument operating point. The appropriate conversion from RF voltage to m/z is selected as the factor that aligns the maximum transmission for ^{86}Kr+ between m/z = 85.75 amu/e and m/z = 86.25 amu/e for the lowest ion energy setting for each sensor.

4.3.1 Mass resolution

Figures 4.3 and 4.4 show 10 mass scans of the krypton mass range for each QMS with and without applied excitation. The mean of each set of scans is indicated as the solid black line. The 1% mass resolution ($R_{0.01}$) for mass peak ^{86}Kr+ is indicated for each case, where $R_{0.01}$ is defined as $(m/z)/(\Delta(m/z))$ where m/z is the measured mass per charge ratio and $\Delta(m/z)$ is the width of the mass peak that is within 1% of its maximum transmission value. This value is not computed for all mass peaks as their relative proximity to one another and measured dynamic ranges may not make

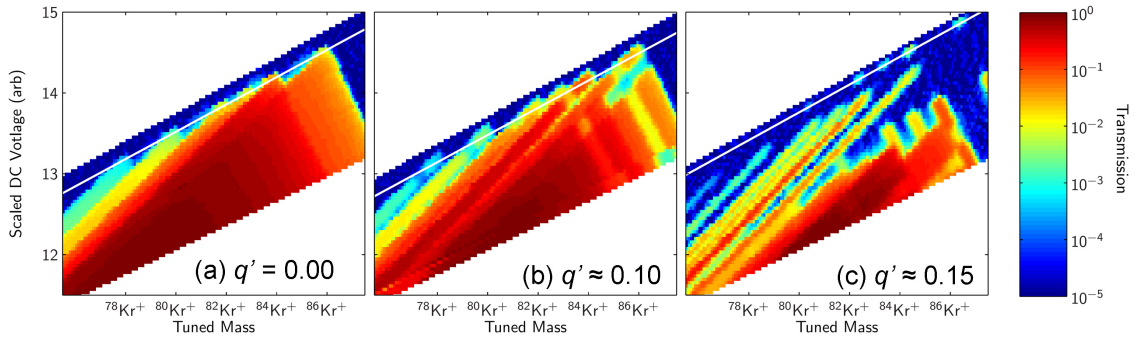

Figure 4.2: Same as Fig. 4.1 but for the circular rod QMS with n ≈70. Similar stability regions are formed but with less sharp boundaries for the non-excitation case, consistent with the low-mass peak tail expected from the circular rod sensor.

such a computation possible in all cases, specifically for the circular rod set operated without auxiliary excitation. The higher resolution scans were set to have a slope that passes through the very tip of the stability region, at the limit of instrument performance.

With added excitation, the circular rods show a factor of two improvement in the $R_{0.01}$ of ^{86}Kr+, consistent with previous measurements of Ar+ using a similar experimental setup (*Gershman et al.* 2011). For the non-excitation scans, the observed $R_{0.01}$ of the circular rods is reduced by the low mass peak tail, which arises as a consequence of imperfections in the quadrupole electric field resulting from non-ideal rod geometry. Auxiliary excitation techniques increase rejection efficiency of unwanted ions near these boundaries, removing this peak tail. However, this tailing effect is not observed in the hyperbolic QMS measured peaks, even for the non-excitation case. Consequently, the $R_{0.01}$ value for the hyperbolic rod set remains unchanged with added excitation, though the highest possible mass resolution for the hyperbolic rod system was slightly larger than that of the circular rod system. These results are consistent with predictions from numerical simulations (*Gershman et al.* 2011; *Konenkov et al.* 2005) that show for ideal quadrupole geometry the overall mass

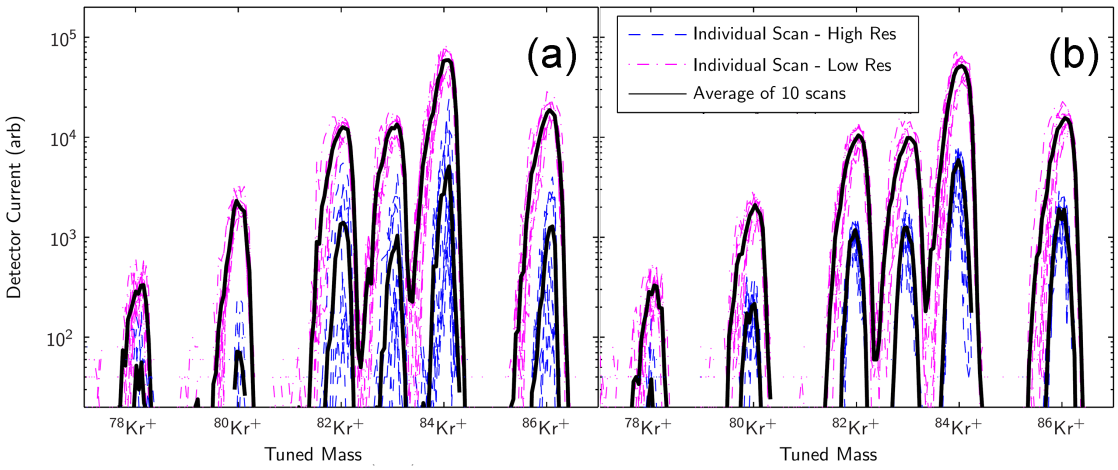

Figure 4.3: Measured peak shapes of Kr+ (a) without and (b) with higher order auxiliary excitation for the hyperbolic rod QMS for n ≈80.. The solid black curves are averages of sets of ten successive mass scans, with the pink dash-dotted curves corresponding to a set of lower resolution, stable measurements and the blue dash curves corresponding to the maximum attainable mass resolution with the given test electronics. The lower and higher resolution measurements in (b) use auxiliary excitation levels of q' ≈0.10 and q' ≈0.15, respectively. While the maximum attainable mass resolution does not increase with applied excitation, the peak stability of the high resolution measurements significantly improves.

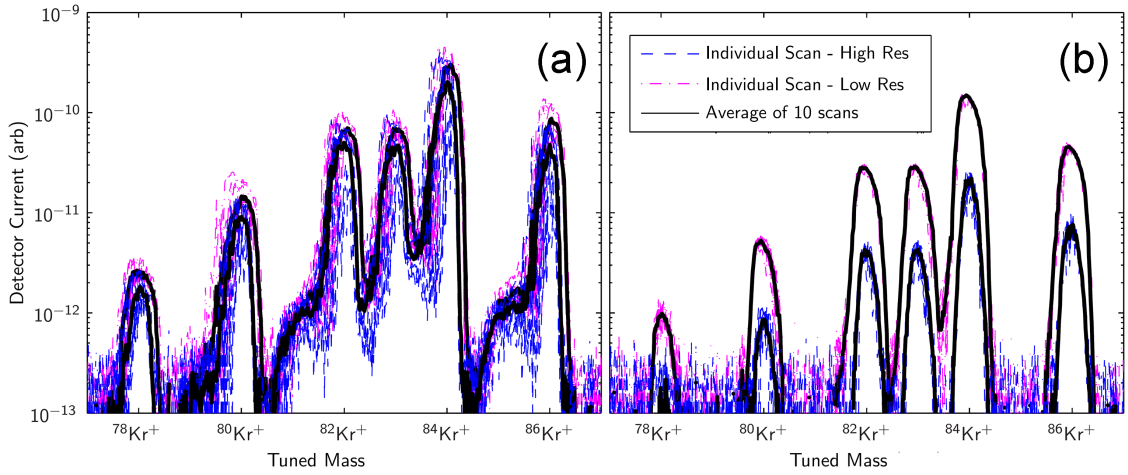

Figure 4.4: Same as Fig. 4.3 but for the circular rod QMS with n ≈70.. Here, not only does the peak stability significantly improve with applied excitation, but the low-mass peak tail characteristic of circular rod sensors is removed, effectively doubling the maximum attainable mass resolution of the instrument.

resolution for a particular ion transmission is only expected to improve for higher ($R \sim$1000) resolution measurements.

4.3.2 Peak variability and stability

Electronics noise causes small shifts in operating point during a mass scan. In the traditional mode of quadrupole operation, for lower resolution scans, these shifts will not create large changes in mass peak shapes. However, for higher resolution scans, when operating scan lines through the tip of stability regions, the mass peak shapes become very sensitive to small changes in operating point. This effect can be observed in the non-excitation peaks of both the hyperbolic and circular rod systems, which become increasingly variable and unstable at high resolution. Scan line operation through the center of an upper stability island, however, mitigates this effect. These islands represent large, symmetric areas of almost constant transmission and resolution. Therefore, small shifts in operating point do not result in significant changes to the mass peak shapes, resulting in more stable peak shapes for both rod geometries, as shown in Figures 4.3 and 4.4.

The accuracy of recovered isotopic ratios of krypton is used to quantify this variability. The 10% transmission, $T_{0.1}$, is the average transmission between the two points defining the 10% resolution of a mass peak. This value was determined for each of the mass peaks corresponding to singly charged krypton isotopes, ^{78}Kr+, ^{80}Kr+, ^{82}Kr+, ^{83}Kr+, ^{84}Kr+, and ^{86}Kr+, and was normalized by the total $T_{0.1}$ for all measured masses, giving the measured relative abundance of each isotope for each scan. The mean of the derived relative abundances of the 10 scans are shown in Figure 4.5 and are compared to the standard literature values (*Lide* 2004). The error bars in Figure 4.5 indicate the standard deviation of the 10 scans, a proxy for the variability of the mass peak shapes. The measurements of ^{78}Kr+ had a very low signal to noise ratio for both the circular rod and hyperbolic rod higher resolution scans, resulting

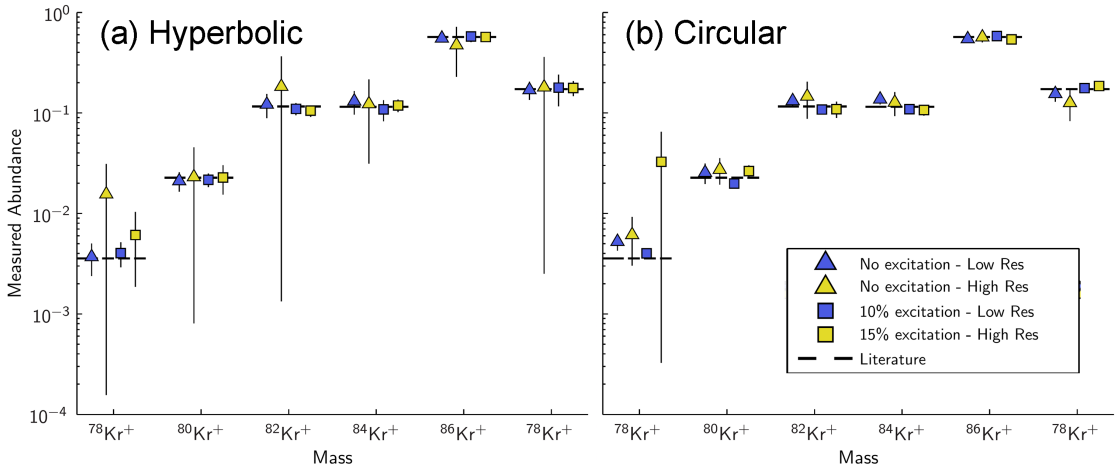

Figure 4.5: (a) Measured relative abundances of krypton isotopes averaged over 10 scans of the hyperbolic rod QMS compared with accepted literature values (*Lide* 2004). The average $T_{0.1}$ of each mass scan is indicated by the shaped marker. The standard deviation, i.e. variability, of $T_{0.1}$ is represented by the error bars around each marker. (b) same as (a) but for the circular rod QMS. For all cases, the peak variability at the limit of the test electronics (high resolution modes) significantly improves with applied excitation.

in misleadingly high calculations of variability. For other peaks, counting statistics result in pointwise errors on the order of 10% of the peak value. The transmission for both the excitation and non-excitation curves are similar, so counting statistical error would have a similar effect on both sets of peaks, indicating that the increase in stability is due to the use of auxiliary excitation. The 1% transmission, $T_{0.01}$, was not used for this analysis due to insufficient separation of mass peaks for the case of circular rods operated without excitation.

The extremely high variability of the non-excitation high resolution hyperbolic rod mass scans is likely a result of operation at a sharp tip of the stability region, where small changes in operating point create drastic changes in transmission combined with the limitations of our interfaced electronics as listed above. This effect is likely less pronounced in the circular rod system since the tip of this region is blunted by non-linear resonant effects (*Schulte et al.* 1999).

While the average measured abundances do not significantly change with excitation, the variability of each scan is significantly reduced for high-resolution operation with applied excitation. This reduction in variability directly translates into a shorter required scanning time for instrument high resolution scans. For example, the average high resolution variability of the hyperbolic rods with applied excitation was 6 times lower than the non-excitation case, leading to a 6 times shorter required scan time. Likewise, the circular rod high resolution scanning time could be reduced by a factor of 2. This increased peak stability is consistent with previous experimental work, as well as predictions from numerical simulations (*Gershman et al.* 2011; *Konenkov et al.* 2005).

4.3.3 Ion rejection efficiency

From both numerical simulations (*Konenkov et al.* 2005) and previous experimental work (*Konenkov et al.* 2001), the number of RF cycles required for sufficient rejection of undesired ions is expected to decrease when operating with applied excitation, i.e. the rejection efficiency should increase, and consequently the attainable resolution for a given ion energy (*Dawson* 1997). This improvement is likely the reason for the improved maximum attainable mass resolution in Section 4.3.1 for the circular rod quadrupole. The maximum attainable resolution was achieved at the limit of the test electronics. In order to characterize the effect of excitation on mass peaks independent of electronics precision effects, the rejection efficiency is measured using the lower resolution instrument settings from Section 4.3.1 (i.e. the blue curves from Figures 4.3 and 4.4).

As discussed in Section 4.2.2, the number of RF cycles an ion spends traversing the rods can be adjusted by changing the QMS rod bias. The ranges of rod biases from Section 4.2.2 were scanned for each system with and without applied excitation. Figures 4.6 and 4.7 show the effect of changing the rod bias on mass peak shapes

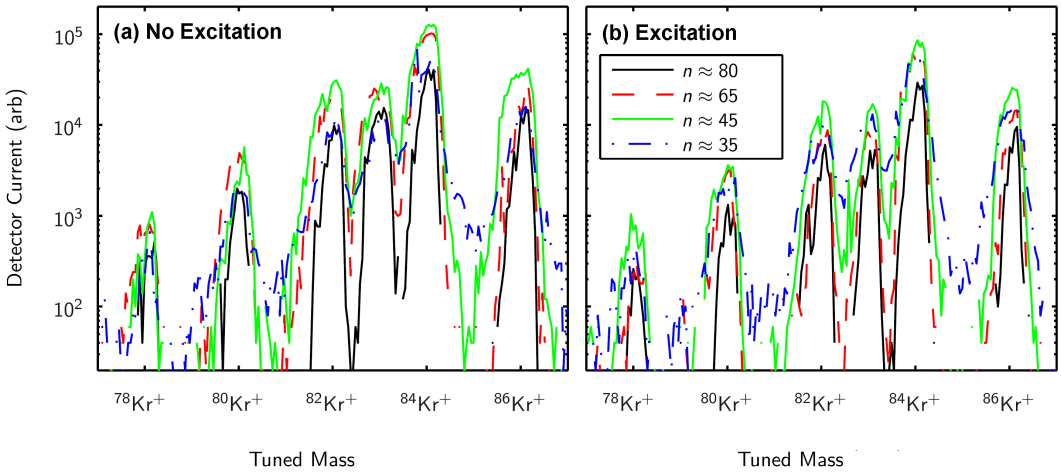

Figure 4.6: Peak shapes for Kr+ measured with the hyperbolic rod QMS for various separation times for (a) no applied excitation and (b) $q' \approx 0.10$ as in Fig. 4.3. With decreasing number of RF cycles, n, the measured mass peaks exhibit tailing on both the low-mass and high-mass sides. Such tailing is mitigated with applied excitation.

for each sensor. For reduced n, for both the hyperbolic and circular rod systems, significant peak tailing occurs on the low mass side of the Kr+ mass peaks. However, this effect is less pronounced for operation with applied excitation.

Because of insufficient separation between adjacent mass peaks, it is difficult to directly calculate the changes in mass resolution of Kr+ without fitting models of the mass peaks to the observed spectra. Instead, to quantify these effects, the crosstalk of ^{85}Kr+ relative to the transmission of the true ^{86}Kr+ peak is calculated. ^{85}Kr+ does not correspond to any known naturally occurring stable isotope of krypton (*Lide* 2004). Therefore any transmission between m/z = 84.75 amu/e and m/z = 85.25 amu/e is considered to be mass crosstalk. This transmission, T^{85}, is normalized by the transmission of the adjacent peak, ^{86}Kr+, between m/z = 85.75 amu/e and m/z = 86.25 amu/e, and plotted in Figure 4.8 for both rod sets.

The ^{85}Kr+ mass crosstalk in the non-excitation circular rod measurements is high due to the significant low mass peak tail that is present even for large values of n. This

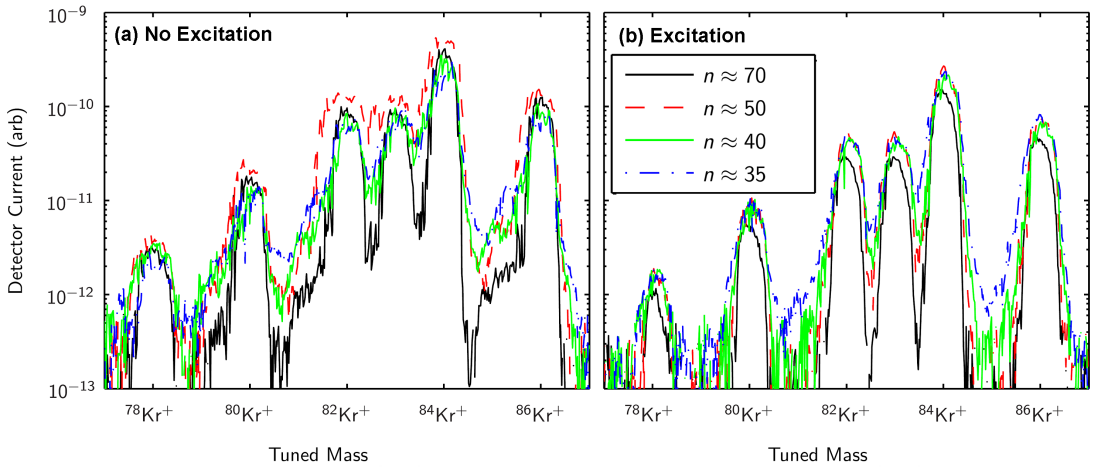

Figure 4.7: Same as Fig. 4.6 but for circular rod QMS using operating points from Fig. 4.4. Just as for the hyperbolic rod QMS, the increased peak tailing as a result of reduced n is mitigated with applied excitation, including the peak tail that appears at all n-values as a consequence of circular rod geometry.

tail is eliminated with the use of auxiliary excitation for all ion energies. Consequently, the relative reduction of mass crosstalk due to auxiliary excitation (about an order of magnitude for all n) is highest for the circular rod system. However, even for the hyperbolic rod set, the relative reduction when using auxiliary excitation is at least a factor of two and up to a factor of 20.

As seen in Figure 4.8, with added excitation, both QMS sensors require only approximately 67% of the n-value to obtain performance comparable to the non-excitation cases. Lowering n has two major implications for QMS sensors: (1) the quadrupole rods themselves can be shortened to create a smaller, lighter instrument without sacrificing performance and (2) higher energy ions may be analyzed, simultaneously relaxing design constraints on the ion source and creating an increased flux of particles at the detector, thereby increasing instrument sensitivity and enabling shorter measurement times.

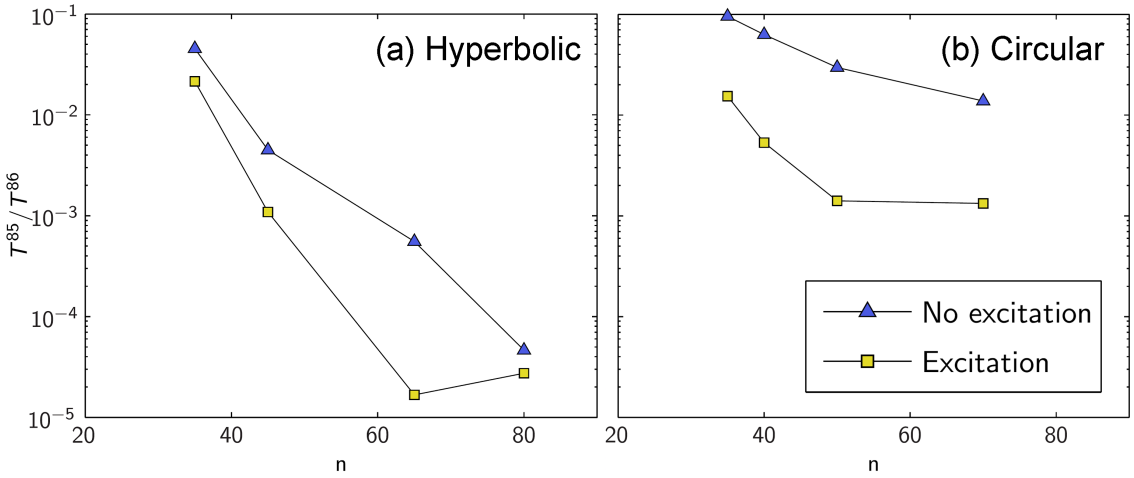

Figure 4.8: Crosstalk of ^{85}Kr+ as a function of separation time for (a) hyperbolic and (b) circular rod QMS sensors with and without applied excitation for the measured mass peaks from Figures 4.6 and 4.7. The crosstalk is reduced by factors of 2-20 for both measurement geometries.

4.4 Conclusions

High order auxiliary excitation significantly improved the performance of the Pfeiffer QMG422 commercial circular rod quadrupole in all areas, doubling the maximum attainable $R_{0.01}$ for the given test electronics, providing a two-fold increase in peak stability and variability, and improving ion rejection efficiency by an order of magnitude. Although the maximum attainable $R_{0.01}$ of the MSL/SAM hyperbolic QMS system did not improve with applied auxiliary excitation, six and two-fold improvements were obtained for peak stability and ion rejection efficiency, respectively. These results are all consistent with previous experimental work and numerical simulations related to auxiliary excitation techniques. Consequently, the results of these simulations, specifically the improved predicted performance of hyperbolic rod sensors for high resolution modes are expected to hold for sensors employing higher precision electronics than those used in this study. One of the most significant improvements for ideal geometry sensors is that of enhanced peak stability obtained from operating voltage scan lines through the center of formed upper stability islands. As of this

work, the only power efficient way of implementing such a technique at high resolutions is to use higher order auxiliary excitation in conjunction with a multiresonant tank circuit.

Based on the derived resolution of these measurements, it is difficult to distinguish between mass peaks obtained from different rod geometries when auxiliary excitation is applied. This similarity in observed performance is even more remarkable when considering the quality of the respective ion sources used for each sensor. The hyperbolic rod ion source is one of the most precise and stable of its kind, with the circular rod QMS using a standard, commercially available ion source with a comparatively wide entrance aperture and ion beam size, factors that significantly affect the performance of the sensor in its traditional mode of operation. Higher order auxiliary excitation techniques are clearly a powerful tool for mitigating effects of imperfections in a quadrupole system, whether they arise from fringing fields, non-linear resonances, or wide ion energy ranges. These potential improvements to circular rod systems could have tremendous impact on the next generation of quadrupole sensors, as circular rod QMS sensors could be created that are suitable for high resolution spaceborne or laboratory applications. We note, however, that the highest attainable mass resolution was measured on the hyperbolic rods operating with auxiliary excitation. We therefore suggest that these techniques could be used to improve the stability and performance of existing hyperbolic sensors, extending their practical mass ranges, and enabling their use for future planetary and cometary missions.

CHAPTER V

Modeling extreme ultraviolet suppression of electrostatic analyzers

This chapter is taken from: Gershman, D. J. and T. H. Zurbuchen (2010), Modeling extreme ultraviolet suppression of electrostatic analyzers, *Rev. Sci. Instrum., 81*, 045111.

Abstract

In addition to analyzing energy-per-charge ratios of incident ions, electrostatic analyzers (ESAs) for spaceborne time-of-flight mass spectrometers must also protect detectors from extreme ultraviolet (EUV) photons from the Sun. The required suppression rate often exceeds $1 : 10^7$ and is generally established in tests upon instrument design and integration. This paper describes a novel technique to model the EUV suppression of ESAs using photon ray tracing integrated into SIMION, the most commonly used ion optics design software for such instruments. The paper compares simulation results with measurements taken from the ESA of the Mass instrument flying onboard the Wind spacecraft. This novel technique enables an active inclusion of EUV suppression requirements in the ESA design process. Furthermore, the simulation results also motivate design rules for such instruments.

5.1 Introduction

Time-of-flight mass spectrometers (TOF-MS) have become a standard for measuring energetic charged particles in various space environments. Since the Suprathermal Energy Ionic Charge Analyzer (SULEICA), the first spaceborne time-of-flight instrument, was launched on the AMPTE IRM spacecraft in 1984 (*Möbius et al.* 1985b), TOF-MS have flown on almost every mission to a planetary body in our solar system. This technology has enabled unprecedented observations of plasma distribution functions in planetary ionospheres (*Young et al.* 2004; *Andrews et al.* 2007; *Williams et al.* 1992) and provided the first isotopic composition measurements of the solar wind (*Gloeckler et al.* 1995b, 1992, 1998).

The frontend of these instruments, the electrostatic analyzer (ESA), serves to filter out ions with a specific energy-per-charge ratio. For a given voltage setting, shaped electrodes analyze incident ions, as indicated in Fig. 5.1, which as an example, sketches the overall function of the WIND/MASS instrument as described in detail by *Gloeckler et al.* (1995b). The MASS ESA is a spherical analyzer; each plate consists of a portion of a sphere mounted in a concentric fashion. Figure 1 also demonstrates that this ESA is combined with additional subsystems to enable mass-resolved measurements of solar wind ions through determination of the time-of-flight of ions in a harmonic electrostatic potential. This TOF measurement is preceded by a post-acceleration voltage of -10kV and then consists of coordinated measurements of a start-signal from secondary electrons that are emitted from a very thin carbon foil and deflected onto a microchannel plate (MCP), and a stop-signal from the direct incidence of the measured ion onto another MCP, a technique developed by *Gloeckler and Hsieh* (1979). Other spaceborne mass spectrometers may have varying shapes and forms. For example, some plasma instruments use MCPs entirely without TOF

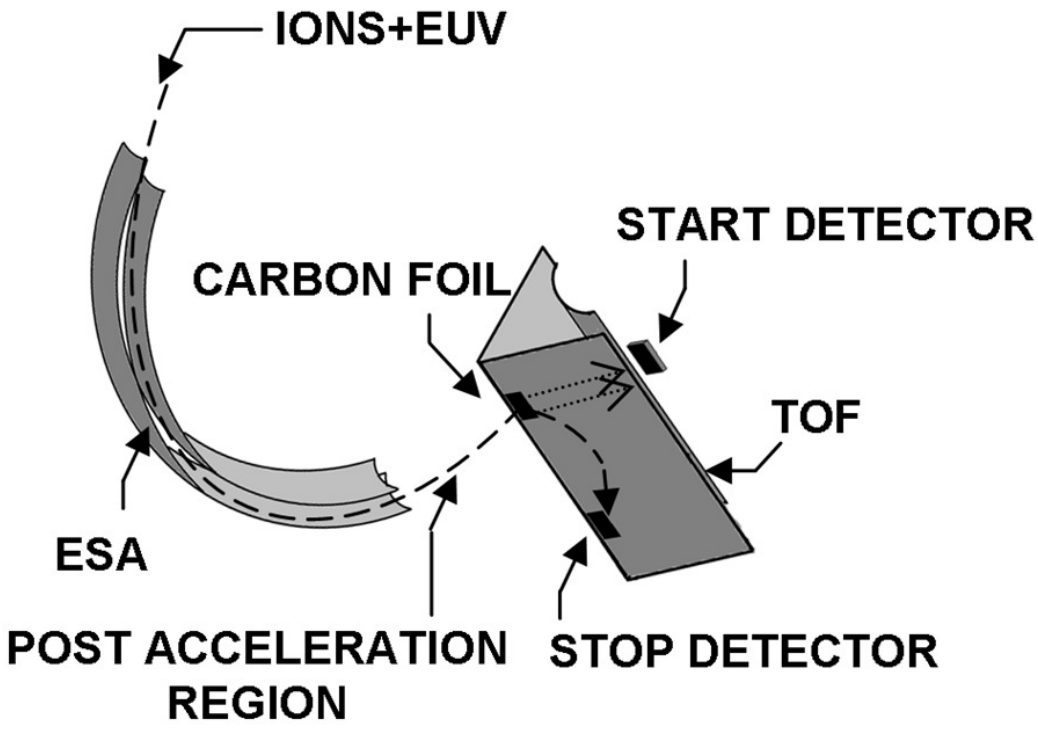

Figure 5.1: Schematic view of the MASS instrument on the WIND spacecraft adapted from *Zurbuchen et al.* (1995). An electrostatic deflection system is combined with a time-of-flight unit with a carbon foil and two microchannel plate detectors which are sensitive to both charged particles and EUV radiation.

measurements, (*McComas et al.* 1998a) and some may include solid-state detectors (SSDs) (*Gloeckler et al.* 1998).

There are two important sources of background caused by EUV which limit the signal to noise in all instruments. First, EUV photons that penetrate the ESA can hit the carbon foil and liberate an electron, resulting in an erroneous ion start-signal (*Hsieh et al.* 1980). Second, EUV photons can hit the radiation detectors directly, resulting in an erroneous ion stop-signal (*Gruntman* 1997). It is of note that a good ion time-of-flight measurement is a double or even triple coincidence (i.e., start, stop, and solid-state detector energy) whereas a photon can cause only a single count. While it can sometimes be possible to determine which count in a particular timing window is

erroneous due to unphysical times of flight, a photon striking the detector at a similar time to either the start electron or the ion-stop signal can cause ambiguity in that particular measurement and thus substantially affect the signal to noise performance ratio, one of the key quality metrics of any ion sensor. Visible light, despite being substantially more abundant in flux than EUV, does not have the energy to release electrons from carbon foils or to generate photo-electrons for amplification in MCPs (*Paresce* 1975). While both visible light and EUV can generate currents in SSDs (*Korde and Geist* 1987), a relatively constant photon flux will lead to an increased SSD noise floor, as opposed to a false pulse as generated in MCPs. Thus, in addition to driving requirements of the ion-optics performance of a given space-instrument, ESAs also have to suppress EUV.

To develop such a requirement, consider a specification of radiation suppression in WIND/MASS that is flying near the Lagrangian point L1 near Earth (*Acuña et al.* 1995). MASS was designed to measure the solar wind heavy ion isotope composition. The total solar wind particle flux at 1AU is approximately $3 \times 10^8 \mathrm{cm}^{-2} s^{-1}$ (*Schwenn and Marsch* 1990) . Heavy ions such as silicon and iron have abundances relative to H^+ on the order of 10^{-5} with certain isotopes one or two orders of magnitude smaller (*Bame et al.* 1975; *Wiens et al.* 2004). The geometric factor specification of MASS is $0.35\mathrm{cm}^2$, resulting in tens of counts per second for minor species isotopes (*Gloeckler et al.* 1998).

In order to obtain an EUV induced noise count levels of an order of magnitude below that of the signal of minor species measurements, erroneous measurements must be limited to a few counts per second. A typical solar EUV flux at 1AU is about 4×10^{11}photons cm^{-2} s^{-1}, which enter the MASS aperture (*Bauer and Lammer* 2004). Upon passing through the electrostatic system, photons can impact the MASS carbon foil of thickness of $2\mu\mathrm{g}$ cm^{-2} and release electrons with a probability of 10^{-3}, triggering a false MCP signal (*Gloeckler et al.* 1998; *Hsieh et al.* 1980). Therefore, to

limit the contribution of EUV photons to only a few counts per second, the MASS ESA suppression to the carbon foil ($\Phi_{\text{foil}}/\Phi_{\text{in}}$) is specified to be better than 10^{-8} where Φ_{in} is the flux of photons entering the ESA and Φ_{foil} is the flux of photons incident on the foil. For MASS, this was achieved using a rough black surface applied to the ESA. For the Solar Wind Ion Composition Spectrometer (SWICS), this was achieved by using a combination of blackening, serrations, and a light-trap design (*Gloeckler et al.* 1992). SWICS has an estimated background suppression factor better than 10^{-12}.

Regardless of the technique employed for meeting EUV suppression requirements, the process by which photon flux is reduced is either absorption or scattering away from the instrument's radiation sensitive parts. The probability of either of these cases occurring significantly increases each time a photon reflects off of a surface inside the ESA. Generally, an experimentally established empirical, 'three-bounce-rule' is used. Apparently, sufficient suppression is achieved if transmitted photons must undergo at least three scattering events under any angle before reaching any sensitive parts of the instrument (*Zurbuchen et al.* 1995; *Moore et al.* 2000; *Herrero* 1992). This rule can be directly tested using CAD programs, but modern, more challenging ESA geometries (*Andrews et al.* 2007) are difficult to analyze completely, or, they make an idealized adherence to the 'three-bounce-rule' impossible. In addition, this simple analysis does not take into account the reflective properties of any particular material used for such sensor designs.

Here we describe a novel technique which allows for the simultaneous automation of photon trajectory creation through a Monte-Carlo process and the incorporation of the material detailed reflective properties in the determination of the EUV suppression properties of an ESA. This method is implemented into a software package, SIMION, which is already used for most ion-optics designs. Our method is rigorous in its application, but also provides unprecedented and integrated EUV suppression

analysis for spaceborne electrostatic systems during the design process where EUV suppression requirements may be dominant. Section 5.2 will discuss the methodology and algorithms used in this application and Section 5.3 describes the detailed implementation into SIMION. Section 5.4 then describes the application of this new method to WIND/MASS, and Section 5.5 summarizes the key results of these calculations and the design rules that can be deduced from them.

5.2 Photon Tracing in SIMION

In order to adequately simulate the EUV suppression of an ESA, a scene representative of a three-dimensional instrument model must be rendered accurately. Many rendering techniques exist in the field of computer graphics, but perhaps the most realistic results can be obtained from rendering photons one at a time while maintaining the statistically correct behavior of each photon, a process known as Monte Carlo ray tracing (*Cook et al.* (1984)). This method, in conjunction with appropriate mathematical modeling of the material interaction in a scene, produces very successful results (*Cook et al.* 1984; *Whitted* 1980). The intense computational cost of tracing individual photons in a scene has traditionally limited the use of photon ray tracing for real-time rendering. However, this method can be a powerful tool to effectively predict the EUV transmission and response of electrostatic analyzers since results are not needed in real-time. Other rendering techniques such as photon mapping (*Jensen* 1996) or radiosity (*Goral et al.* 1984) do not typically accurately model the details of specular reflection, which is often an important contribution to the background of transmitted EUV in ESAs and are hence not suitable for our needs.

In this paper, we introduce a photon tracing process which relies on an existing framework of particle optics in SIMION, commonly used to simulate ion optics properties of instrument models. SIMION uses finite-differencing techniques to solve the Laplace equation in three dimensions, and uses the resulting electric fields to guide

particle trajectories. Geometries in SIMION are defined by potential arrays, which assign boundary potentials to specified points corresponding to instrument electrodes. To utilize SIMION for photon tracing applications, neutral particles are used which are unaffected by electric fields. A local surface normal is computed from the electric field vector at the location of each particle-electrode collision. Mathematical models of curved electrodes can be used to aid the calculation of surface normals, which can be difficult in SIMION due the electric field only being known in the frame of reference of each simulated particle. Higher resolution two-dimensional models of the same instrument can more accurately determine local normal vectors of curved surfaces, taking advantage of symmetries which may exist in the instrument geometry.

5.3 Scattering and Reflection Model

Once general photon ray tracing capabilities and geometry determinations are built into software, a scattering model must be developed. There are two critical ingredients for the analysis of scattered particles in an ESA. First, the geometry of all scattering members of the scene need to be accurately modeled. This scene includes the electrostatic analyzer plates, insulators, and also structural members of the instrument which may not be part of the electrostatic design, but which can contribute to the scattering of radiation in these systems. Second, the accurate scattering properties of each surface have to be known. The absorption and reflection ratios are determined by the electromagnetic properties of each surface, (*Jackson* 1998) as well as the statistical properties of the surface structure (*Ogilvy* 1991). Inelastic effects do not generally play a significant role in our applications.

5.3.1 Scattering Angular Distribution

The angular distribution of the statistical scattering of a given incident photon can be measured by the bireflectance distribution function (BRDF) of that surface

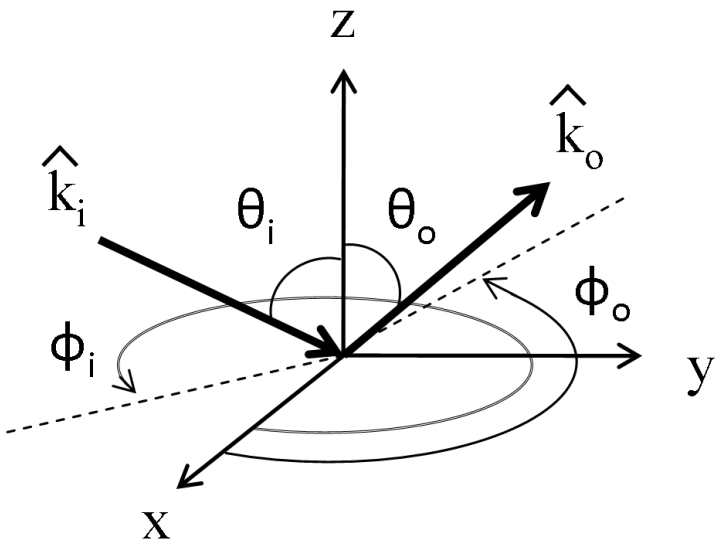

Figure 5.2: Definition of incident and reflected photon directions. $\theta_{\rm i}$ and $\phi_{\rm i}$ measure the photon incident direction; $\theta_{\rm o}$ and $\phi_{\rm o}$ measure the photon scattered direction. $\hat{k_{\rm i}}$ and $\hat{k_{\rm o}}$ are unit vectors corresponding to the incident and scattered directions respectively. The z-axis is defined as the unit vector normal to the scattering surface.

for a given frequency. The BRDF relates the probability of scattering for a given incidence direction ($\theta_{\rm i}$,$\phi_{\rm i}$) and scattering direction ($\theta_{\rm o}$,$\phi_{\rm o}$), as defined in Fig. 5.2.

Using the definitions in Fig. 5.2, the BRDF is defined as the differential output radiance ($dI_{\rm o}$) over the differential input irradiance ($I_{\rm i}cos(\theta_{\rm i})$) per unit solid angle ($sin(\theta_{\rm i})d\theta_{\rm i}d\phi_{\rm i}$), (*He et al.* 1991)

$$BRDF(\theta_{\rm i},\phi_{\rm i},\theta_{\rm o},\phi_{\rm o}) \equiv \frac{dI_{\rm o}}{I_{\rm i}cos(\theta_{\rm i})sin(\theta_{\rm i})d\theta_{\rm i}d\phi_{\rm i}} \tag{5.1}$$

Rearranging equation (5.1) and integrating yields the well known 'rendering equation' (*Kajiya* 1986), which relates the incident and scattering light intensity.

$$I_{\rm o} = \int_0^{2\pi} d\phi_{\rm i} \int_0^{\pi/2} d\theta_{\rm i} BRDF(\theta_{\rm i},\phi_{\rm i},\theta_{\rm o},\phi_{\rm o})\ I_{\rm i}\ cos(\theta_{\rm i})sin(\theta_{\rm i}) \tag{5.2}$$

Consider a single photon incident from a source direction. The incoming radiance

is generally modeled as scaled delta function in spherical coordinates.

$$I_{\mathrm{i}} = \frac{I_{\mathrm{s}}\delta(\theta_{\mathrm{i}} - \theta_{\mathrm{s}})\delta(\phi_{\mathrm{i}} - \phi_{\mathrm{s}})}{sin(\theta_{\mathrm{i}})} \tag{5.3}$$

Here, I_{s} is the magnitude of the incoming radiance and θ_{s} and ϕ_{s} define the direction of the incident light source. When integrating radiance over all possible incident directions $(\theta_{\mathrm{i}}, \phi_{\mathrm{i}})$, only the photon source direction $(\theta_{\mathrm{s}}, \phi_{\mathrm{s}})$ makes a contribution.

Substituting equation (5.3) into the integral in equation (5.2) and dividing both sides by $I_{\mathrm{s}}cos(\theta_{\mathrm{s}})$ yields,

$$BRDF(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta_{\mathrm{o}}, \phi_{\mathrm{o}}) = \frac{I_{\mathrm{o}}}{I_{\mathrm{s}}cos(\theta_{\mathrm{s}})} \tag{5.4}$$

I_{o} is the amount of output radiation energy per projected unit area. Therefore, multiplying I_{o} by a factor of $cos(\theta_{\mathrm{o}})$ yields the amount of total light energy reflected from a surface. Likewise, the ratio $\frac{I_{\mathrm{o}}cos(\theta_{\mathrm{o}})}{I_{\mathrm{s}}cos(\theta_{\mathrm{s}})}$ is the ratio of scattered light energy to incident light energy. In the case of a single photon reflecting off of a surface, this ratio represents the probability of a photon scattering in a particular direction. Therefore, multiplying the BRDF by a factor of $cos(\theta_{\mathrm{o}})$ forms a four dimensional probability distribution for the scattering of an incident photon.

However, each scattering event carries with it not only a change in direction, but a probability that a photon is absorbed, effects which are both implicitly included in the BRDF. In our application, the BRDF is used to determine the scattering direction of an individual photon rather than the total scattered radiance in all directions. By using the BRDF in this manner, the probability of photon absorption is lost.

Since photon absorption and surface scattering should be independent processes, we model the BRDF as the product of two probability distribution functions,

$$BRDF(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta_{\mathrm{o}}, \phi_{\mathrm{o}}) = \rho(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta_{\mathrm{o}}, \phi_{\mathrm{o}}) \times BRDF_{\mathrm{s}}(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta_{\mathrm{o}}, \phi_{\mathrm{o}}) \tag{5.5}$$

where $BRDF_{\mathrm{s}}$ is the scattering distribution function of interest and ρ is the surface reflectivity whose contribution must be removed.

The scattering distribution can be normalized by integrating $BRDF_{\mathrm{s}}cos(\theta)$ over angles in the upper hemisphere as defined in Fig. 5.2. It is important to note, however, that since θ and ϕ are considered variables of a probability distribution rather than of solid angle, the integration of interest is over $d\theta d\phi$ rather than $sin(\theta)d\theta d\phi$.

Probability distributions $\mathrm{P}(\theta_{\mathrm{o}})$ and $\mathrm{P}(\phi_{\mathrm{o}})$ can be formed by integrating $BRDF_{\mathrm{s}}cos(\theta)$ over one variable (θ_{o} or ϕ_{o}) and normalizing:

$$P(\theta_{\mathrm{o}}, \theta_{\mathrm{s}}, \phi_{\mathrm{s}}) = \frac{\int_0^{2\pi} BRDF_{\mathrm{s}}(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta_{\mathrm{o}}, \phi')cos(\theta_{\mathrm{o}})d\phi'}{\int_0^{2\pi}\int_0^{\pi/2} BRDF_{\mathrm{s}}(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta', \phi')cos(\theta')d\theta' d\phi'} \tag{5.6}$$

$$P(\phi_{\mathrm{o}}, \theta_{\mathrm{s}}, \phi_{\mathrm{s}}) = \frac{\int_0^{\pi/2} BRDF_{\mathrm{s}}(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta', \phi_{\mathrm{o}})cos(\theta')d\theta'}{\int_0^{2\pi}\int_0^{\pi/2} BRDF_{\mathrm{s}}(\theta_{\mathrm{s}}, \phi_{\mathrm{s}}, \theta', \phi')cos(\theta')d\theta' d\phi'} \tag{5.7}$$

Using these distribution functions in conjunction with randomly generated numbers will yield values of scattering angles θ_{o} and ϕ_{o} for a given incident direction defined by θ_{s} and ϕ_{s}. Note that Eqs. (5.6) and (5.7) can be individually applied only if θ_{o} and ϕ_{o} are independent of one another. After each photon scattering direction is chosen, the ρ function is applied to determine the probability that the photon was absorbed by the surface.

5.3.2 Surface BRDF Modeling

A BRDF creates a probability distribution which dictates the angular scattering from incident photons on a surface. However, an appropriate BRDF must be chosen for each modeled surface in order to obtain realistic scattering. Consider a monochromatic plane wave incident onto a rough surface. Waves with the same angle of incidence with respect to the local surface normal but at different heights will interfere with one another, due to the phase difference between the two scattered rays.

The Rayleigh criterion states that if the phase difference between two rays is less than $\pi/2$, the surface appears smooth and will specularly reflect light (i.e. angle of incidence is equal to angle of reflection). If the phase difference is larger than $\pi/2$, the surface appears rough and incident light will be scattered in many directions (*Ogilvy* 1991). This phase difference depends on both the incidence angle and the ratio of surface roughness to the wavelength of incident radiation. The Rayleigh criterion is,

$$cos(\theta_{\mathrm{i}}) < \frac{\lambda}{8\sigma} \tag{5.8}$$

where θ_{i} is the angle of incidence, λ is the wavelength of incident radiation, and σ is the roughness of the surface (*Ogilvy* (1991)).

Three cases are now discussed. All of them are relevant for this discussion.

5.3.2.1 Example 1. Very Rough Surface Scattering

Scattering from very rough surfaces ($\sigma/\lambda >> 1$) can be modeled as Lambertian/Diffuse, which is modeled by a constant BRDF (*Goral et al.* 1984). Absorption by very rough surfaces is due to multiple scattering events. A single integrated reflection coefficient for all photon-surface interactions is used as an approximation, making ρ constant. Therefore, $BRDF_{\mathrm{s}}$ is also a constant.

Given $BRDF_{\mathrm{s}}$, we can derive from equation (5.6) that the probability function for θ_{o} results in a cosine profile of the angle between the observer and the surface normal, i.e. a photon should be more likely to be reflected along the surface normal than perpendicular to it.

5.3.2.2 Example 2. Specular Reflection

Smooth surfaces ($\sigma/\lambda << 1$) act as specular reflectors, where the angle of reflected radiation equals the angle of incident radiation. The BRDF as shown in equation (5.9)

is therefore composed of delta functions. (*He et al.* 1991):

$$BRDF_{\text{specular}} = \frac{F(\theta_s, \phi_s)\delta(\theta - \theta_s)\delta(\phi - (\phi_s + \pi))}{cos(\theta)sin(\theta)} \tag{5.9}$$

where F is the Fresnel reflectivity of the surface at the incident direction (θ_s, ϕ_s). This reflectivity factor represents the ρ contribution to the total specular BRDF. Therefore, the scattering BRDF for specular reflection becomes:

$$BRDF_s = \frac{\delta(\theta - \theta_s)\delta(\phi - (\phi_s + \pi))}{cos(\theta)sin(\theta)} \tag{5.10}$$

Note that the delta functions here carries an additional factor of $cos(\theta)$ when compared with equation 5.3). This factor arises due to the fact that $BRDF_s cos(\theta)$ must be normalized rather than simply $BRDF_s$.

5.3.2.3 Example 3. Moderately Rough Surface Scattering

Purely diffuse or specular reflection are limiting cases of a more general theory of scattering. One theoretical simplification of that intermediate case ($\sigma/\lambda \approx 1$) is Kirchhoff theory, which models the surface as an array of small flat surface elements, but neglects multiple scattering (*Ogilvy* 1991). Futher, the statistical height distribution of these surface structures are modeled as a Gaussian with a standard deviation σ and an autocorrelation length τ, indicating a scale for height and the extent of surface roughness. For a Gaussian height distribution, the standard deviation σ is equivalent to the surface roughness σ defined in Section 5.3.2. *He et al.* (1991) developed what is regarded as one of the most accurate BRDF models to date using the vector form the Kirchhoff theory. However, due to its complexity, the He-Torrance BRDF is not practical to implement in SIMION. A more recent model by *Li and Torrance* (2005) simplifies some components of the He-Torrance BRDF with empirical formulas, and will be used here.

A complete scattering model includes specular, directed diffuse, and uniform diffuse reflection components as well as subsurface effects (*Li and Torrance* 2005). The uniform diffuse and subsurface reflection terms corresponds to multiple scattering events. *Li and Torrance* (2005) provide guidelines for when each term should be included. For example, for conducting surfaces, subsurface and multiple scattering effects can be neglected.

The BRDF for a moderately rough surface becomes a model with four parameters which are material dependent. These parameters are the surface roughness σ, the correlation length τ, and the real and complex parts of the index of refraction $n + i\kappa$ for a given wavelength. These have been found for a number of materials commonly used for space instrument applications (*Zurbuchen et al.* 1995). Additional parameters may be required to appropriately weight the contribution of multiple scattering. As in the case of specular reflection, for single scattering events, the $BRDF_{\mathrm{s}}$ function can be derived from the full BRDF by dividing by the appropriate Fresnel reflectivity.

5.3.3 Calculating Suppression

To generate an EUV suppression metric for a particular geometry, large numbers of individual photon trajectories are calculated. Changes in trajectory for each photon is determined by the $BRDF_{\mathrm{s}}$ distribution function at the location of each photon-surface collision through a Monte-Carlo process. However, the probability of absorption is carried along with each photon through its entire trajectory, and applied post-flight. In this way, fewer photons need to be rendered in order to obtain suppression estimates. For example, a single photon which reaches a detector after three reflections with corresponding reflections coefficients of 10% represents a suppression of 10^{-3}. Incorporating photon absorption as a Monte-Carlo process would require 10^3 rendered photons to achieve the same result.

A new suppression estimate can be calculated after each rendered photon. With

enough photons, this estimate should converge to the EUV suppression of the geometry of interest. When a rendered photon reaches the detector, the suppression estimate will increase. When a rendered photon does not reach the detector, the suppression estimate will decrease. In this way, the suppression estimate will perform damped oscillations of sorts, eventually converging to the true EUV suppression value.

5.4 WIND/MASS Instrument Simulations

With a fully developed model for the scattering of light from a surface, we can now consider an application to WIND/MASS which was fully analyzed in end-to-end EUV suppression measurements, as well as the detailed BRDF measurements of surface samples. WIND/MASS therefore provides a very good example to apply and test EUV scattering algorithms developed in this paper.

Lyman-α photons are considered the primary source (approximately 80%) of solar EUV at 1AU (*Bauer and Lammer* 2004). Therefore, laboratory testing was performed at the Lyman-α wavelength of 121.6nm at the University of Bern prior to launch (*Zurbuchen* 1992). Futhermore, measurements were conducted with differing surface finishes of the same analyzer. The first system used untreated aluminum plates. The second system's plates were coated with Cu_2S, which blackened and roughened the surfaces as shown in Fig. 5.3. BRDF measurements in the plane of incidence of these surface samples are available in *Zurbuchen et al.* (1995).

An MCP was placed at the exit aperture of each WIND/MASS ESA system to measure photons which would reach the carbon foil. Unpolarized light entered the ESA on-axis, parallel to both sets of partial spherical plates. A three-dimensional SIMION model of the WIND/MASS simulation can be observed in Fig. 5.4, including all relevant electrodes and insulators. EUV photons scattered outside of these electrodes are considered lost for the simulation. We implemented the process described

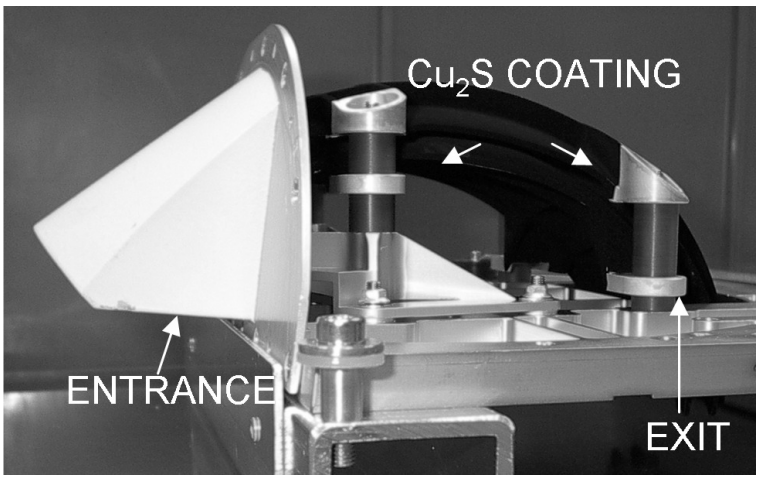

Figure 5.3: Side view of the WIND/MASS electrostatic analyzer system with Cu_2S treated surfaces. This geometry provides the basis for the simulations performed in this paper.

in Section 5.3 and performed simulations analgous to the laboratory experiments.

5.4.1 WIND/MASS with Sandblasted Aluminum Surfaces

The sandblasted aluminum is a moderately rough surface as described in Section 3.2.3. The Li-Torrance model was applied using values from *Zurbuchen et al.* (1995) of $\lambda = 121.6$nm, $\sigma = 44$nm, $\tau = 210$nm, $n = 0.855$, and $\kappa = 0.607$. Figure 5.5 shows the calculated BRDF for several angles of incidence. As expected, specular reflection becomes more significant at higher angles of incidence.

A sufficient number of photons must be rendered in order to obtain an accurate representation of the simulated EUV suppression. The calculated EUV suppression converged to $9.7(\pm 0.2) \times 10^{-6}$ using 2×10^6 photons with a maximum of ten reflections per photon allowed in the simulation.

5.4.2 WIND/MASS with Cu_2S Surface Treatment

For the Cu_2S surface, which is very rough, a pure Lambertian reflectance was assumed with constant reflection coefficient of 0.02 from *Zurbuchen et al.* (1995). This reflection coefficient is the integrated reflectivity of the Cu_2S measured normal

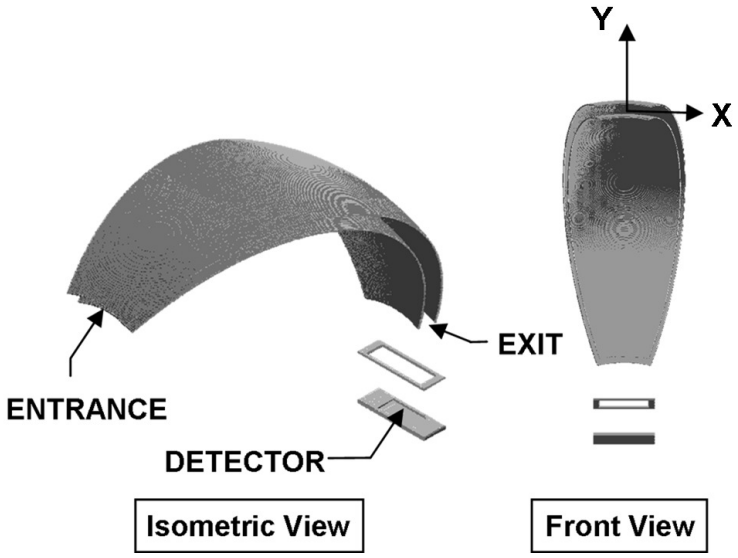

Figure 5.4: SIMION 3D model of WIND/MASS electrostatic analyzer with exit plates and photon detector. This geometry is identical to the experimental test-setup to which the simulations are compared. The coordinate axis (X,Y) of the entrance aperture is shown in the front view.

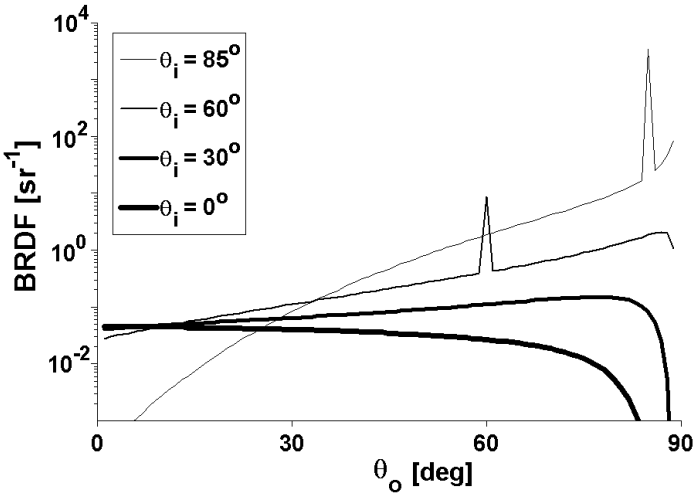

Figure 5.5: Calculated BRDF for sandblasted aluminum using the Li-Torrance model and surface scattering parameters from *Zurbuchen et al.* (1995). Curves are shown for radiation scattered in the plane of incidence ($\phi_o = \phi_i + \pi$) for various values of incident angle θ_i. Specular reflection ($\theta_o = \theta_i$) becomes increasingly important at higher angles of incidence as shown in the $\theta_i = 60°$ and $\theta_i = 85°$ curves. For lower angles of incidence, specular reflection is negligible causing no visible specular peaks to appear in the $\theta_i = 0°$ and $\theta_i = 30°$ curves.

incidence. As discussed in Section 5.3.3, using this value as the reflection coefficient is an approximation. However, this is necessary since only two-dimensional plane of incidence BRDF was measured for the Cu_2S surface and we must extend that into three dimensions. 10^7 photons were rendered yielding an EUV suppression factor of $5.5(\pm 1) \times 10^{-13}$.

5.4.3 WIND/MASS with Polished Aluminum Surfaces

Although no laboratory data is available for comparison, it is instructive to also examine the EUV suppression of the ESA with a smooth surface ($\sigma \to 0$), so that purely specular reflection of radiation is the only contribution to photon flux at the detector. With this assumption, a EUV suppression of $1.33(\pm 0.01) \times 10^{-2}$ is obtained for 10^5 rendered photons, which is, as expected, higher than both the sandblasted aluminum and Cu_2S cases.

5.4.4 WIND/MASS Suppression Measurements

The laboratory measured EUV suppression factors of WIND/MASS were 5.5×10^{-5} and 3.3×10^{-9} for the sandblasted aluminum and Cu_2S treated plates, respectively. The Cu_2S system was chosen for flight since it met the 10^{-8} suppression requirement established for the mission. Count rates for the MASS start detector are available data products of the WIND mission and are sectored into time periods when the instrument is facing towards and away from the sun. The count rate of the highest energy bin (10keV/e), where minimal particles are expected, was accumulated for a ten day period during solar minimum (August 2, 2008 to August 12, 2008). The average dark count rates were 5.12 counts/sec and 4.78 counts/sec for the sun facing and non-sun facing sectors respectively. The difference can be attributed to EUV flux at the carbon foil, yielding a photon induced count rate of 0.34 counts/sec, corresponding to an in-flight EUV suppression factor of 3.5×10^{-9}, consistent with

the laboratory measurements of the Cu_2S system.

5.4.5 Simulation/Measurement Comparison

The mean sandblasted aluminum simulation result of 9.7×10^{-6} is in agreement, to within approximately a factor of 5, with the measured suppression of 5.5×10^{-5}. However, the model parameters σ and τ from *Zurbuchen et al.* (1995) are only accurate to within factor of 1.5. A $\sigma/1.5$ value of 30nm and yields a suppression of $3.06(\pm 0.08) \times 10^{-4}$ with τ held constant and a 1.5τ value of 315nm yields a suppression of $1.13(\pm 0.04) \times 10^{-4}$ respectively with σ held constant. In each case, the measured suppression of the aluminum plates is within the error of the model parameters, considering the simplicity and limitations of the model used.

The mean Cu_2S simulation result of 5.5×10^{-13}, however, is several orders of magnitude below the measured suppression. The large discrepancy is due to the BRDF modeling of the Cu_2S surface. The assumption of completely diffuse reflection may be valid for lower angles of incidence, but forward scattering peaks appear at grazing incidence in the measured Cu_2S BRDF (*Zurbuchen et al.* 1995). Due to the geometry of the WIND/MASS entrance system, on-axis light reflects at such angles.

Cu_2S is a very rough dielectric surface such that Kirchhoff theory is clearly no longer applicable. We can, however, construct a crude model which takes into account forward scattering at high angles of incidence by defining a threshold analagous to the Rayleigh criterion: above a certain angle of incidence the scattering of radiation becomes specular; below this threshold, the scattering is Lambertian. Since the initial angles of incidence are all high ($\gtrapprox 70^o$), this system exhibits specular scattering, as in the aluminum mirror case. Reflected radiation from the Cu_2S surface results from subsurface and multiple scattering, requiring the constant integrated reflection coefficient of 0.02 to be applied instead of the corresponding Fresnel reflectivity as discussed in Section 5.4.2. This method yields an EUV suppression of $7.0(\pm 0.1) \times$

10^{-8}, which can represent an upper bound of the EUV photon flux at the detector. The previous Cu_2S simulation result is a lower bound of the EUV flux. The measured EUV suppression of the Cu_2S system falls between these two cases as expected.

This example illustrates the difficulty in accurately modeling scattering from very rough surfaces, especially at grazing angles of incidence, well beyond the range of simplifications like the ones made in Kirchhoff theory. More comprehensive BRDF measurements of rough surfaces at Lyman-α could be easily implemented in our method in the form of lookup tables as opposed to relying on parameterized models. However, these measurements would have to extend in all directions rather than just the plane of incidence as in *Zurbuchen et al.* (1995) in order to accurately map the three-dimensional scattering from a surface sample.

5.5 Analysis

With the presented photon tracing method, analysis beyond simply computing EUV suppression for a proposed ESA design is possible. It is equally important to understand what factors drive that suppression. By tracking individual photons, the path of light hitting the detector can be directly traced back to the ESA entrance aperture, or any other location of interest. EUV sensitive areas of the ESA can then be identified and mitigated. In addition, EUV suppression can be computed for a number of designs to develop a trade-relation between EUV suppression and other design drivers such as energy resolution or geometric factor.

5.5.1 Response Across Aperture

Figure 5.6 shows the EUV sensitivity of the WIND/MASS ESA at its entrance for the aluminum mirror surface. For each specular reflection inside of a spherical surface, the angle of incidence remains constant. The consequence for the aluminum mirror case is that discrete bands of photons will reach the detector. Each of these bands

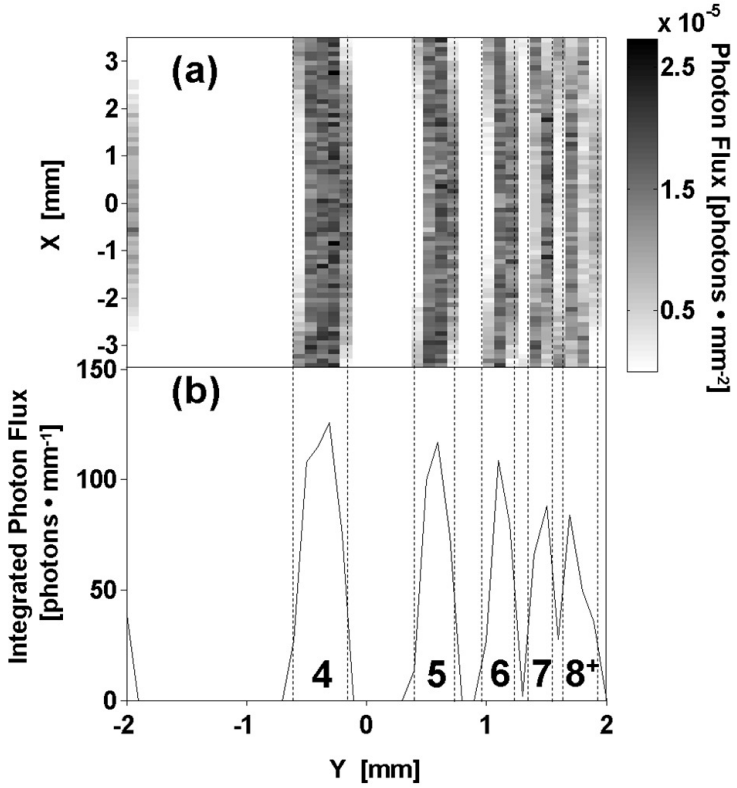

Figure 5.6: (a) Photon transmission as a function of photon entrance coordinate (X,Y) for the aluminum mirror simulation (See Fig. 5.4). (b) The photon intensity is binned according to entrance height. Each peak corresponds to a certain number of reflections a photon much undergo before it reaches the detector. The number of reflections corresponding to each peak is labeled. The dashed vertical lines highlight the passbands of photons at the entrance aperture.

represents a region where a photon must undergo a particular number of reflections before exiting the ESA. Due to the focusing of the spherical analyzer plates, the EUV suppression is almost independent of the entrance X coordinate of the aperture. (See Fig. 5.4 for coordinate system). For the case of SIMION, this would allow for the use of higher resolution two-dimensional models as discussed in Section 5.2.

There is a clear ordering of transmission according to bounce numbers; the smaller number of scattering interactions, the larger the transmitted radiation. However, even photons which undergo a large number of reflections (i.e. >5) can significantly

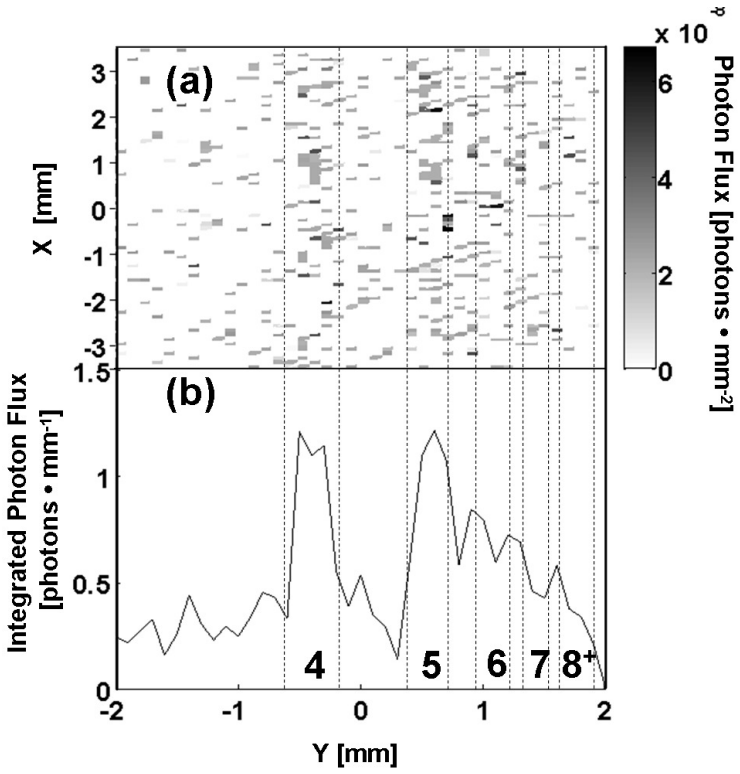

Figure 5.7: Same format as Fig. 5.6 but for sandblasted aluminum case.

contribute to EUV flux reaching the detector since the Fresnel reflectivity of the material typically increases with angle of incidence. For the spherical geometry of WIND/MASS, increased numbers of scattering events corresponds to higher angles of incidence.

The sandblasted aluminum case seen in Fig. 5.7 shows bands forming around the same locations of the specular bands in Fig. 5.6. Since light is diffusely scattered around the specular direction, the bands are wider than the specular case and diminished in peak amplitude. However, it is clear that most of the EUV flux at the detector comes from the light that reflects almost specularly through the ESA.

There are regions at the ESA aperture through which photons are much more likely to successfully reach the detector, reducing the effectiveness of EUV suppression. From this analysis, it is clear that simply scaling down the entire size of the aperture

may not be the best way to improve EUV suppression. For example, from Fig. 5.7 it can be seen that reducing the height of the aperture by 1mm from top eliminates the contribution from photons which undergo more than 5 specular reflections before reaching the detector, whereas reducing the height of the aperture by 1mm from the bottom only eliminates some of the diffusely scattered photon flux. For reference, a 1mm reduction in aperture height (from either the top or the bottom) lowers the geometric factor of the ESA by approximately 30% (*Funsten and McComas* 1998).

5.5.2 Effects of Plate Spacing

The quantitative dependence of the EUV suppression on the analyzer gap-distance, a key quantity that strongly affects the ion performance of ESAs, is now modeled for both specular reflection and sandblasted aluminum. In both cases, we use a range of plate spacings between 2mm and 12mm. A sufficient number of photons were rendered to obtain a convergent EUV suppression. Figure 5.8 shows the total suppression factor as a function of plate spacing. Error bars are shown for the sandblasted aluminum data points. Convergence errors are small enough for the aluminum mirror case so that error bars are not visible.

Figure 5.8 demonstrates the very strong and near stepwise dependence of the overall aluminum mirror transmission curve on the gapwidth. These steps are caused by the locations of the bands of photon flux in Fig. 5.6. There are expected to be large jumps in EUV suppression corresponding to when a new band becomes visible, such as at spacings of 4mm, 6mm, and 11mm where four, three and two bounce photons are able to reach the detector, respectively. Between these points the EUV suppression is expected to remain relatively constant, and is in fact shown to decrease slightly with an increasing gap, due to the lower Fresnel reflectivity given by a lower angle of photon incidence. While similar bands are shown to exist for the sandblasted aluminum in Fig. 5.7, the scattering of photons is not strictly limited to

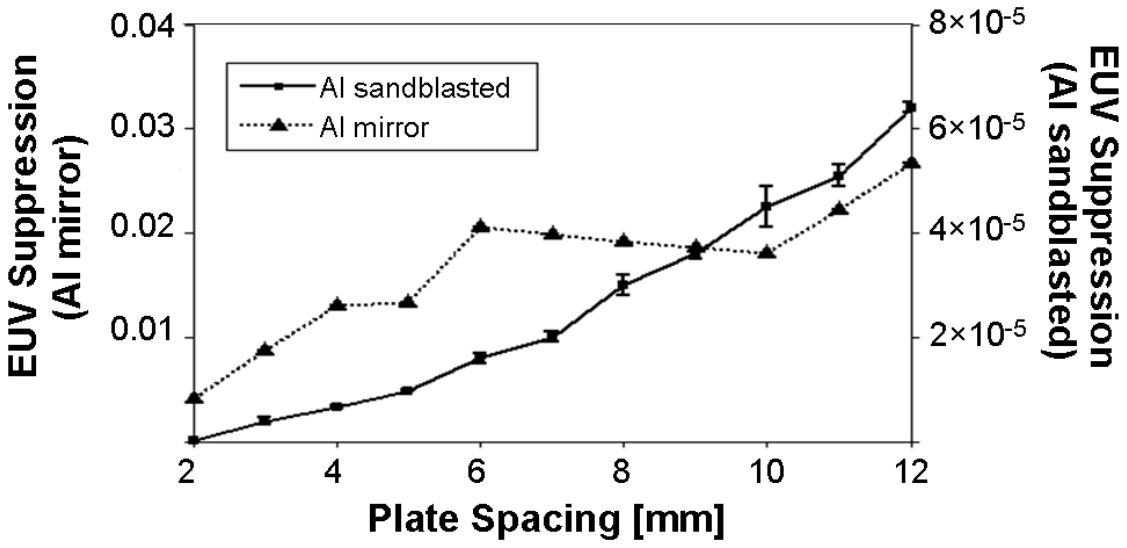

Figure 5.8: EUV suppression factor as a function of WIND/MASS analyzer electrode spacing for aluminum mirror and sandblasted aluminum simulations. Error bars are shown for the sandblasted aluminum. Convergence errors are small enough for the aluminum mirror case so that error bars are not visible.

the same regions as for the pure specular reflection case. This causes the transmission to increase almost linearly with plate spacing as shown in Fig. 5.8.

With a known relationship between EUV suppression and electrode spacing, EUV suppression can become a driving design factor, rather than an afterthought. Relationships between the spacing and other ESA properties such as energy resolution and geometric factor are well known for spherical analyzers (*Paolini and Theodoridis* 1967; *Theodoridis and Paolini* 1969). For example, for the aluminum mirror case, there are regions where a higher bandwidth will not lead to a significant increase of EUV flux reaching the detector. For the roughened aluminum, both energy resolution and EUV suppression increase linearly with electrode spacing. A trade-relation between EUV suppression and other important ESA properties can be developed which will incorporate EUV suppression more tightly into the analyzer design process.

5.6 Conclusions

This paper presented a novel technique to model the EUV suppression properties of electrostatic analyzers in SIMION. Using measured surface scattering data, the model was shown to provide estimates of EUV suppression when compared with laboratory measurements of the WIND/MASS analyzer. These results illustrate that EUV suppression depends heavily not only on the geometry of an ESA design, but also on the reflective properties of a material, exposing a nontrivial relationship between electrode spacing and EUV suppression in the case of WIND/MASS. Employing the presented technique provides new analysis tools for instrument designers, such as the discovery of EUV-sensitive areas of the analyzer and the ability to compare the EUV suppression of different geometries quickly, using pre-existing instrument models.

CHAPTER VI

Noise characterization and design considerations of in space-based time-of-flight sensors

This chapter is taken from: Gilbert, J. A., D. J. Gershman, G. Gloeckler, R. A. Lundgren, T. M. Orlando, J. McLain, and T. H. Zurbuchen (2012), Background noise in space-based time-of-flight sensors - Part I: Characterization and design consideration, *Space Sci. Rev.*, in prep. In addition to assisting with the development and revision of the manuscript text, I was responsible for the electron stimulated desorption contribution to the paper, including the laboratory testing and SIMION simulations.

6.1 Introduction

The composition of space plasmas is often measured in situ using time-of-flight (TOF) mass spectrometers (*Wüest* 1998). These instruments continually make discoveries that enhance our understanding of the space environment. For example, the composition of pickup ions (*Gloeckler and Geiss* 1998b; *Kallenbach et al.* 2000) contains information about the properties of the interstellar medium, planetary at-

mospheres and comets, as well as weathering processes on planetary bodies. Heavy (mass per charge ratio $(m/q) > 2$ amu/e) ions in the solar wind (*von Steiger et al.* 2000) reveal information about its origin and about the formation and evolution of coronal mass ejections. In addition, the composition and velocity distributions of the plasma environment around various planets, comets, and moons can provide insights in their origin, key physical processes, and evolution.

The Solar Wind Ion Composition Spectrometers (SWICS) onboard the Advanced Composition Explorer (ACE) (*Gloeckler et al.* 1998) and the Ulysses (*Gloeckler et al.* 1992) spacecraft have discovered most of the known heliospheric pickup ion species (e.g., *Gloeckler et al.* (1993); *Geiss et al.* (1995); *Gloeckler and Geiss* (1996); *Gloeckler et al.* (2000a)) as well as elucidated the compositional differences between the steady (usually fast, compositionally like the photosphere) and unsteady (compositionally like the corona) solar wind (*von Steiger et al.* 2000). Other TOF sensors include the Fast Imaging Plasma Spectrometer (FIPS), which is part of the Energetic Particle and Plasma Spectrometer suite (*Andrews et al.* 2007) on the Mercury Surface, Space Environment, Geochemistry, and Ranging (MESSENGER) spacecraft (*Solomon et al.* 2007); the Rosetta Orbiter Spectrometer for Ion and Neutral Analysis (ROSINA) (*Balsiger et al.* 2007) on the comet-exploring Rosetta mission; and the Cassini Plasma Spectrometer (CAPS) (*Young et al.* 2004) in Saturn‘s magnetosphere. The scientific value provided by TOF mass spectrometers is immense, and a new generation of instruments will enable discoveries that are not possible with the present technology.

Often, the answers to important science questions are found in measurement regimes where the signature of measured ions is obscured by significant fluxes of unwanted ‘noise’ events that complicate data analysis. For the purposes of this paper, we define noise in an instrument to be any measured event that would be eliminated in a perfect instrument design. By identifying and characterizing some of the common sources of noise in TOF spectrometers, simple design considerations can be

implemented to greatly increase the sensitivity of future instruments, enabling the scientific analysis that may have otherwise been lost in a measured dataset. Some of these noise sources can also affect instruments that do not rely on TOF for particle identification, such as electron analyzers. Nonetheless, the principles of background noise suppression developed here should provide benefits to such sensors.

This paper, the first of a two-part study, will discuss the most common noise sources in spaceborne instruments and will discuss design guidelines that can mitigate their effects before, or in some cases, during, flight operation. Two TOF sensors will be used as examples here, SWICS on ACE and FIPS on MESSENGER; these sensors are introduced in detail in Section 6.2. In Section 6.3, we describe each noise source and their effect on the measured in-flight datasets of SWICS and FIPS, with additional discussion in Section 6.4. The second part of this study (*Gershman et al.* 2012a) will introduce and implement data analysis techniques that characterize, forward model, and remove noise events from already measured TOF datasets.

6.2 Description of example TOF sensors

Many of the noise sources detailed in this paper afflict a variety of instrument designs. To facilitate the discussion, two specific TOF system designs are used here as examples, both of which utilize secondary electron emission for TOF triggers: (1) grid-free TOF telescopes (*Gloeckler et al.* 1985, 1992) and (2) TOF telescopes that use electrostatic mirrors (e.g., *Wilken* (1984); *Zurbuchen et al.* (1998); *Andrews et al.* (2007)). Other designs, such as linear-electric-field (LEF) TOF (e.g., *Hamilton et al.* (1990); *McComas and Nordholt* (1990); *Gilbert et al.* (2010)) and reflectron TOF (e.g., *Kissel and Krueger* (1987); *Balsiger et al.* (2007)), may be affected by some of these background signatures as well. A comprehensive review of TOF sensors, including a chronicle of their history, a discussion of their respective limitations, and a list of instruments and the missions on which they have flown, can be found in *Wüest*

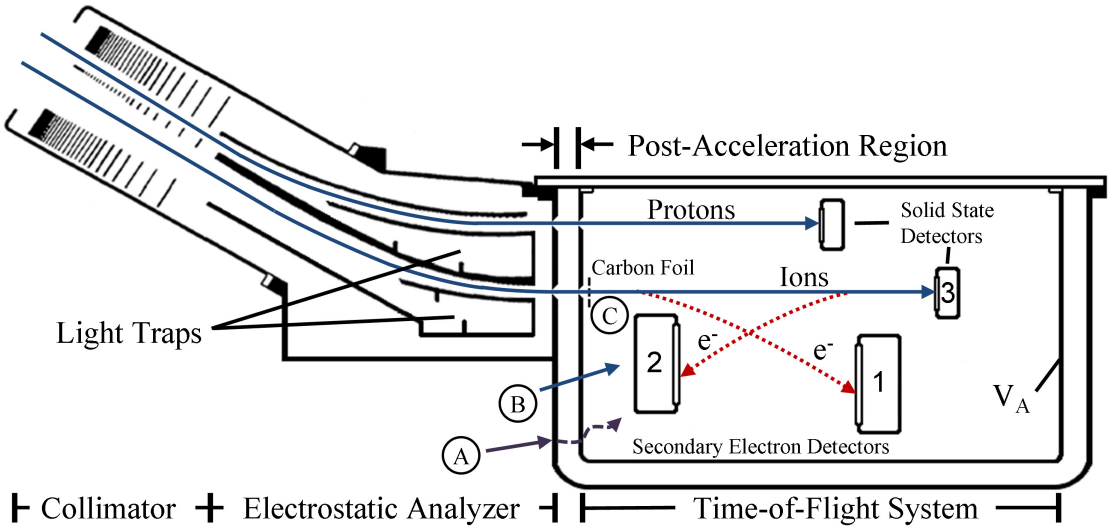

Figure 6.1: Schematic representation of the ACE-SWICS ESA and TOF sections (adapted from *Gloeckler et al.* 1998). Ion and secondary electron trajectories are shown, as well as the three detectors used for triple-coincidence measurements: (1) start MCP, (2) stop MCP, and (3) energy SSD. Some sources of noise are also illustrated: (A) Bremsstrahlung radiation, (B) penetrating ions, and (C) energy straggling.

(1998).

6.2.1 ACE/SWICS

The SWICS (*Gloeckler et al.* 1998) instrument, as shown in Figure 6.1, consists of a small-angle deflection electrostatic analyzer (ESA) that uses shaped electrodes with applied voltage potential to filter ions according to their energy-per-charge ratios (E/q), and a grid-free TOF analyzer that records both the time of flight and the total particle energy (E). In the ESA, voltage settings on the electrode plates are sequentially adjusted to scan over a range of ion E/q values. Light traps and serrated, black-coated surfaces on the deflection plates attenuate incident photons through a series of reflections.

Ions that pass through the ESA enter a post-acceleration region where they are

accelerated in proportion to their charge state (q) by a high voltage $V_{\rm A}$. This increased energy reduces the amount of energy straggling and angular scattering that an ion experiences as it passes through a thin carbon foil at the end of the acceleration region. In many cases, this acceleration also allows the particle to register a measurement above the energy threshold of a solid state detector (SSD).

Ions that pass through the foil cause secondary electrons to be ejected from both the entrance and exit surfaces of the foil (*Allegrini et al.* 2003). The electrons that follow the ion into the TOF system are electrostatically guided towards a detection assembly consisting of stacked microchannel plates (MCP) in front of an anode. When this detector is triggered, a start signal for the flight time is recorded. The original ion continues in a straight trajectory through the TOF system, often in a neutralized charge state, but occasionally with charge state distributions that vary with the incident energy and the elemental species (*Funsten et al.* 1993; *Bürgi et al.* 1993; *Funsten* 1995). The residual energy (E) of the ion is recorded when it impacts the SSD, which has an ion-dependent detection efficiency that reduces to zero below the energy detection threshold of ∼50 keV for heavy elements and 25-35 keV for H and He (*Ipavich et al.* 1978; *von Steiger et al.* 2000). Secondary electrons are released from the SSD and electrostatically guided toward an MCP assembly where they trigger the stop signal for the ion flight time.

In Figure 6.1, the start (1), stop (2), and energy (3) measurements are the three signals required for a valid 'triple-coincidence' event. The measured TOF, the E/q setting of the ESA, and the measured E can be used to calculate the mass (M), Mass-per-charge ratio (M/q), and the charge (q) of the ion. If the SSD is not triggered during an ion measurement, which is common for heavy ions with low charge states such as pickup ions, the start and stop TOF signals can still create a 'double-coincidence' event from which the M/q can be calculated (*Gilbert et al.* (2012)). For later reference, some sources of noise events are marked in the figure: (A) Bremsstrahlung

radiation, (B) penetrating ions, and (C) energy straggling at the carbon foil.

6.2.2 Fast Imaging Plasma Spectrometer

The FIPS sensor also consists of an ESA and a TOF section, with electrostatic mirrors used for secondary electron deflection rather than the grid-free design of SWICS. Ions are accepted into the sensor through an annular aperture and guided through an ESA that filters them according to their E/q ratio by sequentially stepping the voltage settings on the electrostatic deflection plates. The inner surfaces of the plates are serrated and plated with black chrome (*McDonald* 1975) to improve the photon attenuation. In addition, the detailed design of the ESA ensures that any incident photon will undergo multiple surface reflections before arriving at the carbon foil. As ions traverse the foil, secondary electrons are emitted that are guided toward an MCP, triggering a start signal for the time of flight measurement and subsequently recording the impact position on a wedge-strip-zigzag anode. The ions strike an MCP to trigger a stop signal, completing the TOF measurement. The trajectories of secondary electrons are guided by an electrostatic mirror harp (*Busch* 1980) wound from 0.001-inch diameter wire. The 45° angle of the harp with respect to the carbon foil ensures that the incident positions of ions passing through the foil will be effectively imaged by secondary electrons on the start anode.

Figure 6.2 shows a schematic of the FIPS sensor, with examples of three potential sources of background noise marked. Electrons emitted from high-voltage wires, such as the mirror harps or electrostatic suppression grids, can strike the MCP and caused electron-stimulated desorption (D) of surface ions. Ions such as these can strike surfaces within the instrument and cause ion-induced electron emission (E), creating secondary electrons that could possibly trigger signals on the start or stop MCPs. During times of high flux, the probability of accidental coincidence measurements (F) increases, in which the start and stop timing signals can be created by two different

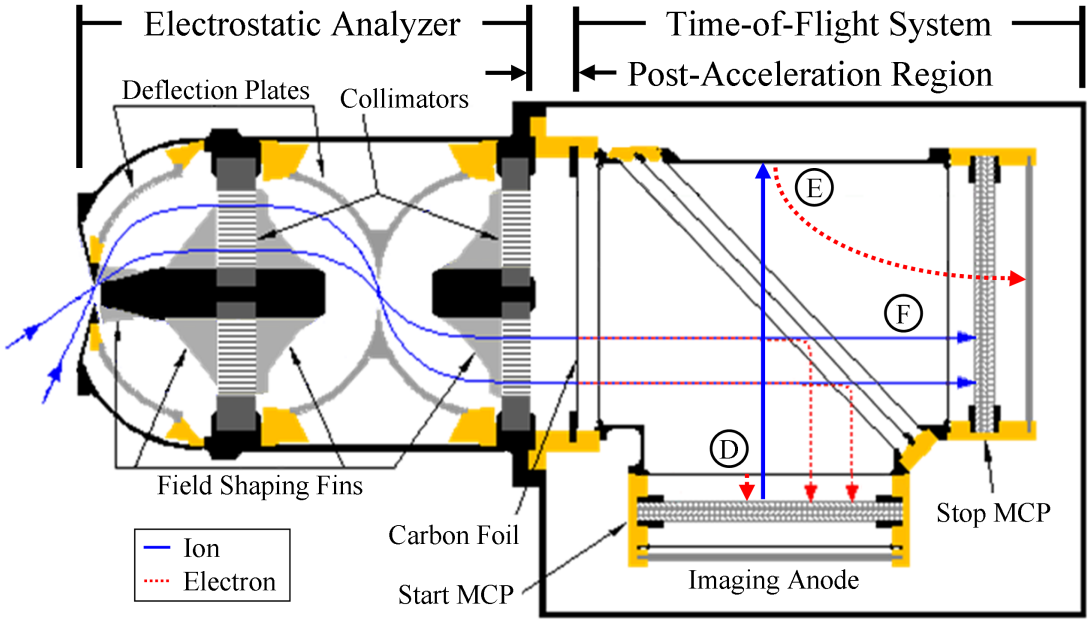

Figure 6.2: The FIPS instrument (illustration adapted from *Andrews et al.* 2007) guides ions through an ESA that is optimized for UV reduction, through a post-acceleration region that accelerates ion to sufficient energies to penetrate a thin carbon foil, and into an electrostatic mirror TOF system where a double-coincidence measurement is made. Some sources of background noise are marked: (D) electron stimulated desorption, (E) ion-induced electron emission, and (F) accidental coincidence detections.

particles.

6.3 Sources of noise events

The concept of noise in a data set can apply to a broad range of topics. Rather than provide a comprehensive description of every conceivable classification of noise events observed in TOF measurements, we will focus on some of the most common categories, their causes, and techniques for mitigating their effects. Table 6.1 categorizes sources of noise events that affect TOF mass spectrometers with their corresponding identifiable signatures in measured data sets. These signatures will be introduced

Table 6.1: Possible sources of background noise events and their symptoms in TOF mass spectrometers.

	Source	**Effect**
Non-ESA filtered	Penetrating radiation	Signals generated at all TOFs, independent of E/q settings
	Electron stimulated desorption	TOF events that correspond to trajectories of ions or neutral particles desorbed from surfaces within the instrument, independent of E/q settings
	Ion-induced electron emission	TOF signals corresponding to particles sputtered or emitted due to ion impacts
	UV photons	Events recorded at all E/q settings and any TOF
ESA filtered	High incident flux	Accidental coincidence measurements covering all TOF values in specific E/q steps.
	Energy straggling	In the E/q steps corresponding to an ion species, events have the expected TOF for that species, and slower.
Other	Electronics	Periodic or recurrent signals, patterns, etc.

here and discussed in more detail in later subsections. We classify noise events into three main categories: (1) events that are not filtered by the voltage settings of the ESA; (2) events that are filtered in the ESA but still trigger noise events; and (3) erroneous events related to electronics hardware.

Category (1) includes Bremsstrahlung radiation and high-energy particles (e.g., cosmic rays or solar energetic particles) that penetrate the housing of an instrument and cause detectors to record an event. Particles generated inside of the instrument, such as those initiating from electron stimulated desorption (ESD) of ions from instrument surface, or ion-induced electron emission, will also be independent of the ESA voltage settings. ESD will release low-energy particles, both ion and neutral, that, can trigger detectors and create a clear signature in the distribution of mea-

sured events. Electron emission induced from surface by incident ions can also be a source of measured events, either by impacting the detectors directly or by ESD. Also included in this category are photons, in particularly Lyman-α, can also traverse the ESA through a series of reflections (*Zurbuchen et al.* 1995; *Gershman and Zurbuchen* 2010), stimulating emission in the carbon foil or triggering particle detectors. The effect of these processes is to increase detector background event rates independent of the E/q step of the ESA. Detections coincident with incident ions passing through the instrument can also effect TOF measurements.

The second category of noise events includes accidental coincidences due to high incident ion flux and delayed flight times due to energy straggling of ions as they pass through the foil. Accidental coincidences from high incident ion flux appear at the E/q step of incident ions, but at times of flight that do not correspond to their true M/q. Energy straggling of ions is a consequence of the energy loss that an ion experiences when passing through a thin carbon foil, a process that has been characterized extensively (*Ipavich et al.* 1982; *Allegrini et al.* 2006; *Ziegler et al.* 2010; *Gilbert et al.* 2012). These ions will have significantly longer measured TOF values than what is typical for their species, contaminating mass peaks of higher M/q ions. In addition, ions that strike instrument surfaces can liberate secondary electrons that, in turn, can trigger ESD event generation processes to create events with specific TOF signatures at the E/q steps of incident particles.

Finally, the third category consists of signatures due to instrument electronics. These sources, mores than the physical processes in categories (1) and (2) that create erroneous events, may be highly instrument-specific. There are, however, some generalized sources that may be widely applicable to TOF sensors, and will be discussed.

6.3.1 Penetrating radiation

High-energy particles such as omnidirectional cosmic rays, as well as energetic protons and energetic electrons, have energies sufficient to penetrate the housing of an TOF instrument and trigger signals on the detectors. Energetic electrons can also generate Bremsstrahlung radiation upon impact, which can also be detected (*Vampola* 1998). Mass constraints on flight instruments often limit the amount of physical shielding that can be used, resulting in penetrating particles a common source of background noise for spaceborne instrumentation, not limited to TOF sensors.

6.3.1.1 Signatures of penetrating radiation

For a TOF spectrometer, any detection of penetrating radiation will be independent of the ESA voltage setting, i.e., counts will appear in the measured data at all E/q steps. In some cases, a penetrating particle or photon will trigger either a start or stop signal coincident with another detection - another penetrating particle, an ion or electron emitted within the instrument, or one that is filtered by the ESA as intended. These two nearly simultaneous events can result in an accidental-coincidence TOF measurement.

In the case of FIPS, protons with energies >13 MeV can penetrate the instrument housing and reach the MCP detector assemblies in the TOF telescope. When the penetrating flux is sufficiently high for ions to strike both the start and stop detectors within the limits of the pre-set valid timing window, a TOF measurement is recorded. Since these ions are not filtered by the voltage settings of the ESA, the TOF signal they produce is unrelated to the E/q values assigned to the voltage steps. They will be evident in the data as signals that appear in all E/q channels, with flight times that follow the pattern of accidental coincidence measurements (described in more detail in Section 6.3.5).

As an example, Figure 6.3 shows plasma data from the FIPS sensor during a

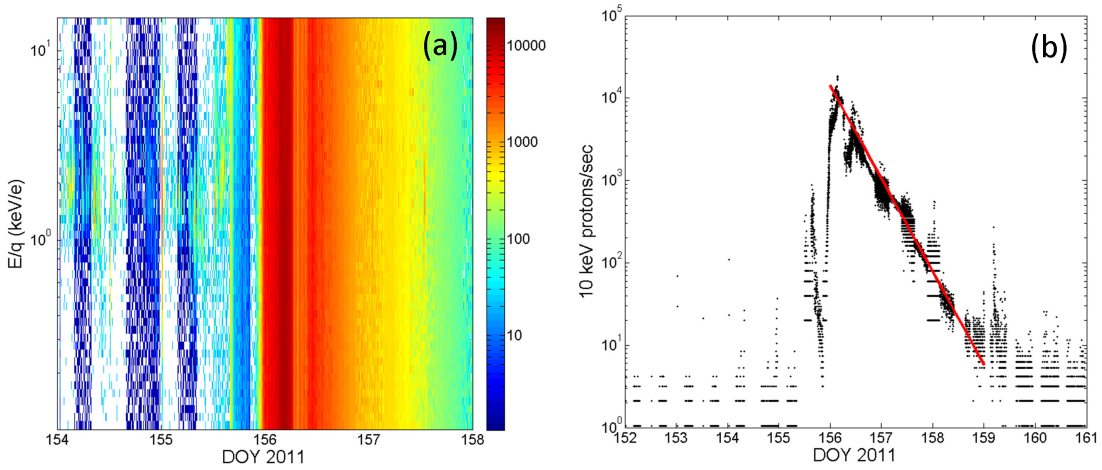

Figure 6.3: Proton count rates in the FIPS instrument during a 2011 June 4 ICME show the effects of penetrating radiation. (a) High-energy ions penetrated the instrument housing of FIPS, saturating the detectors at all E/q steps. Proton counts per second are shown with their E/q value plotted over time. (b) Proton count rates at the 10 keV/e step during the ICME show the effect of penetrating ions.

passing interplanetary coronal mass ejection (ICME) that began on 2011 June 4 (DOY 155). Figure 6.3a shows the proton event count rates at each E/q step. Penetrating radiation from the ICME briefly saturates the detectors, then decays over the course of the next several days. The E/q = 10 keV/e proton event rates are shown in Figure 6.3b, with added background noise from penetrating radiation several order of magnitude over that of the real signal.

6.3.1.2 Mitigation strategies for penetrating radiation

To some extent, background noise events from penetrating particles and radiation can be reduced with extra shielding. However, this solution is usually hindered by instrument mass constraints. An alternative approach used by several instruments is the use of 'anti-coincidence' techniques. Examples of instruments that have successfully utilized this design include the electron electrostatic analyzer (EESA-H) and the ion electrostatic analyzer (PESA-H) (*Lin et al.* 1995) on the Wind spacecraft,

which use an anti-coincidence scintillator surrounding the MCP assembly, and the Solar Energetic Particle Ionic Charge Analyzer (SEPICA) (*Möbius et al.* 1998) on ACE, which uses an anti-coincidence scintillator in conjunction with a silicon SSD. If an event triggers both the particle detector and the anti-coincidence detector, it can be discarded as penetrating radiation, substantially improving the signal to noise ratios of the TOF data set. Triple-coincidence measurements are also effective in mitigating the contributions of penetrating radiation, as will be discussed in more detail in Section 6.3.5.

6.3.2 Electron stimulated desorption

Desorption induced by electronic transitions (DIET) refers to the general phenomena of non-thermal desorption of atoms, molecules, and ions from surfaces. When electrons are the primary excitation source, DIET is more generally referred to as electron stimulated desorption (ESD). An important aspect of ESD is that the forces leading to desorption are due to the repulsive nature of the excited electron states. This is in contrast with typical inelastic ion sputtering, which relies on the momentum exchange between ejecta and an incoming particle (*Behrisch* 1981). In ESD, the original incident high energy electron triggers a cascade of a material (*Madey* 1971) and produces secondary electrons, some of which backscatter and escape the material surface. Consequently, high energy electrons that penetrate deep into a subtrance will deposit energy either directly or indirectly at the surface. The inelastic scattering of these electrons can excite the adsorbate-substrate complex via direction ionization, direct electronic excitation and resonant attachment processes. These leads to the desorption of cations, neutrals, and anions, respectively. Particles liberated from ESD of bulk material must diffuse to the surface between stimulating a desorption event. In addition, more reactive species produced in the bulk may recombine or may react to form more complex molecules (*Menzel and Gomer* 1964).

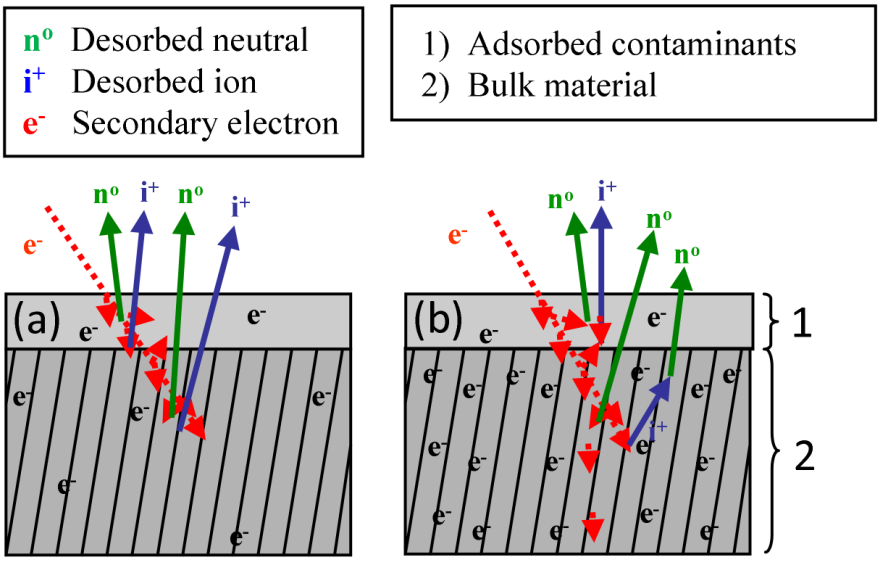

Figure 6.4: Illustration of ESD processes in (a) an insulating material and (b) a conducting material. Both surfaces will have organic and water molecules adsorbed to the surface, but only the conducting material will significant numbers of free electrons on the surface of the bulk material that enhance the recombination of desorbed ions.

Desorption can apply to either substances that are adsorbed onto the surface of a material (i.e. organic contaminants and water), or to the terminal surface molecules that are extensions of the bulk material itself (*Madey* 1971). For the case of space instruments, ESD is relevant for both insulating and conducting surfaces. In the case of an insulator (Figure 6.4a), incident electrons can readily desorb both neutral and ion species from both the adsorbed contaminent monolayers and the bulk material. For conducting material (Figure 6.4b), however, the high electron occupancy at the Fermi level significantly enhances the probability that a desorbed ion from a metal will neutralize (*Menzel and Gomer* 1964). Therefore, it is expected that the majority of the ions yielded from ESD on a conducting surface are from the adsorbed contaminants rather than the underlying metal.

In TOF systems, there are several sources of electrons suitable to cause desorption of instrument surface material, such as ion-induced electron emission from surfaces (*Hasselkamp* 1981), forward and backward secondary electrons from the carbon foil

(*Ritzau and Baragiola* 1998), or field emission of electrons from high potential wire grids or sharp corners in the instrument (*Fowler and Nordheim* 1928). These secondary electrons are created with only 1-2 eV with respect to their surface potential (*Madey* 1971), but can be accelerated within the instrument to energies that are sufficient for ESD. In fact, most ESD processes have an energy threshold on the order of only 10 eV (*McLain et al.* 2011). However, to be a source of background noise, the ions desorbed by ESD must trigger a double-coincidence event. Such a process is most likely to originate in two areas of an instrument: the post-acceleration region, discussed in conjunction with ion-induced electron emission in Section 6.3.3, and the surface of the start MCP, discussed here.

6.3.2.1 ESD and microchannel plates

The MCPs commonly used in start and stop detector assemblies contain arrays of angled channels with high length-to-diameter ratios. These channels are composed of leaded glass treated to be semiconducting and to have a high secondary electron emission (*Wiza* 1979). The MCP surfaces are covered with a conductive coating to allow a potential to be applied across each plate. An incident particle initiates a cascade of secondary electrons in one of the channels which, due the electric field across the plate, accumulates and accelerates out of the other side of the MCP as a charge cloud that can be readily measured.

The initial electron as well as cascaded secondary electrons can desorb ions from the MCP channels and surface (*Then* 1990; *Kulov et al.* 2005). Just as electrons are accelerated through the MCP, newly created ions are accelerated back towards the electron direction of incidence. This effect is well known for photon detection applications, and is called ion-feedback or ion-poisoning, as desorbed glass ions from the MCP channel walls cause damage to the input photocathode (*Pan et al.* 2010). For TOF applications, however, the effects are more subtle. The electron cascade can

not only create a valid start signal, but also can desorb ions from the MCP that may trigger a correlated stop signal, as illustrated in Figure 6.2. If the start and stop MCP assemblies are located directly across from each other, the ions or neutral particles desorbed from the start MCP could impact the stop detector directly, increasing the background noise level in the data.

In order to determine the types of desorbates expected to be observed in flight, laboratory ESD analysis was performed on a flight-spare MCP from FIPS using the experimental setup of *McLain et al.* (2011), which enabled measurements of desorbed ion species using an incident electron beam with energies up to 1 keV. The M/q spectra desorbed from the MCP surface by incident 1 keV electrons is shown in Figure 6.5a. The majority of the desorbed ions are from organic contaminants (C_x H_y) and water group ions with H^+ being the desorbate with the strongest measured signal. Only trace amounts of MCP glass were observed, with small peaks at M/q=16 amu/e for O^+ and M/q=28 amu/e for Si^+. The relative strength of ESD with varying electron energy is shown in Figure 6.5b. From 10 eV through 1 keV, ion desorption yields of all species remain relatively constant. Unlike ion sputtering processes, higher energy incident electrons are still expected to deposit significant energy near the surface of the sample, and therefore similar mass spectra are expected even at energies of several keV.

Due to the rapid formation of hydrocarbon monolayers when exposed to air (*Chambers* 2005), all instrument surfaces will be coated by them until they are removed through some means, such as in-flight high-temperature bake-outs, desorption by incident electrons or photons, or ion sputtering. Since the majority of the measured desorbates are from adsorbed monolayers rather than surface material, a M/q spectra similar to that shown in Figure 6.5 is expected not only from ions desorbed from MCPs, but from other instrument materials and surfaces as well.

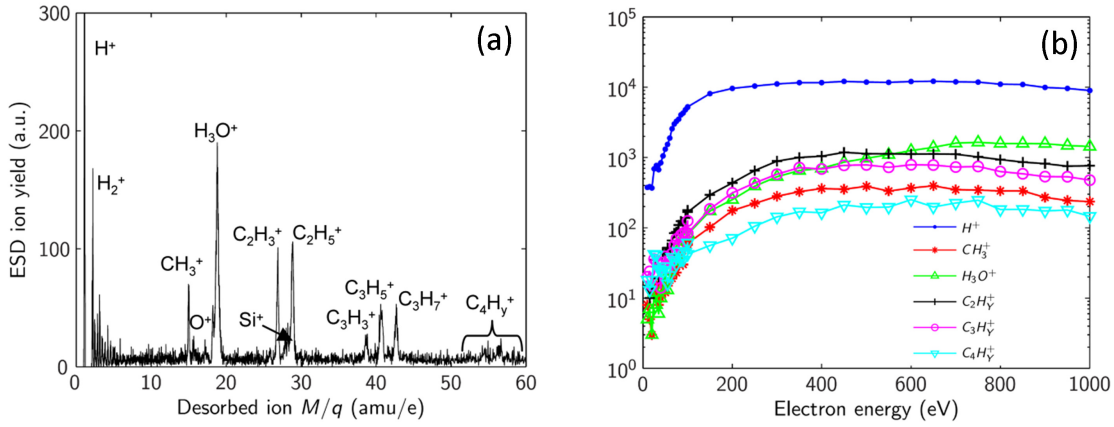

Figure 6.5: (a) The mass spectra of desorbed ions from a FIPS flight-spare MCP using 1 keV incident electrons shows trace amounts of MCP glass ions and a variety of hydrocarbons desorbed from the surface. (b) The energy dependence of desorbed ion yields are shown for various ion species.

6.3.2.2 Signatures of ESD noise

Background noise events from ESD processes can be observed in data from FIPS. The grids used for electric field suppression and for the electrostatic mirror harps are held at high voltage potentials, and therefore emit electrons even at room temperature. FIPS also contains unshielded high voltage cables that can be a source of field emission electrons. Although the TOF chamber is nominally a field-free region (*Andrews et al.* 2007), field penetration through the suppression grids occurs, creating a small attractive potential that draws low energy secondary electrons toward the start MCP, as shown in (D) of Figure 6.2. These two sources of electrons, the wire grids and the high voltage cables, create two distinct groupings of ESD-based background noise events in the FIPS data, both of which can be characterized and removed in post-processing (*Gershman et al.* 2012a).

Electrons emitted from the wire grids will be guided toward the start MCP, creating an almost spatially uniform desorption of ions and neutrals from the channels and surface of the MCP. Desorbed ions will be accelerated to energies of ∼5 keV

towards the roof of the FIPS TOF chamber, where they cause secondary electrons to be emitted, as shown in (E) of Figure 6.2. Field penetration through the mirror harps in the 'stop' half of the TOF system guides secondary electrons towards the stop MCP, resulting in a correlated start-stop pair, i.e., an ESD-based event. If the TOF system design had the start and stop MCPs facing each other, then this ion feedback process would create ESD events even more easily.

Beams of electrons emitted from nearby high voltage cables can reach the edges of the start MCP on FIPS, forming a 'hotspot' of ESD-based noise events. Ions desorbed near the edges of the MCP are subject to fringing field effects such that their trajectories may not reach the top of the TOF chamber, but rather strike the sides. In this case, secondary electrons are expected to be emitted and create a stop signal, generating a valid event that has a fraction of the TOF of ions that reach the top of the TOF chamber. This creates two sets of ESD-based events, a higher-TOF group of ions desorbed by the wire grid emission and a lower-TOF group desorbed at the edge of the MCP.

To characterize these ESD-based noise events in flight, measurements were taken with the ESA set to reject incoming ions and with a variety of voltage settings in the TOF section. The two sets of ESD-based noise are apparent in the solid black line of Figure 6.6a. The shaded region shows three peaks in the lower-TOF group of ions that are desorbed mainly from the hotspot of electron emission visible near the edge of the MCP position image of Figure 6.6b. The locations of the higher-TOF peaks match the travel times of the desorbates from Section 6.3.2.1 from the surface of the start MCP to the roof of the TOF chamber, as determined in Appendix B.1 using a three-dimensional ion-optics model of FIPS in SIMION software (*Dahl* 2000), described in Appendix B.1.

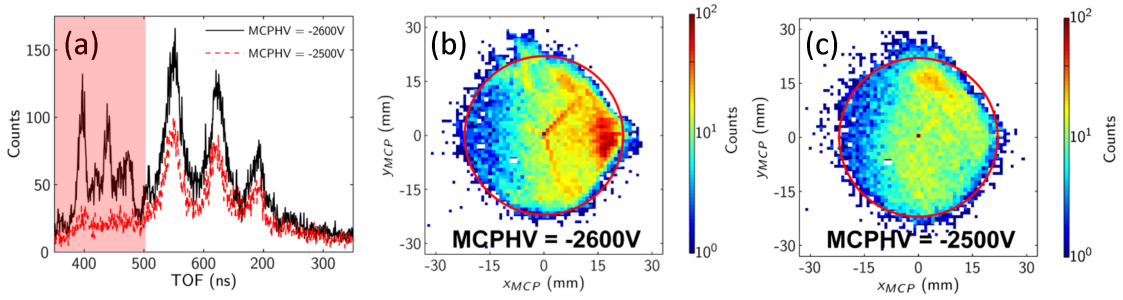

Figure 6.6: Significant reduction of localized ESD events was achieved by modifying the stray electric fields near the edge of the start detector assembly by lowering the MCPHV bias. (a) The TOF spectra as a function of MCPHV shows sets of peaks from the hotspot area (shaded) and from the wire grid electron emission. Certain voltage settings reduced the occurrence of hotspot events. The spatial location of the TOF start signal for two voltage settings indicates when noise events (b) were and (c) were not generated from the ESD hotspot.

6.3.2.3 Mitigation strategies for ESD

The wire corresponding to the FIPS start MCP bias voltage (MCPHV) is in close proximity to the location of the electron beam that forms the ESD hotspot. When MCPHV was adjusted by only 100V, the signature of these hotspot events disappeared, as shown in Figure 6.6a and 6.6c, likely due to changing fringing fields redirecting the electron beam away from the MCP active area. Such an adjustment is an example of a successful in-flight mitigation strategy of ESD noise.

Reducing the surface contamination from organic compounds and water molecules can also help reduce the effects of ESD. However, this reduction is extremely difficult to accomplish. In order to release adsorbed molecules from surfaces, the instrument must be baked at high temperatures, upwards of 600 K. While this is possible in a laboratory setting, contaminant monolayers will quickly accumulate on the surface once removed from vacuum. Therefore, only a high temperature bake-out in flight could potentially serve to clean the instrument surfaces. Backfilling the instrument with pure nitrogen gas until launch may reduce the exposure of surfaces to contam-

inants, but monolayers form so quickly at atmospheric pressure that it may not be possible to truly limit their adsorption.

The most effective time to take precautions against ESD noise is during the design and simulation phase of instrument development. Potential sources of electrons in the instrument should be identified and modeled. Suppression grids can be added to prevent electrons from backscattering from the carbon foil. The front of the MCPs can be set at a negative potential with respect to the surrounding TOF chamber or a physical barrier can be put in place to prevent ion feedback between the MCPS, as was done on each of the three SWICS instruments(*Gloeckler et al.* 1992, 1995b, 1998), for example. Simulations of electron and desorbed ion trajectories (Appendix B.1) can be invaluable in predicting significant ESD noise in instruments. Finally, during calibration, instrument data should be accumulated without incident particles to best characterize any background sources of electrons and desorbed ions.

6.3.3 Ion-induced electron emission

Ions can initiate background noise events when combined with electron stimulated desorption. For instruments that use post-acceleration, there may be a clear noise signature in the data that is attributable to a combination of ion-induced electron emission and ESD. A sketch of a sample post-acceleration region from SWICS is shown in Figure 6.7 with key events numbered.

(1) An ion exiting the ESA is accelerated through the post-acceleration region to reduce the amount of energy straggling and angular scattering at the carbon foil, and to raise the ion energy above any SSD energy threshold. The resolution of the instrument typically increases with higher post-acceleration voltage, V_{A}. On ACE-SWICS, for example, this voltage is set to -25.9 kV. The carbon foil is held at the post-acceleration voltage, and a grid between the carbon foil and the high voltage housing is biased slightly negative with respect to the foil. Secondary electrons emitted from

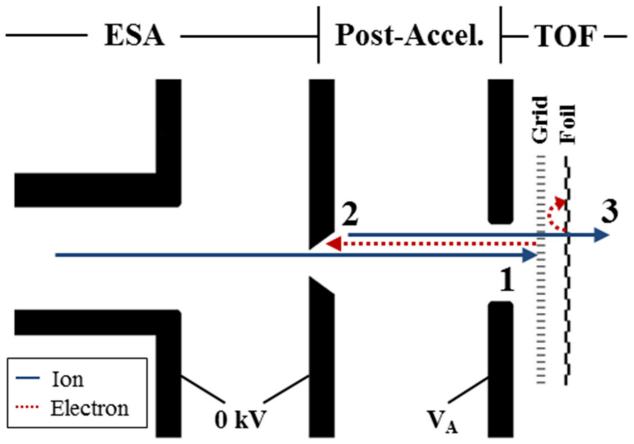

Figure 6.7: Trajectories of desorbed protons in the SWICS post-acceleration region. (1) Incident ions can strike a wire grid whose purpose is to suppress secondary electrons back-scattered from the carbon foil. (2) These ion-induced electrons are accelerated back toward the ESA, where they may strike the ESA housing and desorb ions from its surface. (3) These desorbed ions are accelerated and may enter the TOF system, where they trigger a measurement.

the foil during ion transit are prevented from entering the post-acceleration region by this grid. While the open area of the grid allows the majority of ions to pass freely through, some incident ions will strike the grid, leading to ion-induced electron emission (*Hasselkamp* 1981). These secondary electrons will be repelled by the grid and guided into the post-acceleration, where they will be accelerated back toward the ESA. (2) If the secondary electrons strike the grounded surface of the ESA housing, they can desorb ions via ESD. Positive ions desorbed from the surface will immediately be accelerated toward the carbon foil by VA. (3) These ions are carried through the post-acceleration region into the time-of-flight system where they can trigger a real TOF signal.

6.3.3.1 Signatures of ion-induced electron emission

In the data, these events will be characterized by their highly specific TOF values, i.e., the flight time will correspond to an ion that possesses only the speed gained in the post-acceleration region, accounting for any losses experienced during the transit of the carbon foil. Since the ions are not filtered by the ESA, their measured TOF signals do not correspond to the E/q steps of the instrument, and therefore appear in the data at all E/q values. ACE-SWICS double-coincidence data is shown in Figure 6.8a, where a vertical streak can be seen in a histogram of E/q vs. TOF. This streak appears at TOF values around 48 ns, which corresponds to measurements of desorbed H^+ in SWICS that possess only the post-acceleration energy. This signal is enhanced during times of increased incident particle flux, as expected for an induced desorption process. This same time period is shown in Figure 6.8b for triple-coincidence data, and the vertical proton streak is absent, as expected. The energy gained in post-acceleration alone is usually not sufficient for desorbed H^+ to overcome the SSD energy threshold.

An experimental verification of this noise source can be made by adjusting the post-acceleration voltage and measuring the associated TOF shift in the desorbed peak. In ACE-SWICS, prior to 4 May 2000, a proton with only the energy gained from the 21.3 kV post-acceleration had a time of flight of 51 ns. After that date, the post-acceleration voltage was increased to 25.9 kV, which would give a proton a time of flight of 47 ns. One-year accumulations of ACE-SWICS data are shown for the years 1999 (Figure 6.9a) and 2007 (Figure 6.9b), covering pre- and post- voltage ramp-up. To eliminate the signal of the solar wind and most pickup ions, only data taken during ESA steps with $E/q > 40$ keV/e were used. A comparison between the two years shows one TOF peak that shifts from 51.3 ns to 47.6 ns, indicating an increase in the energy of the incident ion. The TOF values of this shifting peak coincide with those of a proton possessing only the energy gained by the post-acceleration voltage

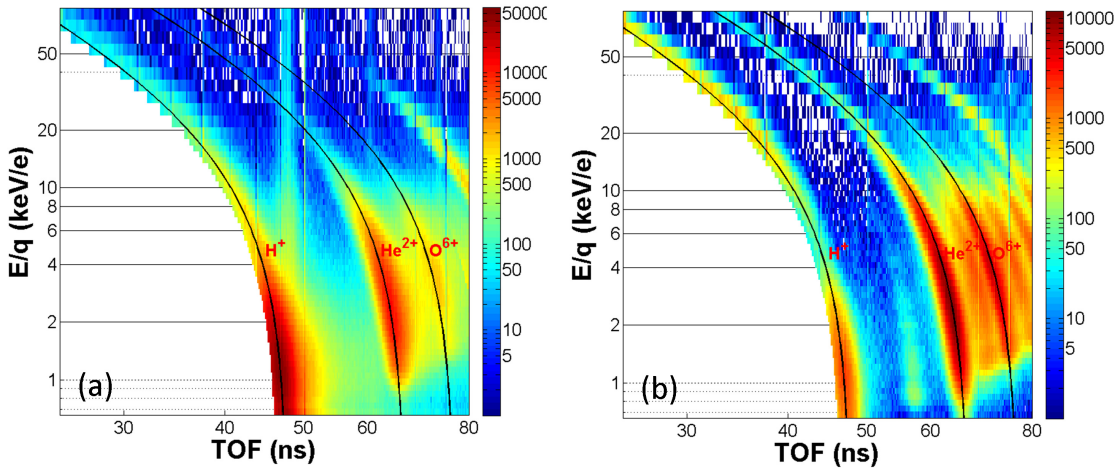

Figure 6.8: (a) SWICS double coincidence accumulations for 2004 showing the E/q-independent desorbed H^+ track centered on 48 ns. (b) SWICS triple-coincidence data for the same time period does not suffer from the desorbed proton background.

used during each of the respective years. The other peaks in Figure 10 did not shift when the voltage was changed, eliminating the possibility that they were caused by ions originating from outside of the TOF section.

6.3.3.2 Mitigation strategies for ion-induced electron emission

Similar mitigation strategies to those described in Section 6.3.2.3 also apply to reducing ion-induced electron emission. If a post-acceleration voltage is used, a triple-coincidence measurement combined with the energy threshold of a solid state detector can filter out particles desorbed in the post-acceleration region.

6.3.4 UV photons

MCP detectors can be triggered by direct photon impact or by photoelectrons released from the carbon foil or other surfaces. The hydrogen emission line of 121.6 nm, Lyman-α, is strongly emitted by the Sun, and has energy sufficient to create MCP signals that are indistinguishable to those of particles (*Wilhelm et al.* 1995). A

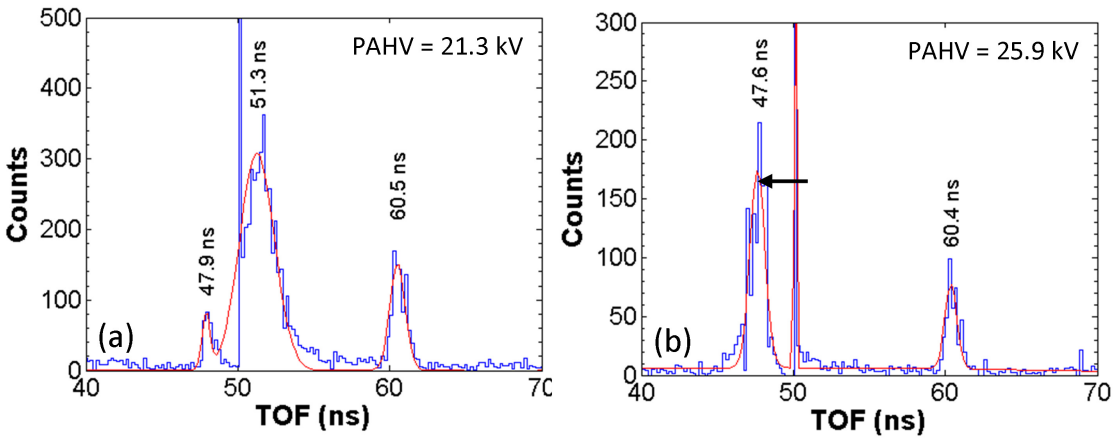

Figure 6.9: The TOF of a proton peak shifted from 51.3 ns to 47.6 ns after the post-acceleration high voltage (PAHV) of ACE-SWICS was increased. These protons possessed only the energy gained by the post-acceleration voltage, and were independent of the instruments E/q setting. Smaller peaks that remained stationary across the voltage increase are caused by other sources.

light leak in an instrument will raise the floor level of background noise at all E/q steps and in all TOF channels. These noise events from UV/EUV can overwhelm any real signal due to the extremely high solar photon flux: 3.5×10^{11} photons cm^{-2} s^{-1} at 1 AU (*Heroux and Hinteregger* 1978), and therefore a rejection rate of $> 1 : 10^{10}$ is desirable for particle sensors with apertures exposed to direct solar UV (*Mukherjee et al.* 2007).

6.3.4.1 EUV induced events

A study by *Gershman and Zurbuchen* (2010) used ray tracing simulations along with models for scattering and reflection off of various surfaces to calculate UV transmission through a model of the high mass resolution spectrometer (Mass) on the Wind spacecraft. They found that sufficient background suppression of photons required many reflections or scattering events before reaching the carbon foil, and was a strong function of material selection and coating.

EUV photons can be absorbed by the carbon foil, liberating an electron that may trigger a start signal (*Hsieh et al.* 1980). If the photon traverses the foil it may trigger a stop signal directly (*Gruntman* 1997). A single photon is therefore not able to trigger a double-coincidence event on its own. However, if the photon flux penetrating the foil is sufficiently high, accidental coincidence events can occur, as described in Section 6.3.5.

6.3.4.2 Mitigation Strategies

Historically, light traps and optical baffles have proven highly effective in reducing UV flux, as have serrated plates and black surface coatings (e.g., *Gloeckler et al.* (1985); *Zurbuchen et al.* (1995)). By creating serrated, or scalloped surfaces, specular reflection of photons is hindered and the scattering efficiency reduces by about a factor of three (*Carlson and McFadden* 1998). The SWICS instrument, which uses a combination of light traps, serrations, and black coatings, has an estimated background suppression of $1 : 10^{12}$ (*Gershman and Zurbuchen* 2010).

Some designs for energetic neutral atom detectors (*Gruntman* 1997; *Mitchell et al.* 2000) or high-energy ion detectors (*Andrews et al.* 2007), allow the foil to be directly exposed to incoming EUV photons. In these cases, solar UV has been suppressed by using thicker foils, foils with specialized coatings (*Hovestadt et al.* 1995), or nanoscale transmission gratings (*Scime et al.* 1995; *Gruntman* 1995). While increasing the thickness of the foil will result in a reduced UV transmission, it will also reduce the particle transmission and increase the degree of scattering. The use of nanogratings to block UV but permit particles to pass unimpeded has been tested successfully in flight (*Pollock et al.* 2000) and offers the possibility of lighter instrumentation, free from the necessity of mechanical light traps or baffles.

6.3.5 High incident flux

High incident ion flux can create background noise events with a couple of different signatures. The times periods tend to be sporadic, as the instrument geometrical factor is designed to limit incident flux to acceptable levels in normal operating conditions. While increased-flux events sometimes correspond to penetrating radiation (Section A), they can also pass through the ESA as designed, thereby limiting the noise signatures to specific E/q steps.

One effect of high incident flux is an increase in secondary electron emission from the foil at the TOF system entrance. In the E/q steps corresponding to the high flux (e.g., solar wind protons at 1 keV/e), the rate of ESD noise events will be enhanced as more electrons strike the start MCP and desorb ions from its surface.

High fluxes also can create conditions in the instrument in which two different ions can coincidentally trigger the start and stop signals for a single TOF measurement within a specified timing window. These accidental coincidences will be found in the data with a well-defined E/q, usually corresponding to the bulk ion species. The characteristic signature of accidental coincidence measurements is an exponential decay in the probability of an event occurring versus the TOF. This exponential decay of accidental background events is present even in times of low flux, but it can be shown analytically, that the exponential term leads to a nearly constant background level in this case. The following discussion of accidental coincidence events follows the derivation of *Bodmer* (1996).

6.3.5.1 TOF signatures of accidental coincidences

Assume that an incident ion penetrates the carbon foil and the resulting secondary electrons are guided toward the start MCP assembly. The detection efficiency of the start MCP assembly leads to a probability p_1 that the electrons will trigger a start signal. After a time τ, the ion will trigger a signal on the stop MCP assembly with some

probability p_2, either by direct impact in the case of a double-coincidence design, or by secondary electron emission from the target SSD in the case of a triple-coincidence design. If many ions of the same species pass through the system, a distribution of flight times $f(\tau)$ is expected due to energy straggling or angular scattering in the carbon foil (Section 6.3.6), the energy passband of the ESA, differences in secondary electron flight time, and electronics noise.

Certain instrument dependent timing limitations are assumed to be in place, such as the restriction for SWICS that no start signal can be linked to a stop signal that occurs outside of a timing window of $T_{\max}$. Furthermore, a start signal must be triggered to open a timing window before a stop signal will be considered valid. If the first stop signal in the timing window is used to determine the TOF, then the probabilities must be found for whether it is a real signal or an accidental coincidence. If $f(\tau)d\tau$ is the probability of a real stop signal detection between τ and $\tau + d\tau$, then we can write the probability of detecting a real stop signal before time $t = \tau$ as

$$F(\tau) = \int_0^\tau f(t)dt, \tag{6.1}$$

and the probability of not having a stop signal in that interval as $1-F(\tau)$. With an incident ion rate of $J_{\rm o}$, the stop signals due to accidental coincidences in the interval 0 to τ follow a Poisson distribution with a mean of $p_2 J_{\rm o}\tau$. Therefore, the probability of not finding an accidental coincidence in that time is $exp(-p_2 J_{\rm o}\tau)$. With this, we can write the probability distribution of no accidental coincidence signal occurring before time $t = \tau$, and the real signal being measured between τ and $\tau + d\tau$ as

$$f_{\rm r}(\tau)d\tau = e^{-p_2 J_{\rm o}\tau}\left(f(\tau)d\tau\right). \tag{6.2}$$

Similarly, the probability distribution of no accidental coincidence or real signal being detected before time $t = \tau$, and the accidental stop signal being measured

between τ and $\tau + d\tau$ is

$$f_{\mathrm{a}}(\tau)d\tau = e^{-p_2 J_{\mathrm{o}}\tau}\left(1 - F(\tau)\right)\left(J_{\mathrm{o}}p_2\right) \tag{6.3}$$

If an approximation is made of an infinitely small TOF resolution and $f(\tau)$ is normalized to p_2, the distribution $f(\tau)$ can be replaced by the delta function $p_2\delta(\tau - \tau_{\mathrm{o}})$. This will simplify the equations for the real and accidental probability distributions to

$$f_{\mathrm{r}}(\tau) = p_2\delta(\tau - \tau_{\mathrm{o}})e^{-p_2 J_{\mathrm{o}}\tau}, \tag{6.4}$$

and

$$f_{\mathrm{a}}(\tau) = J_{\mathrm{o}}p_2 X e^{-p_2 J_{\mathrm{o}}\tau}, \begin{cases} X = 1, & \tau < \tau_{\mathrm{o}} \\ X = 1 - p_2, & \tau_{\mathrm{o}} \leq \tau \leq T_{\max} \end{cases} \tag{6.5}$$

The accidental coincidence probability density (Equation 5) is plotted for the case of low ion flux (Figure 6.10a) and high incident flux (Figure 6.10b) with τ_{o}=90 ns. During times of low flux ($J_{\mathrm{o}} = 0.1$ ions s^{-1}), the background noise level due to accidental coincidence is near constant, as is typically seen in instrument data. When the flux through the instrument is extremely high, ($J_{\mathrm{o}} = 50$ ions s^{-1}) the exponential term has a stronger effect.

6.3.5.2 Mitigation strategies for high incident flux

To prevent high incident flux during routine measurement times, the geometric factor, which characterizes an instruments sensitivity and gathering power, must be designed appropriately. For regimes in which a low flux is expected, either a high geometric factor or a long integration time will be necessary to accumulate a statistically significant signal. However, this may make the instrument more susceptible to

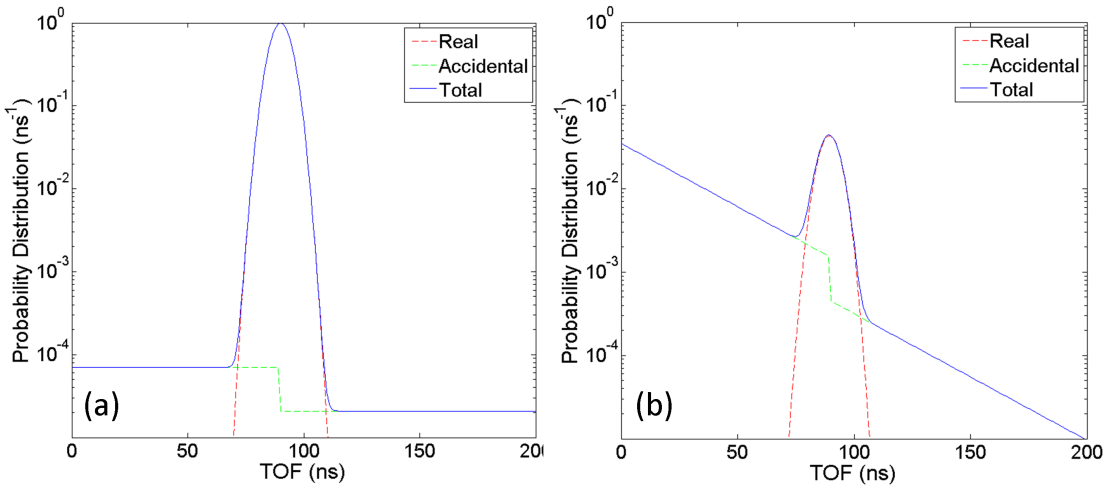

Figure 6.10: Modeled time-of-flight distribution of accidental coincidences for a low incident flux (a) shows a constant background level, while a high incident flux (b) takes an exponential form. Accidental coincidences in flight instruments most often manifest themselves as in the constant background case.

accidental coincidence events. Some sensors (e.g., *Rohner et al.* 2012) have employed an adjustable geometric factor, which allows a reduction in ions reaching the TOF section during times of high incident flux.

In addition to the use of anti-coincidence detectors, as described in Section 6.3.1.2, the likelihood of accidental coincidence events drowning out the signal in the data is greatly decreased when the requirement of triple coincidence is put in place. Figure 6.11 shows single-, double-, and triple-coincidence measurements from ACE-SWICS during a strong ICME in the year 2000. The counts collected in the start MCP signal are shown in Figure 6.11a, binned by their E/q step. Figures 6.11b and 6.11c show the double- and triple-coincidence counts, respectively. The levels of background noise events were quite high during this time, but the signal of the real data can most clearly be seen in the triple-coincidence measurement. While the double-coincidence data still provides interesting information on the plasma environment during this event, the resolution is much lower due to the increased background noise.

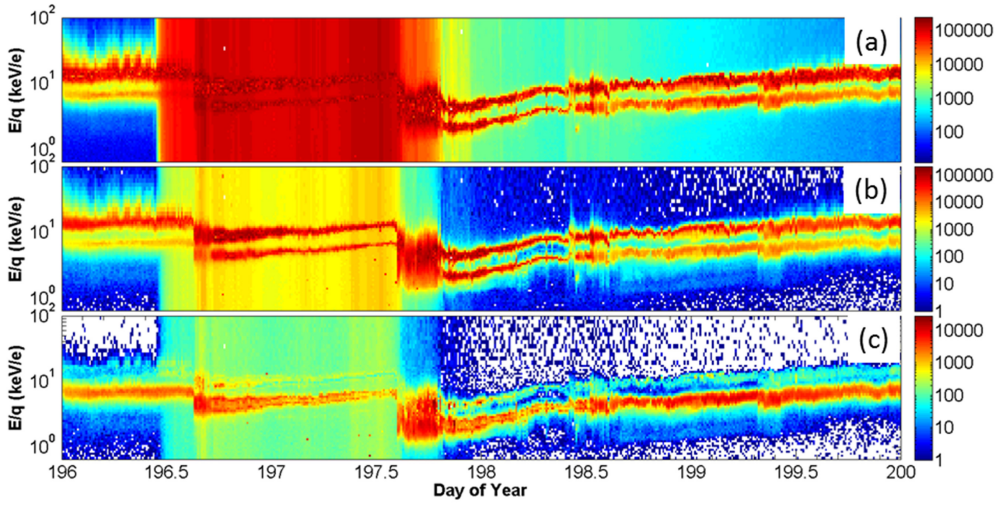

Figure 6.11: ACE-SWICS (a) single- (b) double- and (c) triple-coincidence counts during a large ICME event in 2000. The background level in the triple-coincidence data is significantly lower than the other two.

6.3.6 Energy straggling

The energy lost by ions as they pass through the foil can be derived from experimental data (*Ipavich et al.* 1982; *Allegrini et al.* 2006), or can be calculated using the Stopping and Range of Ions in Matter (SRIM) software (*Ziegler et al.* 2010; *Gilbert et al.* 2012). The distribution of energy loss that an ion experiences can be easily seen in Figure 6.12 which shows results from an SRIM Monte Carlo simulation of 1 keV hydrogen atoms going through a 1.5 μg cm^{-2} (67 $\mathring{A}$ thick) carbon foil. A tail of high energy loss can be seen in which collisions in the foil have greatly de-energized, and thus decelerated, the ions. These ions will pass through the TOF system with a range of measured flight times, corresponding to their decreased speeds. Since they pass through the ESA as intended, these decelerated ions will be found in the data with their typical E/q values, but will span a range of measured times of flight, leading to calculated M/q values that are unrelated to the identity of the original ion.

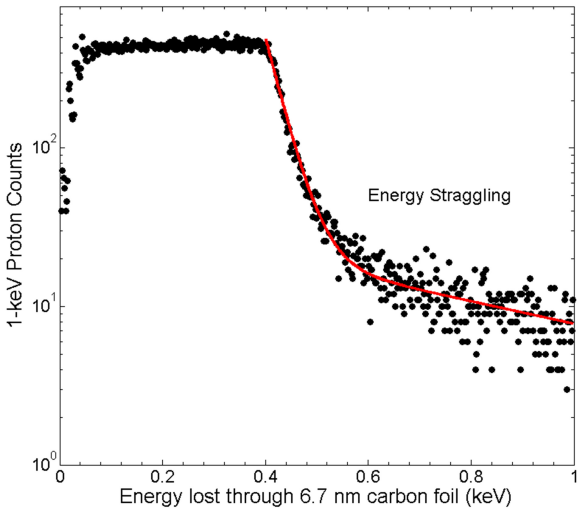

Figure 6.12: Energy loss for 1-keV hydrogen atoms that pass through a 1.5 μg cm^{-2} carbon foil. A tail of high energy loss is visible in the distribution.

6.3.6.1 Signature of energy straggled ions

The effects of energy straggling become apparent in the data when it is viewed as a two-dimensional distribution of E/q vs. TOF. In this parameter space, ions with the same M/q fall in curved tracks; shown in the figure are black lines indicating, from left, H^+, He^{2+}, O^{6+}, and He^+. After being filtered by their E/q, the bulk of incident ions will pass through the carbon foil and trigger a TOF measurement that falls along the calculated tracks. Those in the tail of high energy loss, however, will have slower TOF values. Double-coincidence (Figure 6.13a) and triple-coincidence (Figure 6.13b) measurement techniques of data from a 9 January 2005 ICME both show the effects of energy straggling. These are evident as horizontal data tracks, covering a range of TOF values within the expected E/q step, and overlapping the calculated tracks of all heavier M/q values. Clear in the figure are the energy-straggled ion tracks of H^+ and He^{2+}, but the effect is visible for other ions when the counts are high enough. Also apparent in the figure is a discontinuity due to an onboard priority scheme for SWICS that mitigates the signals of protons and alpha particles for $M/q \geq 3.27$ amu/e to allow a better identification of higher M/q ions.

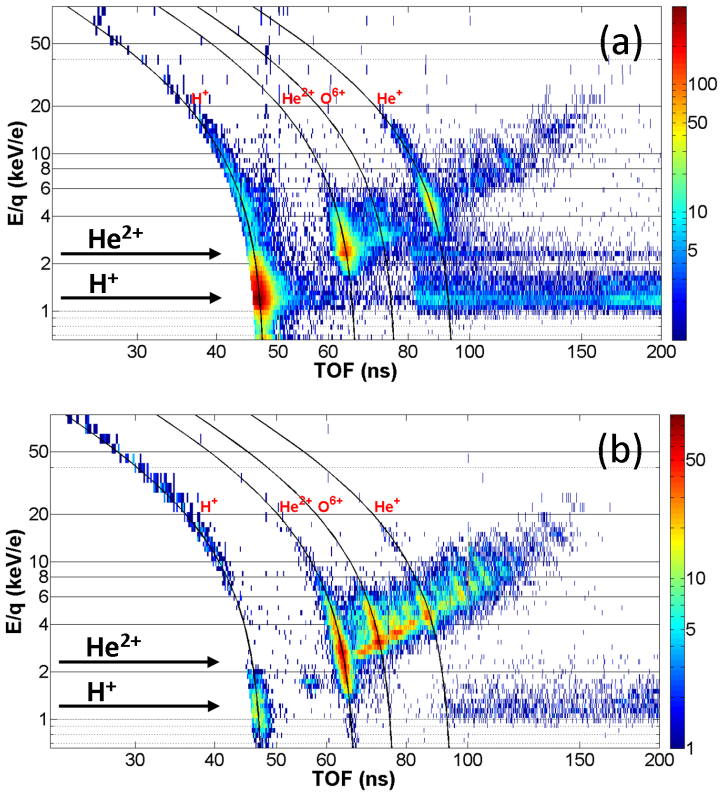

Figure 6.13: ACE-SWICS data during an ICME from January 2005. Energy straggling of H^+ and He^{2+} is evident in the E/q vs. TOF plot as horizontal streaks at consistent E/q values, with longer times of flight than the bulk of the species. The effects of energy straggling can be seen in both (a) double-coincidence and (b) triple coincidence data. A discontinuity due to an onboard priority scheme for SWICS that mitigates the signals of protons and alpha particles for $M/q \geq 3.27$ amu/e to allow a better identification of higher M/q ions.

Energy straggling effects can be easily distinguished from noise due to accidental coincidence events. Since ions do not accelerate through the carbon foil, the energy straggling effects will only cause ions to have longer times of flight than the bulk ions, while accidental coincidences will not be restricted to any TOF range. The lack of a horizontal data streak to the left of the He^{2+} peak in Figure 6.13 is a strong indication that the streak to the right is due to energy straggling rather than accidental coincidence effects.

6.3.6.2 Mitigation strategies

The use of carbon foils for secondary electron generation in straight-through TOF systems leads to an expected and unavoidable background signal due to energy straggling. Because it is restricted to specific E/q steps and TOF values, this noise can be reduced or avoided when post-processing the data. The issue can be avoided altogether, however, by using an energy-isochronous design (*Hamilton et al.* 1990; *McComas and Nordholt* 1990) in which ion trajectories undergo half of a harmonic oscillation in the linear electric field of the TOF system. For ions that exit the carbon foil with a positive charge, the times of flight are independent of the ion energy (*Gilbert et al.* 2010) and energy straggling will therefore have no effect on the mass resolution. However, ions whose trajectories are not reversed by the linear electric field, such as those who exit the carbon foil with neutral or negative charges, will still suffer from energy straggling.

6.3.7 Electronics

Noise events in the data can also be caused by issues with the flight electronics. There are many possible sources of noise events in the electronics, and it is not possible to list them all, as many are instrument-specific. An example from SWICS will serve here as an illustration of how noise events from electronics can affect an accurate

interpretation of the data when not properly characterized.

6.3.7.1 Signatures of electronics noise

In SWICS, periodic vertical streaks can be seen in plots of E/q vs. TOF in both double- and triple-coincidence data. Fourier analysis indicates a periodicity in channels numbered 2^n (n=1:10) and at every linear combination of 2^n. This is likely due to the sample and hold analog-to-digital (ADC) converter used to compute TOF values. A slight discharge of the hold capacitor during the ADC conversion can create a pileup of events from adjacent TOF channels into the nearest 2^n valued channel number. This pileup creates a spike at the TOF $=2^n$ channel and a deficit in the surrounding channels. The conversion of TOF channel into nanoseconds is TOF = 200/1023·(channel number). The largest disturbances occur at channels 256 (28), 512 (29), and 768 (28 + 29), corresponding to TOF values of 50.05, 100.1, and 150.15 ns, respectively.

6.3.7.2 Electronics noise effects on data analysis

Noise events from onboard electronics can have a significant effect on data analysis of species with low countrates. Figure 6.15 illustrates how the ADC-induced counts skew data analysis by creating an artificial peak corresponding to high-energy O^+ pickup ions. When double-coincidence data from days 150-180 in 2007 are examined, an artificial peak in TOF channel 29 (100.1 ns) is apparent at all E/q steps (Figure 6.15). The calculated trace of O^+ crosses this large electronic noise peak between E/q = 60-75 keV/e. When the O^+ data are ordered by their E/q step, there is an apparent increase in high-energy ions, when in fact the counts are an artifact of the electronics (Figure 6.15b).

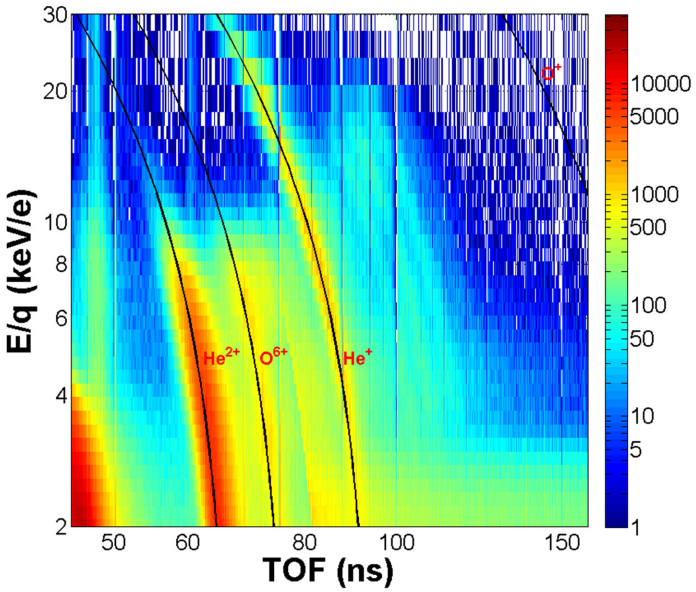

Figure 6.14: The ADC-induced vertical streaks are visible in SWICS TOF data from 2007. The effect is strong at TOF values of 50, 100, and 150 ns, consistent with TOF channels that are a superposition of 2^n values.

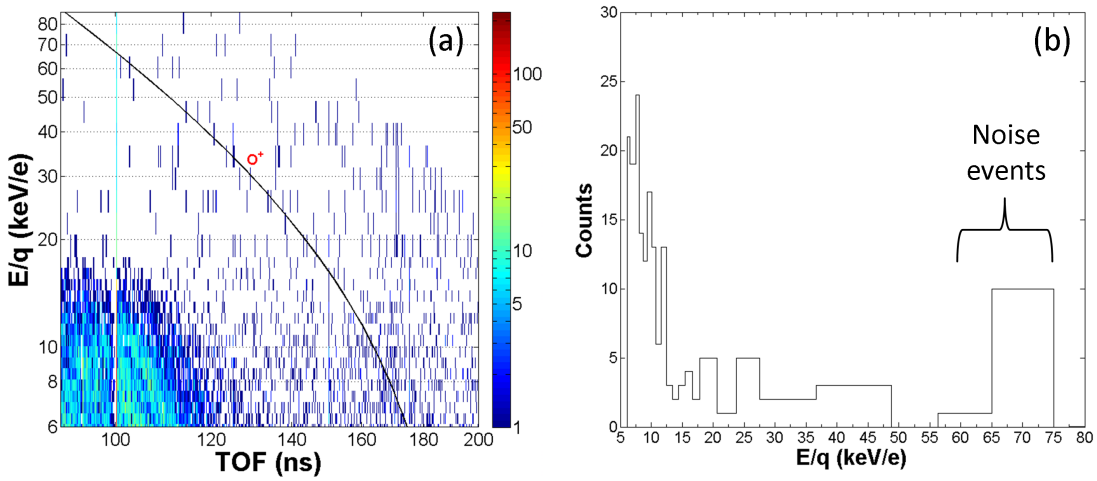

Figure 6.15: An example from 30 days of 2007 SWICS of how electronic noise events will skew data for low-count studies. O^+ is mapped by a solid black line. (a) The strong peak in TOF channel 512 (100.1 ns) crosses the O^+ calculation near E/q = 60-75 keV/e. (b) The O^+ measurements are ordered according to their measured E/q step. A large electronic noise peak between E/q = 60-75 keV/e is revealed.

6.3.7.3 Mitigation strategies for electronics noise

The elimination of electronics noise in-flight is difficult and extremely instrument-specific. Such issues are best addressed in the design, testing, and calibration of instruments in a controlled laboratory setting.

6.4 Conclusions

The mechanisms behind a number of noise sources in TOF spectrometers have been identified, as well as their corresponding signatures in instrument measurements. The variety, subtlety, and complexity of these processes illustrates that the design of an effective spaceborne TOF sensor is far from trivial. Not only must the instrument accurately analyze incident ions, but it must be designed to withstand a harsh and unpredictable space environment that is a unique and often overpowering source of measurement contamination. The sources of noise events discussed in this paper apply to any TOF spectrometer, with many also more widely applicable to other types of particle detecting instruments, e.g. single-coincidence instruments, such as electron detectors or bulk plasma sensors.

Careful design principles such as UV suppression, in-flight bake-out, and the avoidance direct paths for MCP-to-MCP ion feedback can reduce the background noise level seen in the data. However, one of the greatest noise suppression techniques is the use of triple-coincidence detection, a strategy that has enabled sensors like SWICS to provide extremely sensitive data sets of low-flux solar wind heavy ions. Using a small form factor design without triple-coincidence capabilities, such as for FIPS, can lead to additional unexpected sources of noise, some of which are only revealed post-launch.

It is not feasible, nor necessary, to eliminate all noise sources in the design and flight stages of the instrument. In the companion paper to this one, *Gershman et al.*

(2012a) describe noise removal techniques for the post-processing of flight data. These techniques are applied to the FIPS sensor on MESSENGER, demonstrating that valuable science can still be achieved with a compact, double-coincidence instrument. Although *Gershman et al.* (2012a) uses FIPS as a case study, the framework and processing techniques discussed are completely generalizable, and can applied to any TOF instrument to assist with noise forward modeling, and in some cases, removal.

CHAPTER VII

Noise forward modeling and removal through post-processing in space-based time-of-flight sensors

This chapter is taken from: Gershman, D. J., J. A. Gilbert, J. M. Raines, G. Gloeckler, S. C. Solomon, and T. H. Zurbuchen (2012), Background noise in space-based time-of-flight sensors - Part II: Noise event forward modeling and removal through post-processing, *Space Sci. Rev.*, in prep.

Abstract

This paper develops and implements a mathematical framework that enables noise modeling and removal for spaceborne time-of-flight mass spectrometers. Using a kernel density estimation-based Monte Carlo event-processing algorithm, instrument measurements are probabilistically separated into real and noise-based events. Given an accurate noise model, the overall misidentification of events is expected to be less than 25% even for datasets with only few events or low signal-to-noise (SNR) ratios, with significantly improved results expected for progressively larger accumulations of data. These techniques are applied to data from the MESSENGER/FIPS instrument and are shown to successfully recover heavy ion events from in-flight data around the

magnetosphere of Mercury. Such data analysis methods not only enable a more in-depth understanding of sensor operation, but also provide a unique post-processing approach to the improvement of in-flight SNR for any similar sensor without any modifications to instrument settings.

7.1 Introduction

Noise events in spaceborne time-of-flight mass spectrometers (TOF-MS) are undesirable measurements that can arise from a variety of sources, ranging from background processes to incident particle events. These erroneous events contaminate mass and energy spectra, making it difficult to recover accurate ion composition and abundances, especially for heavy ion species whose measured event rates can be low. A number of these noise sources, as well as their effects on data analysis are introduced and discussed in detail in *Gilbert et al.* (2012a), i.e., part I of this study. This paper (part II) focuses on modeling and probabilistic removal of these events, such that the signal-to-noise ratio (SNR) of measured datasets can be increased using software post-processing techniques, rather than adjusting in-flight instrument settings. The techniques described here were used to enable the scientific analyses of *Zurbuchen et al.* (2011) ,*Gershman et al.* (2012d), *Gershman et al.* (2012b), and *Raines et al.* (2012) using observations from the Fast Imaging Plasma Spectrometer (FIPS) (*Zurbuchen et al.* 1998; *Andrews et al.* 2007) on the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft (*Solomon et al.* 2001).

There are two general classes of noise contamination: additive and multiplicative/distortion. Additive noise is an unwanted signature superimposed on top of a real signal (*Tuzlukov* 2002). The most common, and perhaps ubiquitous, form of this contamination is white noise (*Kuo* 1996), a constant spectral density and Gaussian-

distribution of amplitudes overtop an audio, electronic, or image signal. For any additive noise source, however, the underlying signal remains unchanged. Distortion, on the other hand, convolves the uncontaminated signal with a noise function, altering its original properties (*Vaseghi* 2008). A simple example of distortion is image blurring, where the color in a particular pixel is changed based on properties of its neighbors (*González and Woods* 2008).

The vast majority of noise events in a TOF-MS are additive. Although, as will be discussed, the intensity of some noise sources will scale with the number of real events, these noise events are nonetheless superimposed on the measured distribution of incident particles. Therefore, TOF-MS noise removal typically only requires distribution separation rather than dataset deconvolution. Also, TOF-MS noise sources will be characterized using accumulations of instrument data, becoming well-known quantities. A known source is significantly easier to remove from a dataset than an unknown one, and consequently we will not require the use of complex blind signal separation techniques (*Comon and Jutten* 2010). Finally, the ability to post-process measured datasets enables the devotion of significant computational resources to the removal process, allowing each event to be analyzed individually and completely, as opposed to being limited by the real-time processing power of an on-board flight computer.

In the case of many audio and image processing applications, noise sources exhibit specific spectral signatures (*Smirnov* 1999). After a frequency analysis of a particular dataset, high or low pass filtering can be used to significantly reduce the contribution of many noise sources (*Davis* 2002). This strategy, however, may not be best suited for applications to TOF-MS measurements that are a set of discrete, multi-dimensional events whose properties must not be altered as a consequence of any noise reduction process. Instead, we will employ Monte Carlo techniques (*Kalos and Whitlock* 2009) to process instrument measured events. Monte Carlo techniques have been applied

to a wide variety of applications in fields such as statistical physics (*Binder and Heermann* 2010), biology (*Manly* 2007), and finance (*Glasserman* 2004). While these techniques vary in their implementation, they all use a set of sampled random numbers to aid in their computations. Here, the probability that a particular event is noise will be determined using an instrument noise forward model. These probabilities, in conjunction with randomly generated numbers will be used to process sets of TOF-MS observations.

The primary obstacle for accurate noise reduction for TOF-MS is low measured event rates. TOF-MS measurements are discrete datasets and often result in zero recorded events for a given time step, as opposed to an analog audio signal, or a digital image. In fact, some measured TOF-MS data will rarely form a smooth distribution function unless they are accumulated over long time periods. The challenge here is to reconstruct a smooth distribution function, sufficient for processing, based on only a few, discrete events.

Data from MESSENGER/FIPS, which is currently in orbit around Mercury, is used as an example for both noise forward modeling and removal processes. FIPS is a double coincidence TOF-MS, composed of an electrostatic analyzer (ESA) and a TOF section. As illustrated in Figure 7.1, charged particles are guided by an electric field between shaped electrodes of the ESA such that only ions with a particular energy-per-charge (E/q) successfully reach and penetrate a thin carbon foil, liberating a secondary electron. This electron is accelerated by electric fields in the TOF section and electrostatically reflected by a mirror harp assembly onto a microchannel plate (MCP) detector, where it creates a start signal. The incident particle, which is most often neutralized by the foil (*Young et al.* 2004), passes through the mirror harp assembly to another MCP detector, generating a stop signal, resulting in a correlated TOF event. The anode in the start MCP assembly is position-sensitive, such that valid events have five properties: time step t, energy-per-charge E/q, TOF, and start

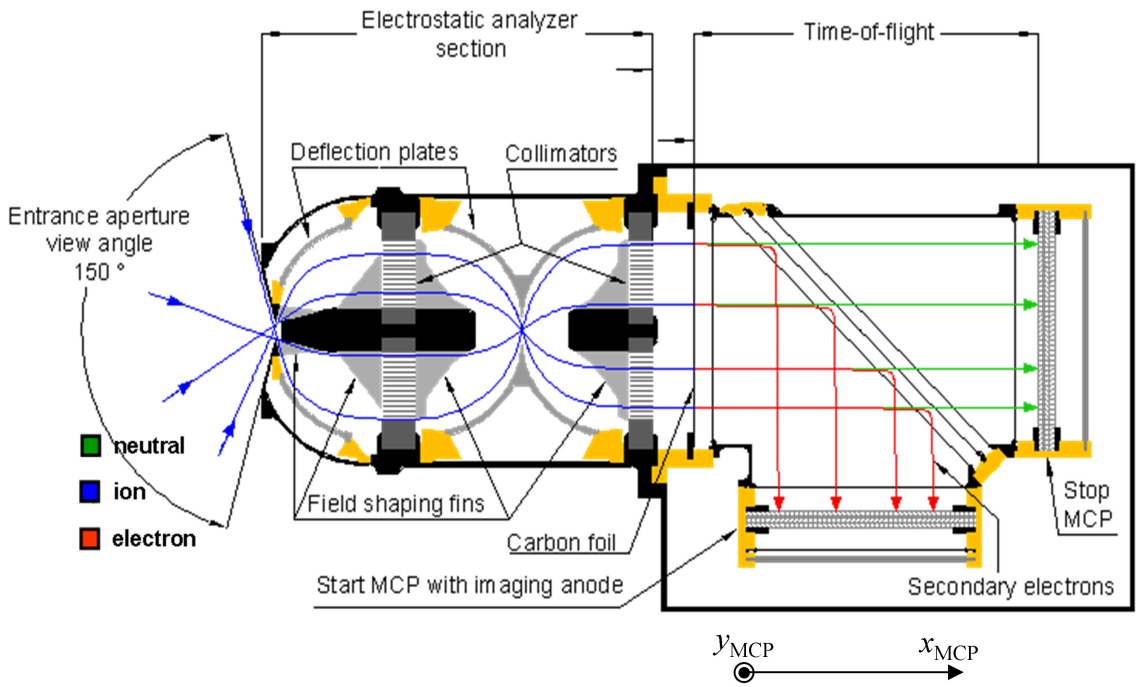

Figure 7.1: Illustration of the MESSENGER/FIPS TOF-MS adapted from *Andrews et al.* (2007). Ions with a particular range of E/q are filtered through an ESA and then analyzed by a TOF section with imaging capabilities on the start detector. The $(x_{\mathrm{MCP}}, y_{\mathrm{MCP}})$ coordinate system of the MCP is defined here.

MCP coordinates $(x_{\mathrm{MCP}}, y_{\mathrm{MCP}})$, with coordinate system geometry defined in Figure 7.1.

Although FIPS is used as a case study, the techniques developed here should be readily generalizable to any TOF-MS. In Section 7.2, we develop a mathematical framework for TOF-MS measurements as well as a noise removal process suitable for datasets with varying SNR and low total count rates. In Section 7.3, we discuss a method to create a complete instrument noise forward model using accumulations of measured data, using MESSENGER/FIPS as an example. Finally, in Section 7.4, we apply the noise forward modeling and removal processes to successfully recover measured heavy ion events in and around the magnetosphere of Mercury from MESSENGER/FIPS in-flight observations.

Table 7.1: Notation used for the mathematical framework of instrument noise modeling and removal processes. Here, Φ is a dummy variable used to represent any of several different quantities and distributions introduced throughout the paper.

Notation	Definition
Φ_{real}	Real signal distribution of Φ
Φ_{noise}	Noise signal distribution of Φ
Φ_{bg}	Background signal distribution of Φ
Φ_{inc}	Incident particle signal distribution of Φ
$\hat{\Phi}$	Normalized distribution of Φ
Φ_{m}	Φ derived from instrument measurement
Φ^*	Φ derived from noise removal processing

7.2 Mathematical framework

In order to model noise in a TOF-MS instrument, a mathematical framework for describing how an instrument measures events must be developed. Such a framework will enable comprehensive noise forward modeling and removal processes. The techniques described here are applicable to any TOF-MS. Table 7.1 introduces notation that is used to describe quantities throughout the paper. Additional subscripts/superscripts that appear throughout the paper that are not in Table 7.1 will be defined in the body of the text. The derivations here, as well as the noise processing developed later in this section are limited to one-dimension. However, as will be discussed, these results will apply to multi-dimensional datasets.

7.2.1 Instrument measured events

Let the distribution of all measured events with respect to some parameter, x, be defined as $f_{\text{m}}(x)$. In general, x is one of the measurable event parameters, i.e., t, E/q, TOF, etc. $f_{\text{m}}(x)$ necessarily can only take integer values, since it results from discretely measured events. The measured events that form $f_{\text{m}}(x)$ can be considered to be a set of random samples of some probability density distribution, $f(x)$. $f(x)$ can have non-integer values, but its total integrated value must be the same as for

$f_{\mathrm{m}}(x)$ such that, for discrete values of x,

$$\sum_{\mathrm{i}} f_{\mathrm{m}}(x_{\mathrm{i}}) = \sum_{\mathrm{i}} f(x_{\mathrm{i}}). \tag{7.1}$$

The instrument-measured $f_{\mathrm{m}}(x)$ is formed through a histogram of discretely measured events. In such an accumulation, each event corresponds to a Kronecker delta function (δ_{ij}) shifted to the appropriate histogram bin center, x_{i}. $f_{\mathrm{m}}(x)$ can be written in terms of a sum of measured events as,

$$f_{\mathrm{m}}(x) = \sum_{\mathrm{i}=1}^{N} \delta_{\mathrm{x,x_i}}, \tag{7.2}$$

where N is the total number of measured events and x_{i} is the histogram bin location of the i^{th} event.

For large numbers of events, $f_{\mathrm{m}}(x)$ and $f(x)$ are almost identical, as shown in Figure 7.2a, which shows a sampling of a distribution $f(x)$ for N = 10,000 events. However, since $f_{\mathrm{m}}(x)$ must necessarily have integer values, for small numbers of events, $f_{\mathrm{m}}(x)$ can differ significantly from $f(x)$, as shown in Figure 7.2b, which shows a sampling of only $N = 100$ events from the same distribution. Therefore, $f(x)$ can be thought of as the large N limit of $f_{\mathrm{m}}(x)$.

Distributions of events can be separated into two subsets: noise-based events, and real particle events, such that,

$$f(x) = f_{\mathrm{real}}(x) + f_{\mathrm{noise}}(x), \tag{7.3}$$

and

$$f_{\mathrm{m}}(x) = f_{\mathrm{m,real}}(x) + f_{\mathrm{m,noise}}(x). \tag{7.4}$$

The total number of noise based events is defined as η, with the total number of

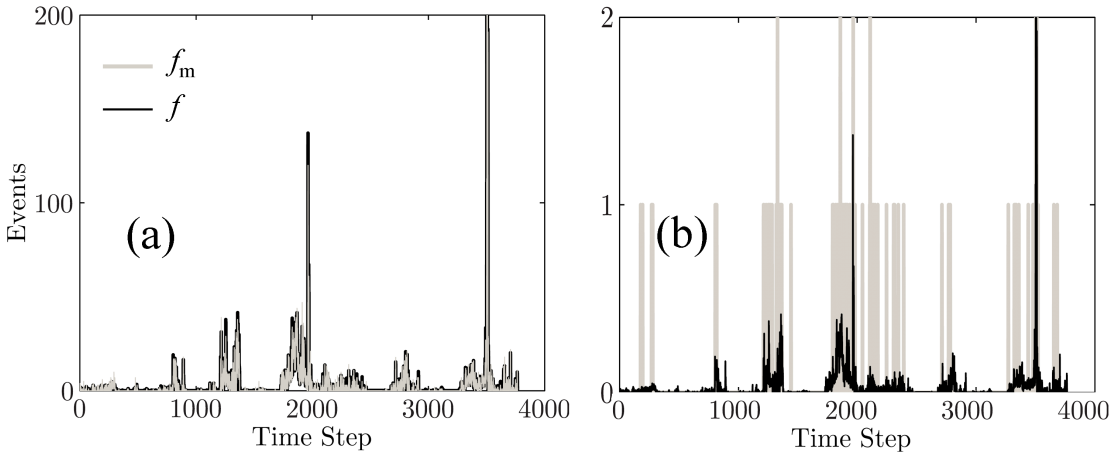

Figure 7.2: Comparison of measured distribution $f_{\mathrm{m}}(x)$ obtained from generating N samples from example distribution $f(x)$ for (a) $N = 10{,}000$ events and (b) $N = 100$ events. For small numbers of events there are large point-wise differences between $f_{\mathrm{m}}(x)$ and $f(x)$. For large N, $f_{\mathrm{m}}(x) \approx f(x)$. For each pair of distributions, $N = \sum_{\mathrm{i}} f_{\mathrm{m}}(x_{\mathrm{i}}) = \sum_{\mathrm{i}} f(x_{\mathrm{i}})$.

real events defined as S, with $S + \eta = N$.

As will be discussed in Section 7.3, a noise forward model will naturally produce the $f_{\mathrm{noise}}(x)$ distribution since average noise event rates, which can be non-integers, are used to generate the expected noise for a given time interval.

7.2.1.1 Signal to noise ratio estimation

In order to best characterize and understand the impact of noise sources on data analysis, a direct estimation of the instrument SNR is needed. For a given accumulation of data, the measured SNR_{m} is defined in equation (5) as the total number of measured real events, S, divided by the total number of measured noise-based events, η,

$$SNR_{\mathrm{m}} \equiv \frac{\sum_{\mathrm{i}} f_{\mathrm{m,real}}(x_{\mathrm{i}})}{\sum_{\mathrm{i}} f_{\mathrm{m,noise}}(x_{\mathrm{i}})} = \frac{S}{\eta} = \frac{N - \eta}{\eta} \tag{7.5}$$

As discussed previously, large accumulations of data are needed to have $f_{\mathrm{m,noise}}(x) \approx f_{\mathrm{noise}}(x)$. However, an accumulation over one or more dataset dimensions should be sufficient to achieve $\sum_{\mathrm{i}} f_{\mathrm{m,noise}}(x_{\mathrm{i}}) \approx \sum_{\mathrm{i}} f_{\mathrm{noise}}(x_{\mathrm{i}})$. In the case that this relationship holds, SNR_{m} is directly calculable and can be used to get a direct estimate of the signal-to-noise ratio for any data product using the total measured data, f_{m}, and forward modeled noise, f_{noise}.

7.2.1.2 Noise event probabilities

Following equation (7.3), the probability that a measured event with property x is a noise-based event can be written as,

$$P_{\mathrm{m,noise}}(x) = \frac{f_{\mathrm{m,noise}}(x)}{f_{\mathrm{m}}(x)}. \tag{7.6}$$

Likewise, equation (7.4) can be used to form the probability,

$$P_{\mathrm{noise}}(x) = \frac{f_{\mathrm{noise}}(x)}{f(x)}, \tag{7.7}$$

which is effectively the large N limit of equation (7.6).

If there is a known probability, $P_{\mathrm{m,noise}}(x)$ that a measured event with property x is noise, then a Monte Carlo processing technique can be used to separate a distribution into real and noise-based subsets as follows: For each measured event, a random number is uniformly generated between 0 and 1. Given the property x of each event, the corresponding probability $P_{\mathrm{m,noise}}(x)$ is compared with the value of the random number. If the generated number is smaller than $P_{\mathrm{m,noise}}(x)$, the event is flagged as noise-based. Otherwise, the event is considered to be real. After all events have been processed, recovered distributions $f^{*}_{\mathrm{m,noise}}(x)$ and $f^{*}_{\mathrm{m,real}}(x)$ are produced. It follows that an accurate estimate of $P_{\mathrm{m,noise}}(x)$ is vital for a successful noise removal process.

Consider several sets of N measured events randomly sampled from the distri-

bution $f_{\text{noise}}(x)$. Although each set originates from the same distribution, for small N, the point-wise differences between them will be significant. No set of events is necessarily better than another; they are simply independent samplings of the same distribution. It is therefore extremely difficult to precisely reproduce an instrument measured $f_{\text{m,noise}}(x)$ distribution from random samplings of $f_{\text{noise}}(x)$, as all possible combinations of samplings are computationally unreasonable to calculate. Therefore, $P_{\text{m,noise}}(x)$ will be extremely difficult to determine directly.

P_{noise}, however, can be readily computed because f_{noise} is directly produced from a noise forward modeling process (Section 7.3), and $f(x)$ can be accurately estimated for a given $f_{\text{m(x)}}$ (Section 7.2.2). Therefore, as an approximation, P_{noise} rather than $P_{\text{m,noise}}$ is used in order to create a computationally reasonable Monte Carlo processing algorithm. To take into account errors in the estimation of P_{noise}, this Monte Carlo technique can be iterative, where the dataset is continually reprocessed using updated values of $f(x)$ and f_{noise} until the total expected amount of noise has been removed, or until the algorithm converges. As detailed in Appendix C.1, using synthetic datasets to characterize algorithm accuracy, this process will necessarily remove the amount of noise predicted by f_{noise}, but may result in the misidentification of up to 20% of all events, regardless of SNR.

7.2.2 Kernel density estimation

In order to determine $P_{\text{noise}}(x)$, the distributions $f(x)$ and $f_{\text{noise}}(x)$ must be known. Noise forward modeling will produce $f_{\text{noise}}(x)$, and for large N, $f_{\text{m}}(x)$ naturally approaches $f(x)$. However, for small N, another technique must be used to convert the measured $f_{\text{m(x)}}$ to a reasonable approximation of $f(x)$. Kernel density estimation (KDE) (*Scott* 1992) is a technique that can accomplish this conversion for small N. In KDE, instead of using delta functions as in equation (2), each event is represented by an arbitrary normalized function, or kernel. Kernels that span multiple histogram

bins enable non-integer representation of measured events. A Gaussian kernel is a natural choice for TOF-MS systems since measured events often accumulate to form a set of Gaussian-like peaks due to their Normally-distributed error in the high count limit.

A KDE-based distribution of measured events, $f_{\mathrm{m,KDE}}(x)$, can be written as a sum of normalized discrete Gaussians:

$$f_{\mathrm{m,KDE}}(x) = \frac{\sum_{\mathrm{i}=1}^{N} e^{-\left(\frac{x-x_{\mathrm{i}}}{w(x)}\right)^2}}{\sum_{\mathrm{i}=-\infty}^{\infty} e^{-\left(\frac{x_{\mathrm{i}}}{w(x)}\right)^2}}. \tag{7.8}$$

Here, w is the bandwidth of the Gaussian kernel whose value must be tuned to a particular type of dataset. In general, w can be a function of x. Proper selection of bandwidth $w(x)$ can be difficult and is distribution dependent (*Silverman* 1986). This choice is further complicated by the fact that the distribution of TOF-MS data changes with time. For this reason, $f_{\mathrm{noise}}(x)$ is used to aid bandwidth selection, since the measured distribution must encompass the noise distribution. The choice of $w(x)$ here is not found analytically, but rather empirically, to provide a reasonable estimation of $f(x)$ guided by measured instrument data.

Our empirical bandwidth will be composed of two parts. First, we recognize that integer values of expected noise, on average, should be accompanied by a corresponding number of measured events. The fractional part of $f_{\mathrm{noise}}(x)$, however, will likely be accompanied with either zero or one measured events. Therefore, the kernel bandwidth should be based off of the fractional part of $f_{\mathrm{noise}}(x)$, $f_{\mathrm{noise}}(x) - \lfloor f_{\mathrm{noise}}(x) \rfloor$, where $\lfloor\ \rfloor$ denotes the integer part of the enclosed quantity (*Graham et al.* (1994)). This factor is well suited for setting bandwidths for datasets composed almost entirely of noise, i.e., . In this case, following equation (7.8), the amplitude of the Gaussian kernel for a single event will be approximately equal to the noise distribution, such

that . However, for distributions with high SNR, the data will tend to overspread, resulting in a poor estimate of $f(x)$. To compensate for this, the factor $1/(1+SNR_{\mathrm{m}})$ is used to reduce the overall spreading of the measured events for increased signal strength. Together, these factors form the empirical bandwidth,

$$w(x) = \frac{1}{f_{\mathrm{noise}}(x) - \lfloor f_{\mathrm{noise}}(x) \rfloor} \cdot \frac{1}{1+SNR_{\mathrm{m}}}. \tag{7.9}$$

Equation (7.9) is meant to be generally applicable, suitable for datasets with both strongly localized and uniformly distributed events and a variety of SNR. Further modifications to this bandwidth may be required to best fit a particular dataset of interest.

A metric that determines how well a generated distribution matches the true distribution $f(x)$ is the relative error between them. The relative error, E, is defined as,

$$E(f, f_{\mathrm{m}}) \equiv \frac{\sum_{\mathrm{i}} |f(x_{\mathrm{i}}) - f_{\mathrm{m}}(x_{\mathrm{i}})|}{\sum_{\mathrm{i}} f(x_{\mathrm{i}}}, \tag{7.10}$$

where $f(x)$ is the true distribution and $f_{\mathrm{m}}(x)$ is the measured distribution. Here, the sum over i only represents the portion of the dataset where there are measured events, since those are the only steps where P_{noise} will be computed.

Some individual steps (x_{i}) will have a very high local SNR, for example when a real event is measured at a step with almost no expected noise. In this case, the bandwidth $w(x_{\mathrm{i}})$ from equation (7.9) will be large, causing overspreading such that $f_{\mathrm{m,KDE}}(x_{\mathrm{i}})$ will be very different from $f(x_{\mathrm{i}})$. While the factor $1/(1+SNR_{\mathrm{m}})$ can help to mitigate this overspreading, high local SNR is still possible for steps with overall SNR_{m} close to zero. However, in such cases the ratio $f_{\mathrm{noise}}(x_{\mathrm{i}})/f_{\mathrm{m,KDE}}(x_{\mathrm{i}})$ can still be very small, leading to $P_{\mathrm{noise}}(x_{\mathrm{i}}) \approx 0$, as desired. The most sensitive SNR

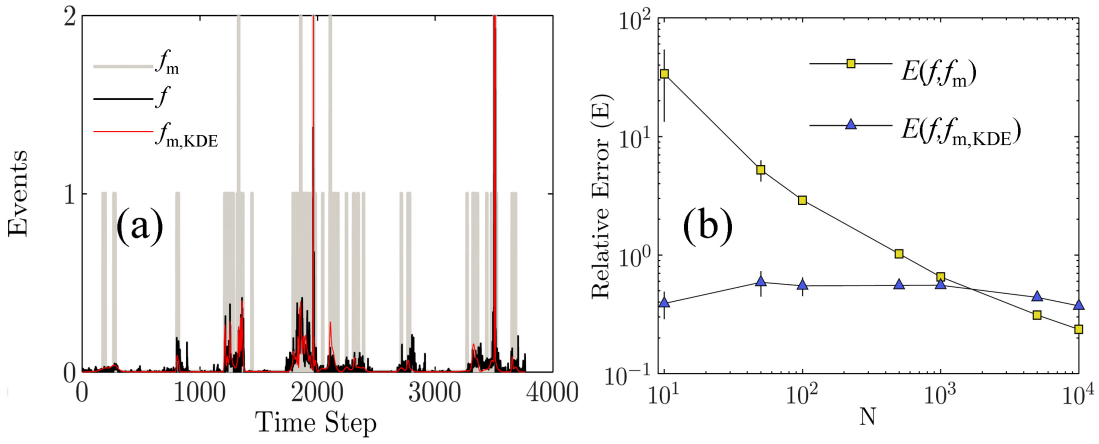

Figure 7.3: (a) An example distribution $f(x)$ with corresponding f_{m} and $f_{\mathrm{m,KDE}}$ distributions for $N = 100$ and $SNR_{\mathrm{m}} = 0$. (b) $E(f, f_{\mathrm{m}})$ and $E(f, f_{\mathrm{m,KDE}})$ as a function of N, showing that KDE techniques provide better estimates for f for small N, with a relative error nearly independent of N. Ten sets of samples were obtained from f for each data point. The mean and standard deviation of E for the 10 sets are shown for each as the data marker and error bars, respectively.

regime for noise removal will therefore be $SNR_{\mathrm{m}} \to 0$, where it is imperative that $f_{\mathrm{m,KDE}}(x_{\mathrm{i}}) \approx f(x_{\mathrm{i}})$.

Figure 7.3 shows the distribution f(x) from Figure 7.2 with corresponding $f_{\mathrm{m,KDE}}$ with $SNR_{\mathrm{m}} = 0$. $f_{\mathrm{m,KDE}}$ clearly creates a better estimation of $f(x)$ than the samples forming $f_{\mathrm{m}}(x)$. To quantify this improvement, $E(f, f_{\mathrm{m}})$ and $E(f, f_{\mathrm{m,KDE}})$ are computed for various N values as shown in Figure 7.3b. For large N, the KDE distributions have a slightly higher relative error than f_{m}, but for small N, are more accurate by 2 orders of magnitude.

In the case of noise modeling for TOF-MS, KDE techniques can be applied for any accumulation, such as time history, TOF, or incident particle direction datasets. For the incident particle direction datasets, the bandwidth can apply to a two-dimensional Gaussian kernel. In general, KDE techniques can be applied to higher dimensioned datasets (*Silverman* 1986), with each dimension permitted its own characteristic

bandwidth. The visualization of such techniques becomes difficult and in the case of TOF-MS, the determination of appropriate bandwidths may become intractable. For this reason, KDE techniques for TOF-MS may be practically limited to lower dimensional datasets. As detailed in Appendix C.2, using KDE techniques with a Monte Carlo event processing algorithm should result in the misidentification of less than 25% of all events, regardless of SNR, only a modest increase in error when compared with using a known $f(x)$ instead of $f_{\mathrm{m,KDE}}(x)$.

7.3 Forward modeling of instrument noise

A good estimate of $f_{\mathrm{noise}}(x)$ is vital for an accurate noise removal. Furthermore, even in the absence of a suitable approximation for $f(x)$, $f_{\mathrm{noise}}(x)$ will provide a direction estimation of instrument SNR. The complete $f_{\mathrm{noise}}(x)$ distribution should have contributions from any and all noise sources in an instrument. Here, MESSENGER/FIPS is used as a baseline, where both background and active noise sources will be discussed and modeled appropriately to demonstrate how to generate $f_{\mathrm{noise}}(x)$ from a set of accumulated observations.

Figure 7.4 shows an unfiltered 50-day accumulation from the day-of-year (DOY) range of 100-149 in 2011 of all MESSENGER/FIPS measured events. The E/q is plotted as a function of TOF, with curved tracks corresponding to real events superimposed with various noise sources. This time period of data will serve as the calibration dataset that will be used to characterize the behavior of f_{noise}. Data from the E/q steps marked as Background will be used to characterize the background noise rates and distributions. The distributions of events from the $E/q - TOF$ steps marked as Protons will be used to model the contamination of the H^+ mass peak into the peaks of higher mass-per-charge ratio (m/q) species. Finally, the events within the $E/q - TOF$ steps marked as ESD will provide the rates of events triggered by increased desorption of material from the start MCP at times of high incident particle

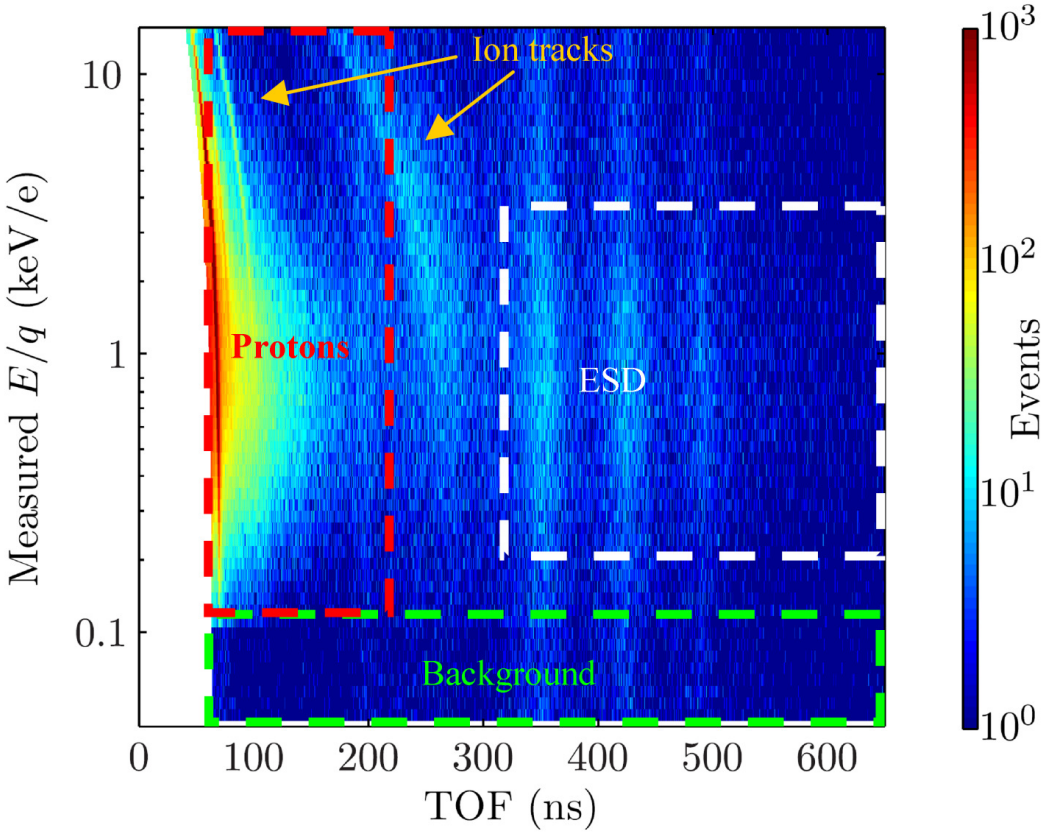

Figure 7.4: (a) $E/q - TOF$ spectrogram of 50 days (2011 100-149) of on-orbit MESSENGER/FIPS heavy ion ($m/q > 1$ amu/e) measurements. Curved tracks correspond to incident ions. Boxes marking the $E/q - TOF$ steps corresponding to background events, proton tail events, and induced ESD events are indicated as Background, Protons, and ESD, respectively, and will be used to characterize each noise source.

flux. Each of the active noise sources will be extrapolated to the TOF region of 200ns 300ns, where neither source is dominant.

7.3.1 Passive noise sources

Background, or passive noise, is constant in time and independent of E/q. However, such noise may have specific TOF and MCP position dependence. Consider as an example, electron stimulated desorption (ESD) processes in the MESSENGER/FIPS instrument. Two background electron sources have been identified, as

described in *Gilbert et al.* (2012a): emission from high voltage wire harps, and a localized field emission source near the edge of the MCP active area. A complete noise model must include distributions corresponding to each source.

The E/q steps marked as Background in Figure 7.4 are the lowest instrument energy steps in which significant numbers of real ion events are not expected, particularly outside of Mercurys magnetosphere. To further avoid possible contamination from any actual low energy events, only time steps where the ratio between the start MCP rate and valid event rate was greater than 10 are used, as real incident events will have ratios of $\sim$2 due to the efficiencies of ion detectors. Due to frequent drifts in instrument operating conditions, most notably ambient temperature, the background rate calculations are performed using daily accumulations of data. The background rates are the 'front rate, i.e. the background signal on the start MCP ($FSR_{\rm bg}$); the 'rear rate, i.e. the background signal on the stop MCP ($RSR_{\rm bg}$); the background proton event rate, i.e., background events with low TOF, ($PE_{\rm bg}$); and the valid event, i.e. double coincidence, background rate ($VE_{\rm bg}$). These rates (in units of $\sec^{-1}$) are used in various stages of the instrument noise forward model. The accumulation time in seconds for each time step, $\Delta(t)$, is readily calculable and can change depending on the operating mode of the instrument. The background TOF and MCP X-Y spectra are obtained using the full 50 day accumulated dataset, since they are not expected to change with time.

The total background distribution must be separated into its harp emission and other field emission subcomponents, denoted with subscripts 'he and 'fe respectively. A simple separation isolates events corresponding to the 'hotspot pixels (*Gilbert et al.* 2012a) in the background MCP X-Y distribution as the generalized field emission source, and attributes the rest to wire harp emission of electrons. Following *Gilbert et al.* (2012a), the TOF distribution of the harp emission source corresponds to the flight times of desorbed ions from the surface of the start MCP to the top of the

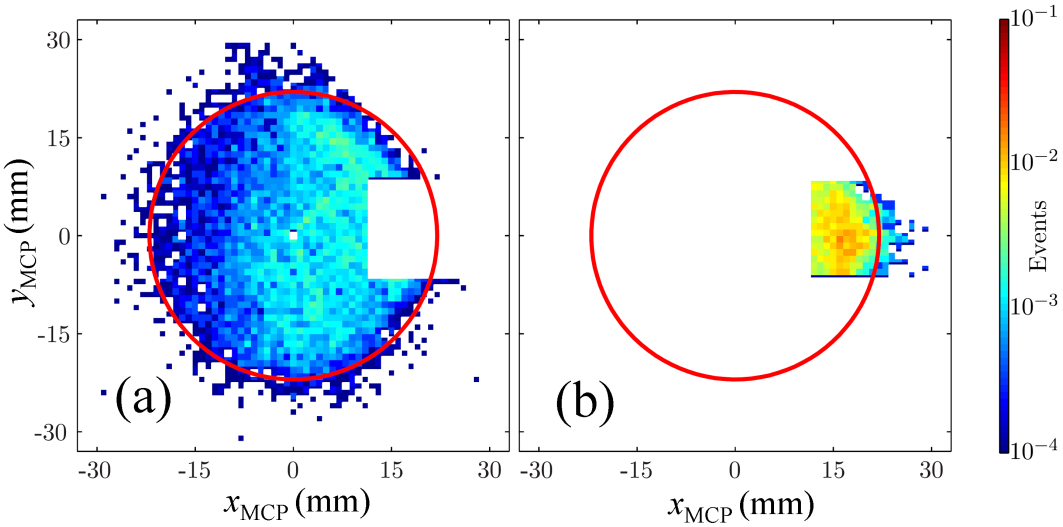

Figure 7.5: Normalized MCP X-Y distribution of background ESD events in MESSENGER/FIPS triggered by (a) harp emission electrons ($\hat{f}_{\text{he}}(x_{\text{MCP}}, y_{\text{MCP}})$) and (b) field emission electrons ($\hat{f}_{\text{fe}}(x_{\text{MCP}}, y_{\text{MCP}})$). The red circle indicates the boundaries of the MCP active area. The field emission electrons are concentrated in a beam forming a set of 'hotspot pixels near the edge of the active area while the harp emission electrons strike almost uniformly over the surface of the MCP.

TOF chamber. The TOF of the field emission source also includes a set of peaks at a constant fraction of the flight times of the harp emission source, indicating that desorbed species are striking the sides of the TOF chamber instead of the top. These accumulated distributions are smoothed to reduce statistical noise. The resulting distributions of MCP X-Y and smoothed-TOF for each source are shown in Figures 7.5 and 7.6, respectively.

The TOF background noise distributions $f_{\text{noise,fe}}(t, TOF, x_{\text{MCP}}, y_{\text{MCP}})$ and $f_{\text{noise,he}}(t, TOF, x_{\text{MCP}}, y_{\text{MCP}})$ can be written for a given accumulation of time from these distributions following equations (7.11) and (7.12),

$$f_{\text{noise,fe}}(t, TOF, x_{\text{MCP}}, y_{\text{MCP}}) = g_{\text{fe}} \cdot VE_{\text{bg}} \cdot \hat{f}_{\text{fe}}(TOF) \cdot \hat{f}_{\text{fe}}(x_{\text{MCP}}, y_{\text{MCP}}) \cdot \Delta(t), \quad (7.11)$$

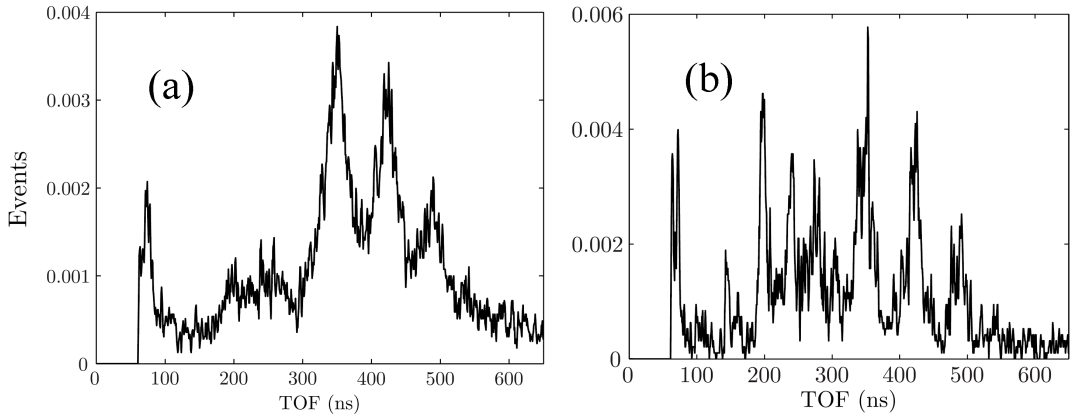

Figure 7.6: Normalized TOF distribution for background ESD events in MESSENGER/FIPS triggered by (a) harp electrons ($\hat{f}_{\rm he}(TOF)$) and (b) field emission electrons ($\hat{f}_{\rm fe}(TOF)$). The harp emission TOF distribution corresponds to the flight times of desorbed species from the surface of the start MCP to the top of the TOF chamber. The field emission events include a contribution from desorbed ions that strike the side of the TOF chamber, leading to flight times at a constant fraction of the corresponding harp emission distribution peaks.

and

$$f_{\rm noise,he}(t, TOF, x_{\rm MCP}, y_{\rm MCP}) = g_{\rm he} \cdot VE_{\rm bg} \cdot \hat{f}_{\rm he}(TOF) \cdot \hat{f}_{\rm he}(x_{\rm MCP}, y_{\rm MCP}) \cdot \Delta(t). \quad (7.12)$$

Here, $g_{\rm fe}$ and $g_{\rm he}$ are the fractions of contributions of both sources, where $g_{\rm fe}+g_{\rm he} = 1$. $VE_{\rm bg}$ is the total background valid event rate in $\sec^{-1}$, and $\Delta(t)$ is the accumulation time for a given time step in seconds. Other background sources of noise can be characterized similarly.

7.3.2 Active noise sources

Active sources create noise as a function of incident particles or photons. Any external stimulus capable of creating signals on the start and stop detectors is con-

sidered an active source. The sources considered here will be ion energy straggling and induced ESD, as discussed by *Gilbert et al.* (2012a). Accidental coincidences due to high flux or penetrating radiation will not be modeled due to the extremely low SNR of measurements during those time periods.

7.3.2.1 Ion energy straggling

Particles passing through a thin carbon foil lose some distribution of their kinetic energy, a phenomenon called energy straggling (*Allegrini et al.* 2006). This energy distribution results in a high TOF peak tail for each species. For a TOF-MS, H^+ will always have the smallest TOF, as it has the smallest possible m/q of 1 amu/e. The tail for H^+, therefore, has the potential to contaminate the TOF channels corresponding to all other ion species, as shown in Figure 7.4. Furthermore, for MESSENGER/FIPS, the most easily measured ion is H^+, which is present in both the solar wind and Hermean magnetospheric environments, and is typically orders of magnitude more abundant than other species (*Zurbuchen et al.* 2008). The rates of heavier ion events can be comparable to the magnitude of the proton peak tail, making it difficult to distinguish between the two sets of particles. Events in the $E/q - TOF$ steps in Figure 7.4 labeled as Protons are assumed to be dominated by the proton peak tail.

While it is possible to obtain an overall good fit to energy straggled TOFs using a kappa distribution (*Allegrini et al.* 2006), for heavy ion species, small errors in the fit can create relatively large errors in f_{noise}. Therefore, rather than use a functional fit for the entire proton peak shape, the accumulation of data from Figure 7.4 is smoothed, analogous to the formation of the background noise distributions in Section 7.3.1. The average proton peak shape is expected to change with time, as changes ambient temperature may affect the energy straggling process through the carbon foil, and the thickness of the carbon foil may not be spatially uniform, leading to different peak shapes for time periods where, on average, particles enter the instrument aperture

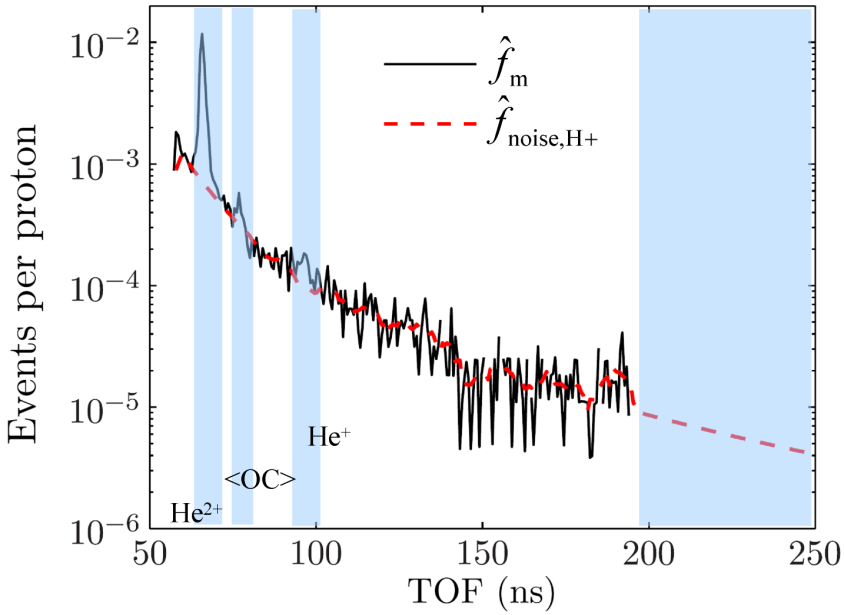

Figure 7.7: TOF distribution of proton tail events from E/q = 1.5 keV/e for 2011 100-109, normalized by the proton event rate. TOFs corresponding to solar wind ion tracks for He^{2+}, He^{+}, and $\langle OC \rangle$, the averaged solar wind heavy ion content (Gershman et al., 2012a), are indicated. The black solid curve is the set of normalized raw measurements . The red dashed curve represents a smoothed that has been interpolated through ion track TOFs, and extrapolated to higher TOFs with a power law, forming the distribution of .

from different directions. For MESSENGER/FIPS, the proton peak shape is updated using 10 day accumulations, providing a balance between statistical significance and the capturing this time variation.

The expected background spectra (calculated in Section 7.3.1) is subtracted away from this accumulation, and TOF channels corresponding to real ion tracks are excluded. The data is linearly interpolated through these data gaps, and extrapolated to higher TOF channels with a power law fit, consistent with a kappa TOF peak shape. An example TOF spectra for a 10 day accumulation at E/q = 1.5keV/e is shown in Figure 7.7, with a corresponding noise curve fit that interpolates through ion track locations and extrapolates to higher TOF values.

The amplitude of the H^+ tail should scale with the total number of measured proton events (PE) measured at lower TOFs, i.e. near the peak of the H^+ TOF distribution. The accumulated data in each energy step is therefore normalized by the total corresponding proton event rate to form $\hat{f}_{\mathrm{noise,H+}}(TOF, E/q)$. The errors in this quantity are greatest for energy steps in which the proton event rate is comparable to the background rate. However, since the expected noise from proton events in these steps will be low, this error is manageable. The X-Y distribution of this noise will be that of the incident protons, $\hat{f}_{\mathrm{inc,H+}}(E/q)$, which can be obtained by accumulating events close to the proton peak at each energy step. The expected f_{noise} therefore becomes,

$$\begin{aligned} f_{\mathrm{noise,H+}}(t, TOF, E/q, x_{\mathrm{MCP}}, y_{\mathrm{MCP}}) = (PE(t, E/q) - PE_{\mathrm{bg}}) \\ \times \hat{f}_{\mathrm{noise,H+}}(TOF, E/q) \\ \times \hat{f}_{\mathrm{inc,H+}}(x_{\mathrm{MCP}}, y_{\mathrm{MCP}}) \cdot \Delta(t). \end{aligned} \tag{7.13}$$

7.3.2.2 Induced ESD

In addition to electrons from field emission or harp emission sources, secondary electrons from incident particles can desorb ions from instrument surfaces, resulting in valid TOF events. For MESSENGER/FIPS, a number of ion species are desorbed from the start MCP and follow trajectories that lead to correlated stop signals. This effect occurs with some probability relative to an incident particle being detected. Therefore, the total number of these events should be proportional to the number of non-background start signals detected in a particular E/q step. Also, desorption rates will not only depend on the number of incident particles, but where those particles enter the instrument.

Assuming, as in Section 7.3.1 that a uniform distribution of harp electrons strikes

the MCP, the harp electron event distribution is proportional to the number of valid events per start signal in each pixel. Not every start signal will not have a corresponding stop signal, so the ratio of the valid event rate to start rate will always be less than one. The incident particle distribution ($\hat{f}_{\rm inc}$) should therefore correspond to the distribution of starts signals on the MCP. Multiplying $\hat{f}_{\rm inc}(x_{\rm MCP}, y_{\rm MCP})$ by the background harp electron event distribution $\hat{f}_{\rm he,bg}(x_{\rm MCP}, y_{\rm MCP})$ should therefore give an estimation of the MCP spatial distribution of induced ESD events.

The TOF distribution for induced ESD events for MESSENGER/FIPS should follow that of the tungsten harp emission based events, . However, it should be noted that incident particles tend to hit closer to the center of the MCP. Ions desorbed from these locations have a larger distance to travel due to the curved top of the TOF chamber. This effect is observed as a small shift in TOF ($TOF_{\rm o}$) of the harp electron TOF distribution. For MESSENGER/FIPS, $TOF_{\rm o}$ can be computed for each daily accumulation through calculation of the change in the first moment of the largest TOF peak (325ns-375ns) from the expected background distribution, following equation (7.14),

$$TOF_{\rm o} = \frac{\sum\limits_{\rm ESD} TOF \cdot (f_{\rm m} - f_{\rm he,bg})}{\sum\limits_{\rm ESD} (f_{\rm m} - f_{\rm he,bg})} - \frac{\sum\limits_{\rm ESD} TOF \cdot f_{\rm he,bg}}{\sum\limits_{\rm ESD} f_{\rm he,bg}}, \tag{7.14}$$

where

$$\sum_{\rm ESD} \equiv \sum_{\rm TOF=325ns}^{375ns} \sum_{\rm (E/q)_{ESD}} \sum_{\rm t} \sum_{\rm (x_{MCP}, y_{MCP})} . \tag{7.15}$$

The distribution of ESD induced noise with respect to E/q step and MCP X-Y can be written as,

$$f_{\text{noise,ESD}}(t, x_{\text{MCP}}, y_{\text{MCP}}, E/q) = c \cdot (FSR(t, E/q) - FSR_{\text{bg}}) \times \hat{f}_{\text{he}}(t, x_{\text{MCP}}, y_{\text{MCP}}) \times \hat{f}_{\text{inc}}(t, x_{\text{MCP}}, y_{\text{MCP}}, E/q) \times \hat{f}_{\text{he,bg}}(TOF - TOF_{\text{o}}) \cdot \Delta(t), \tag{7.16}$$

where c is a constant.

The constant c can now be found from a daily accumulation of data, such that ,

$$\sum_{\text{TOF}_{\text{ESD}}} \sum_{(\text{E/q})_{\text{ESD}}} \sum_{\text{t}} \sum_{(\text{x}_{\text{MCP}},\text{y}_{\text{MCP}})} f_{\text{m}} = \sum_{\text{TOF}_{\text{ESD}}} \sum_{(\text{E/q})_{\text{ESD}}} \sum_{\text{t}} \sum_{(\text{x}_{\text{MCP}},\text{y}_{\text{MCP}})} (f_{\text{noise,ESD}} - (f_{\text{noise,fe}} + f_{\text{noise,he}})), \tag{7.17}$$

where $(E/q)_{\text{ESD}}$ and TOF_{ESD} correspond to the ESD region marked in Figure 7.4.

With a full TOF spectrum of noise known, the ESD spectrum readily applies to the overlap region in TOF between 200ns and 300ns.

7.4 Noise processing for MESSENGER/FIPS

With all noises sources well characterized, f_{noise} can be accurately determined for a given set of measured data fm, as illustrated in Figure 7.8, resulting in a direct estimation of SNR for any size dataset. A KDE estimator, given f_{noise} and f_{m} as inputs, can produce $f_{\text{m,KDE}}$, an approximation to the underlying distribution function f that can be analyzed using a Monte Carlo event-processing algorithm, and the measured distribution of real events, $f^*_{\text{m,real}}$, can be recovered. This model is applied to in-flight data for MESSENGER/FIPS.

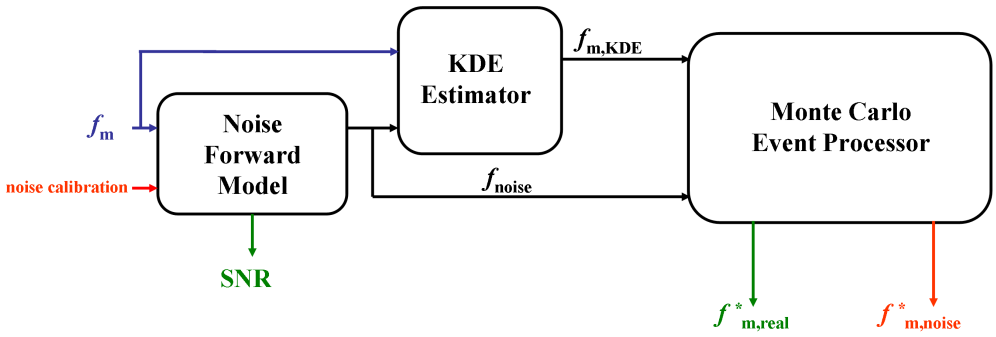

Figure 7.8: Block diagram describing noise modeling and event processing for MESSENGER/FIPS. Measured data as well as noise calibration factors are input into a noise forward model that produces f_{noise}, and consequently, estimates of dataset SNR. The measured data f_{m}, along with f_{noise} is input to a KDE estimator that provides $f_{\text{m,KDE}}$, a distribution suitable for use in a Monte Carlo processor, that separates f_{m} into $f^*_{\text{m,real}}$ and $f^*_{\text{m,noise}}$.

7.4.1 Noise forward modeling

To model the various sources of noise, the relevant noise distributions from Section 7.3 must be added together for each time step. It is important to note that the spatial distribution of incident particles is required in both the models of proton peak tail noise and induced ESD events for MESSENGER/FIPS. In cases where the SNR in each E/q step is low, an iterative noise removal/modeling process may be required to obtain the proper incident spatial distribution. However, in the case of MESSENGER/FIPS the number of expected noise events will be small for a single time step, i.e., $f_{\text{m,noise}}(t) \ll f_{\text{m,real}}(t)$. E/q steps that significantly contribute to active noise sources will have large numbers of incident particles that should dominate the measured X-Y distribution. Therefore the total MCP X-Y distribution is used for an approximation of the spatial distribution of incident particles. The error in this approximation will be large for small numbers of incident particles, but since active noise sources scale with the measured start rate, the contribution from these low count-rate steps to the total accumulated active noise should be negligible.

Due to the number of data dimensions and their corresponding interdependencies, a complete five-dimensional $f_{\text{noise}}(t, E/q, TOF, x_{\text{MCP}}, y_{\text{MCP}})$ distribution may be difficult to feasibly manage and store. Furthermore, as discussed in Section 7.2.2, noise removal using KDE techniques is challenging for higher dimensional datasets. Therefore, instead of producing the complete distribution of noise, several noise data products can be produced that represent accumulations along one or more of the dimensions. Such summation seemingly removes vital information for reconstructing an accurate fm,real. However, as will be discussed in Section 7.4.2.1, with proper choice of accumulations, a reasonable fm,real can be recovered. For MESSENGER/FIPS, the MCP X-Y distribution is measured with the least precision. Furthermore, the majority of the noise sources are either widely spread over the MCP active area, or have similar positional dependence as the incident particles. Therefore, ignoring the MCP X-Y dependence is expected to improve statistical significant of accumulations with minimal impact on the accuracy of modeled distributions.

Two accumulations will be calculated here as an example. The first, f_{noise}, is a $E/q - TOF$ spectrogram for a particular time period. Such an accumulation, while not as useful from a scientific analysis perspective, provides a good evaluation of noise forward modeling performance. The second, f_{noise}, is an energy-dependent time series for a given ion that accumulates over MCP X-Y and several TOF values, and that will be the baseline for all FIPS heavy ion science analysis.

Figure 7.9 shows for 10 days (2011 100-109) an accumulated $E/q - TOF$ distribution (7.9a) with corresponding forward modeled noise (7.9b). Ion tracks for He^{2+}, He^{+}, Na^{+}, and Ca^{+} are marked with dashed white lines. (Note: For clarity, not all ion tracks present in the data have been included in this example.) The two distributions vary only at the locations of ion tracks, where the noise distribution underestimates the total number of events, as it should. There are good matches with the proton peak tail and induced ESD noise in terms of both $E/q - TOF$ location and inten-

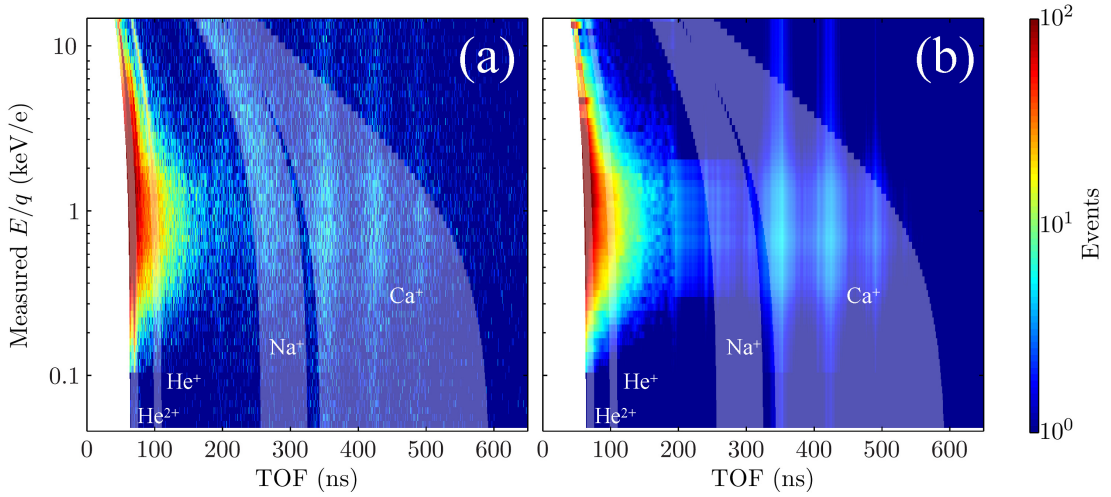

Figure 7.9: (a) 10 day (2011 100-109) $E/q - TOF$ accumulation of MESSENGER/FIPS measured events. Tracks corresponding to heavy ions He^{2+}, He^{+}, Na^{+}, and Ca^{+} are highlighted. (b) Forward modeled $E/q - TOF$ noise distribution for the same time period. There is a good match between the measured data and generated f_{noise} distribution, with the only significant discrepancies occurring in the locations of known ion tracks, where real events are expected to be present.

sity. Synthetic data is used in Appendix C.1 to demonstrate that with an accurate f_{noise}, removal should be accurate to within 25%. Given the quality of the generated f_{noise} shown in Figure 7.9, this estimation of error should be directly applicable to MESSENGER/FIPS measured data.

7.4.2 Noise removal

Once f_{noise} is generated, the set of processed events can be split into sets of recovered real and noise based events. Although it is not possible to calculate the accuracy of this separation for measured datasets, the expected properties of recovered events, both real and noise, can be used to gain confidence in the removal process. On-orbit MESSENGER/FIPS data (2011 100-109) is analyzed using daily accumulations of He^{2+}, He^{+}, Na^{+}, and Ca^{+} at each E/q step, yielding a series of one-dimensional time

series suitable for processing with the algorithm described in Section .

Each time series is split into two parts: 'active and 'quiet, with active times defined as when the proton rate goes above 15 events/s (i.e. inside Mercury's magnetosheath or during FIPS solar wind observations), and more quiet times (typically inside Mercury's magnetosphere) defined as when the proton rate has stayed below 15 events/s for at least 1 hour. The active time periods will be dominated by active noise sources such as H^+ energy straggling and induced ESD, which have larger uncertainty in the forward modeled f_{noise}. If the noise for a particular step is overestimated, real events that occur during quiet times would be thrown out. By splitting up the time series, a more sensitive recovery of events is expected.

Figure 7.10 shows the total number of events from both active and quiet time periods for each ion as a function of E/q, with corresponding recovered real events and predicted f_{noise}. For all ions, the large peak near E/q = 1keV/e is effectively removed, since it corresponds to the proton peak tail/ESD events. From Figure 7.9, where there are clear ion tracks at higher energies in the absence of modeled noise, almost no events are removed.

Statistics will also play a role in the recovery, as the counting error, i.e. $\sqrt{N}$, for low SNR datasets results in some uncertainty on the number of events that should be eliminated. Consequently, recovered signal to counting error threshold, i.e., $S/\sqrt{N}$, may be required to give confidence in the recovery even in the presence of errors in the noise forward model. For example, in *Gershman et al.* (2012d), the threshold used for solar wind heavy ions was $S/\sqrt{N} > 1$ for quiet times and $S/\sqrt{N} > 2$ for active times. The threshold in *Gershman et al.* (2012b) for He^+ pickup ions was $S/\sqrt{N} > 1$ and $S/\sqrt{N} > 10$ for quiet and active, respectively, to ensure that times of increased solar wind proton flux into the instrument do not lead to an erroneously high measurement of He^+ events.

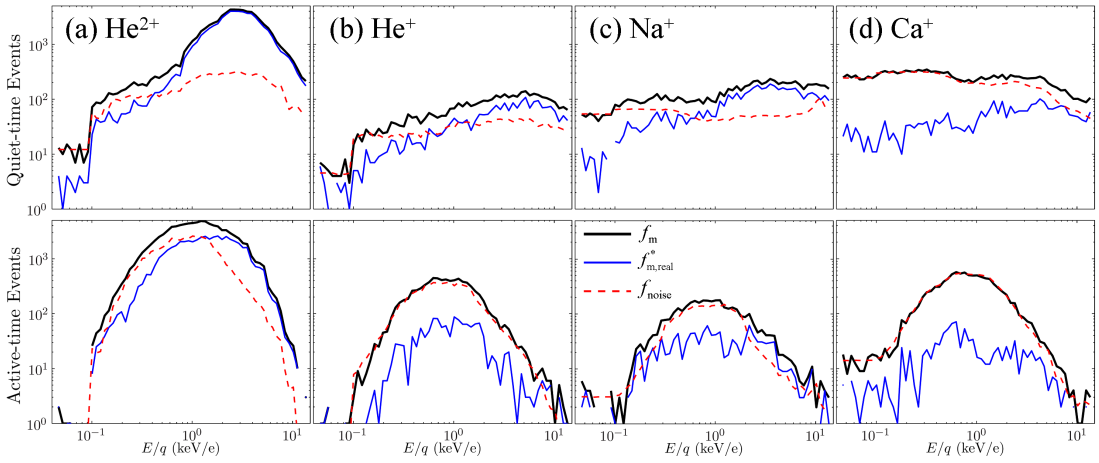

Figure 7.10: Total measured (f_{m}), modeled noise (f_{noise}), and recovered ion events ($f^{*}_{\mathrm{m,real}}$) as a function of E/q for a 10 day (2011 100-109) accumulation of MESSENGER/FIPS on-orbit data for quiet and active periods for (a) He^{2+}, (b) He^{+}, (c) Na^{+}, and (d) Ca^{+}. The noise processing algorithm accurately recovers ion events, resulting in a significant increase in measurement signal-to-noise ratios, especially for E/q steps corresponding to high incident proton flux.

7.4.2.1 Recovery of individual ion events

The output of the noise processing algorithm is a one-dimensional time history of ion events as a function of E/q. However, to enable the most flexibility in scientific analysis of measured instrument data, a list of real individual events in the full five dimensions is desired. Surprisingly, the low measurement statistics is actually advantageous for this recovery. The processed dataset will have at most a few (often only one) measured events per time step. For this case, there is very little ambiguity on which event per step should be identified as noise. However, in cases where there are multiple events per time step, a simple algorithm is used in which the proper number of events with the highest predicted MCP X-Y noise for that particular step are removed. Such a recovery has potential for increased error, but this ambiguity will not be common for low statistics, and in fact datasets with large numbers of events will most likely not require such a recovery. Furthermore, the effects of any errors

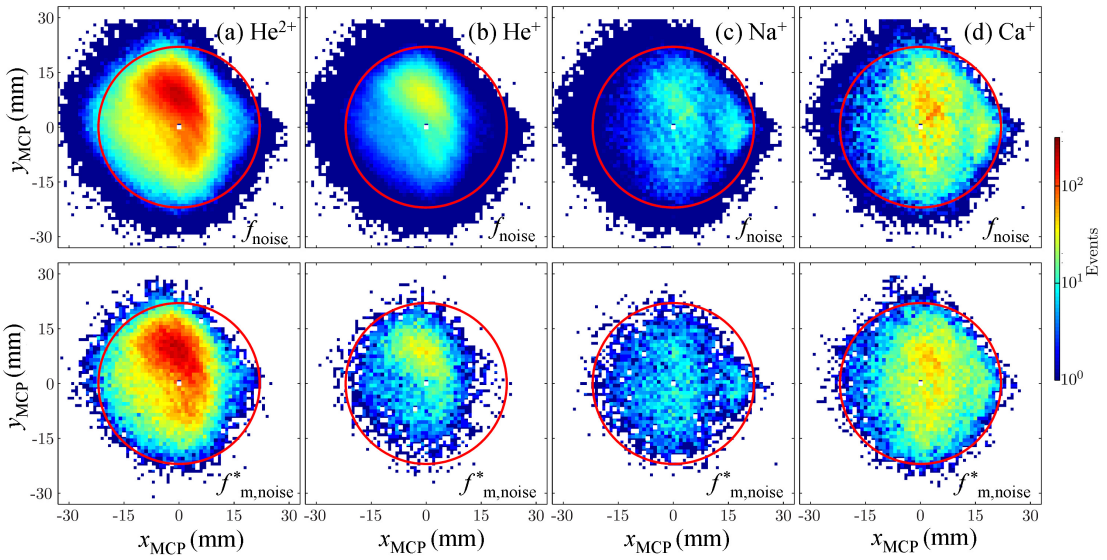

Figure 7.11: Total recovered and modeled noise MCP X-Y distributions for 2011 100-109 for (a) He^{2+}, (b) He^{+}, (c) Na^{+}, and (d) Ca^{+} ions. The recovered noise MCP X-Y distributions are the set of individual events (with full five-dimensional information) that have been processed with the Monte Carlo noise removal algorithm. Although the MCP X-Y dependence was originally ignored in the event processing, the recovered distribution, $f^{*}_{\mathrm{m,noise}}$, still matches the predicted noise, f_{noise}, implying a quality removal.

introduced by incorrect assignment of events are expected to be small, especially since most noise events have similar MCP distributions as the real events.

As an example, Figure 7.11 shows the total modeled MCP X-Y noise distribution for He^{2+}, He^{+}, Na^{+}, and Ca^{+}, and the corresponding X-Y distributions of the recovered noise events, $f^{*}_{\mathrm{m,noise}}$. The overall recovery errors are expected to be less than 25% as predicted by the example distributions from Appendix C.2. The removed noise events exhibit a similar distribution to the predicted noise in terms of both spatial distributions and intensity. Therefore, despite the fact that the MCP X-Y distribution was not considered during the removal process, a reasonable five-dimensional event distribution recovery has been accomplished.

7.5 Conclusions

We have developed a mathematical framework for the characterization, forward modeling, and post-processing of noise events in spaceborne time-of-flight mass spectrometers. Using in-flight data from MESSENGER/FIPS as calibration, a forward model of instrument noise was constructed that provided estimations of data signal to noise ratios in all data dimensions. Kernel density estimation techniques were used to take measure instrument data and recover its underlying continuous distribution function suitable for use with a Monte Carlo event-processing algorithm. With an accurate noise model, measured events were separated into real and noise based events, with an expected overall misidentification of events expected to be less than 25%, even for datasets with low SNR or poor counting statistics. This removal was done after accumulating over incident particle directions that were later recovered to enable the maximum versatility of scientific analysis.

Although applied to MESSENGER/FIPS as an example, the techniques described here are applicable to any such TOF-MS or even other instruments that measure discrete events. An in-depth characterization of instrument noise not only results in an improved understanding of idiosyncrasies of a particular spaceflight instrument, but can also make any changes in instrument response or newly measured scientific phenomena more easily identifiable. These techniques provide a unique way to improve upon the signal-to-noise ratio of instruments whose settings may not be sufficiently adjustable in-flight, a method to mitigate unexpected instrument behavior due to the harsh and unpredictable space environment.

CHAPTER VIII

MESSENGER observations of the spatial distribution of planetary ions near Mercury

This chapter is taken from Zurbuchen, T. H., J. M. Raines, J. A. Slavin, D. J. Gershman, J. A. Gilbert, G. Gloeckler, B. J. Anderson, D. N. Baker, H. Korth, S. M. Krimigis, M. Sarantos, D. Schriver, R. L. McNutt Jr., S. C. Solomon (2011), MESSENGER observations of the spatial distribution of planetary ions near Mercury, *Science, 333 (6051)*, 1862-1865. In addition to assisting with the editing of the manuscript text, I was responsible for the noise processing of all heavy ion data.

Abstract

Global measurements by MESSENGER of the fluxes of heavy ions at Mercury, particularly sodium (Na^+) and oxygen (O^+), exhibit distinct maxima in the northern magnetic-cusp region, indicating that polar regions are important sources of Mercurys ionized exosphere, presumably through solar-wind sputtering near the poles. The observed fluxes of helium (He^+) are more evenly distributed, indicating a more uniform source such as that expected from evaporation from a helium-saturated surface. In some regions near Mercury, especially the nightside equatorial region, the Na^+ pressure can be a substantial fraction of the proton pressure.

Mercurys dipole magnetic field, particularly its small magnitude and near-alignment with the planets rotation axis, defines the planets interaction with the constantly expanding solar atmospherethe solar windand structures the plasma and charged-particle environment of the planet (*Anderson et al.* 2011). By its orientation and strength, Mercurys magnetic field inhibits direct solar wind access to the planetary surface in dayside equatorial regions (*Slavin et al.* 2008), where the average magnetic field orientation is nearly perpendicular to the velocity of the incoming solar wind (*Kabin* 2000). At high latitudes, in contrast, the solar wind interaction with the magnetic field forms northern and southern cusps, funnel-shaped indentations in the magnetopause that capture some of the magnetosheath plasma and guide it to lower altitudes (*Frank* 1971; *Haerendel et al.* 1978). Since Mercury lacks an appreciable atmosphere, this funneling of solar wind plasmas down to the surface is of particular importance because the incident plasma is believed to sputter neutral atoms from the surface into the exosphere and to account for a substantial portion of the exospheres variability (*Killen et al.* 2007). Neutral exospheric particles can also be generated by other processes, such as thermal evaporation off Mercurys surface, desorption stimulated by photons or electrons, and micrometeoroid impact. Less well understood are surface processes that might lead to the direct ejection of ions from the planetary surface (*Killen et al.* 2007). Whether they originate from ionization of the neutral exosphere or from the surface, Mercurys ions subsequently undergo energization and transport by electromagnetic forces that dominate Mercurys space environment (*Slavin et al.* 2008; *Lukyanov et al.* 2004).

During its near-equatorial flybys of the innermost planet in 2008-2009, the MESSENGER spacecraft obtained initial measurements of the structure of the magnetosphere (Fig. 8.1). Mercurys magnetic field is highly distorted by the solar wind. On the dayside, the planetary magnetic field is compressed by the ram pressure of

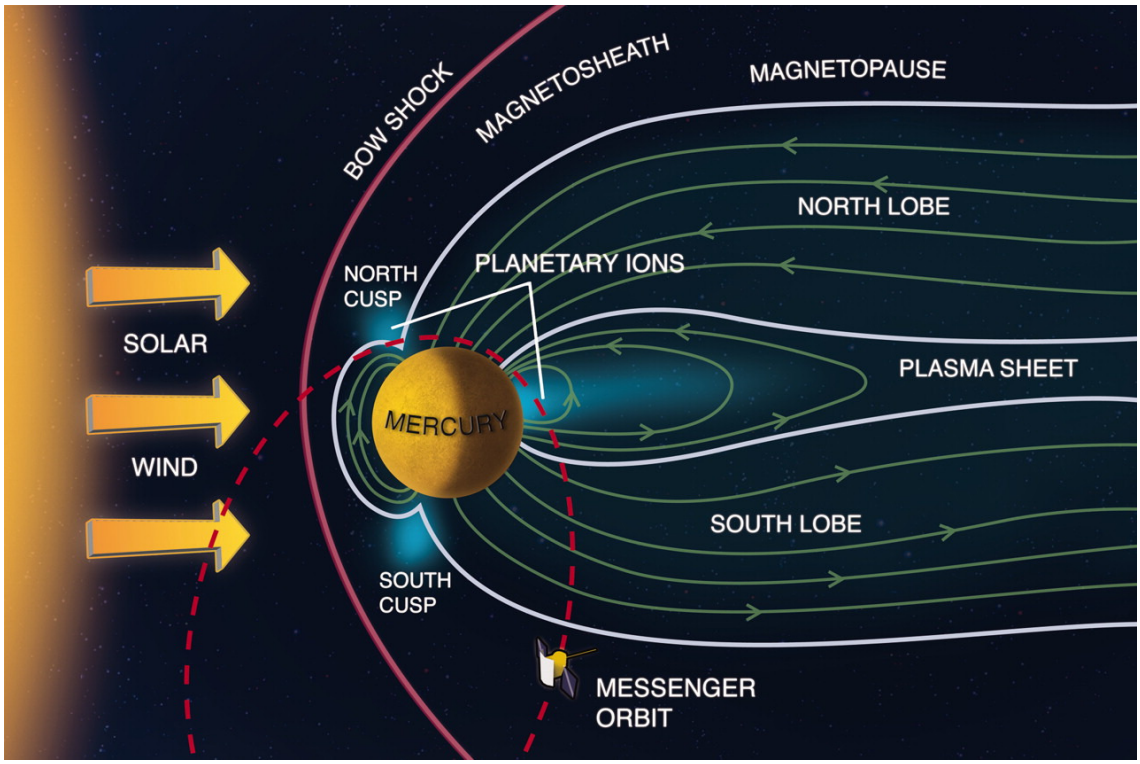

Figure 8.1: This schematic view of Mercurys magnetosphere, derived from measurements made during MESSENGERs three flybys, provides a context for the measurements reported here. Mercurys planetary magnetic field largely shields the surface from the supersonic solar wind emanating continuously from the Sun. MESSENGER has been in a near-polar, highly eccentric orbit (dashed red line) since 18 March 2011. Maxima in heavy ion fluxes observed from orbit are indicated in light blue.

the incident solar wind, whereas on the nightside the magnetic field is pulled back to form a long magnetic tail (*Slavin et al.* 2010). Special attention is called to the northern and southern cusp regions, from which ions from all sources stream along magnetospheric field lines into the northern and southern lobes of the tail, where they drift toward the tails equatorial plane to concentrate and form the plasma sheet (*Delcourt et al.* 2007).

MESSENGER was inserted into orbit about Mercury on 18 March 2011, and here we report the results of near-continuous measurements of planetary ions near Mercury on a global scale. These measurements were made with the Fast Imaging

Plasma Spectrometer (FIPS), the low-energy portion of the Energetic Particle and Plasma Spectrometer (EPPS) instrument (*Andrews et al.* 2007). We focus on the spatial distribution of the most abundant ions with energy per charge E/q between 0.1 and 13 keV/e and with mass per charge m/q between 4 and ∼50 atomic mass units per unit charge. Ions within those ranges constitute more than 80% of the heavy ions (i.e., singly charged ions heavier than H^+) measured by FIPS. Ions with $m/q = 21$ to 30 (here termed the 'Na group' and including Na^+, Mg^+, and Si^+, among others) are dominated by Na^+, the most abundant exospheric ion component (*Zurbuchen et al.* 2008). Ions with $m/q = 14$ to 20 (here termed the 'O group') include O^+ and water-group ions (such as H_3O^+, OH^+, and H_2O^+). Because He in the solar wind is fully ionized, the observed He^+ ($m/q = 4$) flux must largely be produced locally, although it may contain a contribution of helium from interstellar gas, ionized near the Sun and then swept along with the solar wind (*Möbius et al.* 2004).

As shown in Fig. 8.2, He^+ and the Na and O groups of exospheric ions are distributed throughout the entire volume of Mercurys magnetosphere traversed by MESSENGER. Na-group and O-group fluxes peak near the northern polar regions, close to the periapsis of MESSENGERs eccentric orbit and coinciding with the location of the northern magnetic-cusp region. He^+ does not show a distinct enhancement in such regions but is more evenly distributed, with observed fluxes near 60°N latitude comparable to values seen elsewhere, even though Na-group and O-group ions are enhanced by factors of 3 or more. Ions that are strongly associated with surface sputtering effects near the magnetospheric cusps should be concentrated near the polar regions, whereas ionized exospheric components associated with the global neutral exosphere would not be expected to have such a strong correlation with polar regions but would instead be distributed more evenly around the planet. At comparable distances, fluxes near dusk (∼18 hours local time) are identical to those near dawn (∼6 hours local time) to within 5%. A predicted large-scale asymmetry, resulting from

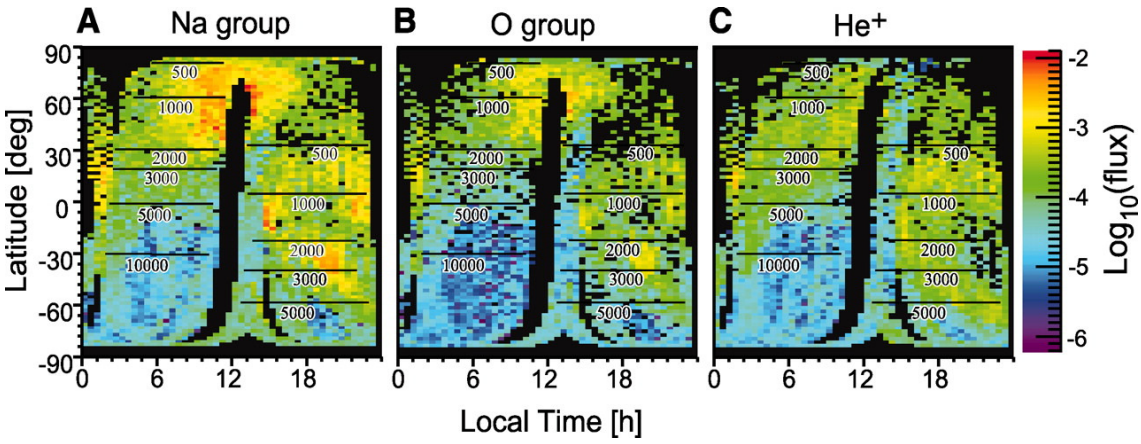

Figure 8.2: Spatial distribution of (A) Na-group, (B) O-group, and (C) He^+ ions versus planetary latitude and local time. Measurements span 66 days (26 March to 30 May 2011) of MESSENGER orbital observations, during which the periapsis local time varied from 17.84 hours to 10.14 hours. Approximate distances from Mercurys surface in km are indicated in black. Colors indicate relative fluxes (in units of $cm^{-2}sr^{-1}s^{-1}$); black areas denote regions not observed, including a swath during which FIPS was mostly powered off. The observed heavy-ion data were collected at 8-s time resolution for most of each orbit and binned by latitude and local time in bins of width 2° and 0.5 hours. Multiple measurements in a single bin were averaged. During these orbits, near Mercury, the planet is generally within the $\sim 1.4\pi$ sr field of view of the FIPS instrument. At distances greater than ~1 Mercury radius, the probability for FIPS to observe ions strongly depends on spacecraft orientation. For example, the solar direction is obstructed at all times by the spacecrafts sunshade. The m/q ratio is derived from the measured ratio of E/q and a velocity measurement, through time of flight, of each ion (*Andrews et al.* 2007).

the transport of exospheric particles in Mercurys magnetosphere (*Lukyanov et al.* 2004), has not been observed. Measured spatial distributions of Na-group, O-group, and He^+ ions might also reflect the spatial distribution of the associated components in a neutral exosphere in which the He originates primarily from evaporation of a He-saturated planetary surface (*Hartle et al.* 1975).

Observations during three dayside high-latitude magnetospheric passes, each lasting 30 to 40 min, provide strong evidence for the cusp association of Na-group and O-group ions (Fig. 8.3). On each orbit, MESSENGER crossed the bow shock and

magnetopause moving anti-sunward and eastward at local times of 09:00 to 10:00 and altitudes of 2000 to 2700 km and 1200 to 1800 km, respectively. The cusp passages are identified by a diamagnetic depression in magnetic field intensity and are characterized by the appearance of 0.1 to 2 keV H^+. These MESSENGER cusp encounters occurred at magnetic latitudes of 65° to 75°N and altitudes of 600 to 800 km, or about halfway between the magnetopause and the surface. The duration of the cusp crossings varied from ∼100 s for the first two passes to ∼150 s for the last. These durations imply that the horizontal dimensions of the cusp along these trajectories were 400 to 600 km.

On 13 April, the heavy ion flux was maximum near the magnetopause, but heavy ions were also present throughout the entire pass, with the lowest fluxes observed after crossing the bow shock and poleward of the cusp. On 14 April, the strongest heavy-ion fluxes were observed within and just before the cusp traversal. However, the most intense and most broadly distributed heavy-ion fluxes were seen during the 16 April pass. Here, the heavy-ion flux exhibited a broad maximum near the magnetopause, but the flux was very high from the middle of the magnetosheath through the cusp region, after which it dropped rapidly. The interplanetary magnetic field (IMF) orientation was quite variable during these passes, and additional studies will be necessary to ascertain the external factors controlling the penetration of heavy planetary ions from the magnetosheath into the magnetosphere and the local production of neutral and ionized species associated with surface sputtering. However, the most probable reason for the enhanced ion flux on 16 April (Fig. 8.3C) is the anti-sunward and northward direction of the IMF measured just before the bow-shock encounter. This IMF orientation is believed to be favorable to magnetic reconnection between the IMF and the magnetospheric magnetic field just tailward of the northern cusp, leading to enhanced solar wind access to the surface and ion sputtering in the northern cusp (*Sarantos et al.* 2001).

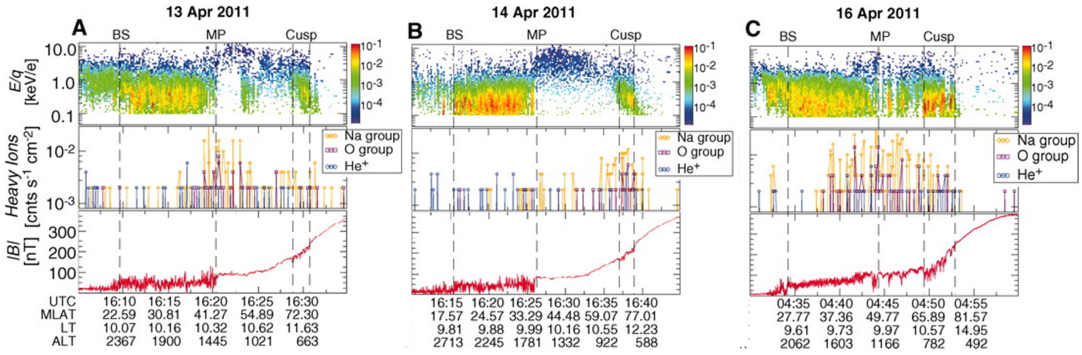

Figure 8.3: MESSENGER observations for three passes through the dayside magnetosphere on (A) 13, (B) 14, and (C) 16 April 2011. For each, the top panel gives E/q spectrograms for H^+ with a time resolution of 8 s covering the E/q range 0.1 to 13 keV/e. The middle panel gives the integrated flux of He^+, O-group, and Na-group ions at the same time resolution and E/q range. The bottom panel gives the magnitude of the measured magnetic field with a time resolution of 50 ms. All quantities are plotted relative to universal time coordinated (UTC), magnetic latitude (MLAT, degrees), local time (LT, fractional hour), and altitude over Mercurys surface (ALT, km). Magnetic latitude includes the northward offset of Mercurys magnetic equator from the geographic equator (*Anderson et al.* 2011). These passes show representative maxima in the H^+ and heavy-ion fluxes at high magnetic latitudes near noon local time (Fig. 8.1). On each orbit, MESSENGER crossed the bow shock (BS) and the magnetopause (MP) and entered the magnetosphere near 10:00 local time at magnetic latitudes 30° to 50°N, and the spacecraft moved closer to Mercury as it headed northwestward to encounter the cusp near noon around 62° to 72°N. The FIPS field of view was stable during all three passes, so the H^+ and heavy-ion fluxes may be compared along each pass and from pass to pass.

Enhancements to the fluxes of Na-group and O-group ions were also observed during MESSENGER passes through the nightside magnetosphere at equatorial latitudes (Fig. 8.2). Three examples of such equatorial passes on 13, 15, and 21 April are displayed in Fig. 8.4. As with the cusp region, these enhancements were often, but not always, associated with diamagnetic decreases in the magnetic field (Fig. 8.4A, B). Although lower fluxes of these heavy ions can be observed at slightly higher and lower latitudes, they maximized near the magnetic equator. The measurements in Fig. 8.4 are consistent with passages through the inner portion of the plasma sheet that separates the northern and southern lobes of Mercurys magnetotail (*Slavin et al.* 2008, 2010; *Delcourt et al.* 2007).

For Maxwellian velocity distributions and low Mach numbers, we determined densities, temperatures, and thus thermal pressures for the H^+ and Na-group ions, where sufficient counts were available (*Raines et al.* 2011). Typically, within these assumptions, the densities and temperatures are accurate to within 20% and are mostly limited by counting statistics. For a given temperature, magnetic field strength, and ion mass-per-charge ratio, the spatial scale of the gyrating motion of a given particle around the field (the Larmor radius) can be calculated. Temperatures of H^+ and Na-group ions imply Larmor radii in the ranges 10 to 30 km and 200 to 700 km, respectively. Both ion groups should therefore be organized by the magnetospheric field, although the Larmor radii of Na-group ions are of the same order of magnitude as the horizontal dimension of the cusp. For the cusp passage of 16 April (Table 8.1), we find a thermal pressure ratio of 2%. The heavy ion particle distributions during the passages of 13 and 14 April are non-Maxwellian, and a pressure calculation is not possible with this methodology.

All magnetospheric passes in Fig. 8.4 exhibit strong maxima in the heavy-ion flux near the equator. In all cases, there was a maximum in the Na-group flux near the equator, yet the extent of these heavy-ion maxima varied by a factor of $\sim$3 among the

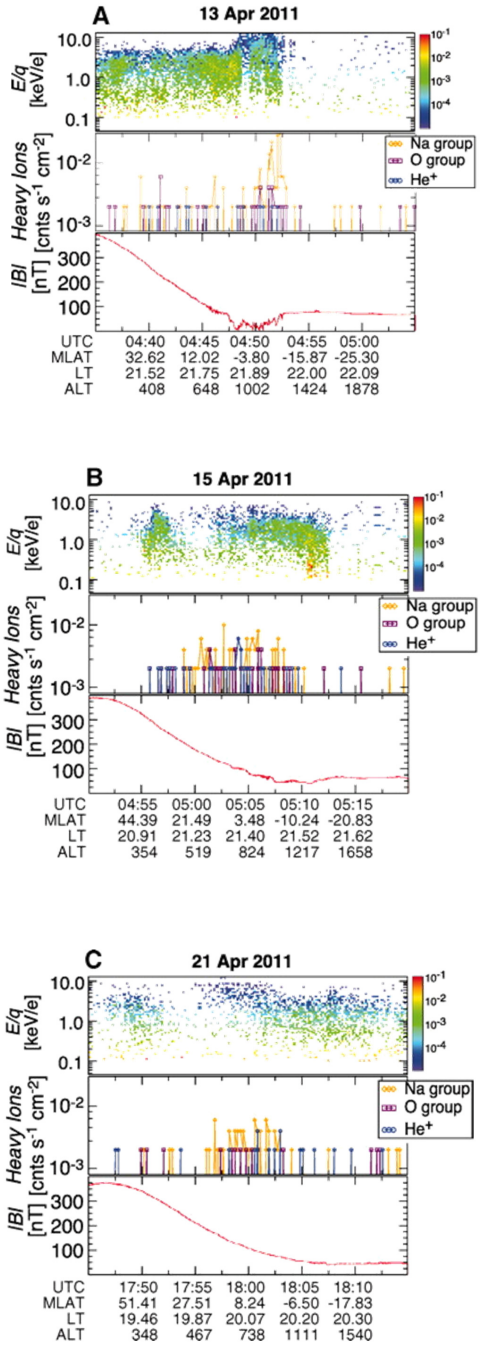

Figure 8.4: MESSENGER observations for three passes through the premidnight magnetosphere at mid to low latitudes on (A) 13, (B) 15, and (C) 21 April 2011 in the same format as in Fig. 8.3. These passes show representative maxima in the H^+ and heavy-ion fluxes near low magnetic latitudes (Fig. 8.1). On each orbit, MESSENGER moved southward and away from Mercury until it crossed Mercurys equatorial plane at altitudes of 800 to 1000 km.

Table 8.1: Estimates of ion density n, temperature T, and pressure P for the cusp (C) and equatorial (E) regions shown in Figs. 8.3 and 8.4.

Date		H^+			Na group		
(2011)		n (cm^{-3})	T (MK)	P (nPa)	n (cm^{-3})	T (MK)	P (nPa)
13 April	C	12	1.3	0.22	*	*	*
14 April	C	20	0.79	0.21	*	*	*
16 April	C	40	0.79	0.44	0.5	1	0.008
13 April	E	1.5	5.1	0.11	†	†	†
15 April	E	1.6	3.6	0.079	0.2	1-3	0.003-0.008
21 April	E	0.28	7.8	0.030	0.3	3	0.01

* Too few counts.

† Substantial bulk velocity violates assumptions for data inversion.

three passes. We determined H^+ and Na-group thermal pressure ratios for the 15 and 21 April equatorial passes (Table 8.1): = 4 to 10% and 33%, respectively (Table 8.1). The nightside magnetospheric plasma is expected to be transported preferentially through the magnetically dominated lobe regions and concentrated in the plasma sheet at low latitudes. The plasma sheet is threaded by closed planetary magnetic fields such that its northward component combined with the dawn-to-dusk electric field imposed by the global solar wind interaction result in a sunward magnetic stress. This stress is balanced by a tailward-directed gradient in the plasma pressure (i.e., plasma density and temperature decrease with increasing downtail distance).

The heavy-ion observations provide important constraints for Mercurys neutral exosphere and its temporal variability and spatial distribution. Single-particle calculations indicate that He^+ and O^+ ion fluxes could be comparable if the near-surface density of neutral oxygen does not exceed 200 atoms cm^{-3}, as predicted by exosphere models (*Sarantos et al.* 2001; *Wurz et al.* 2010). From the observed spatial distributions, the energy of the observed ions, and comparisons with observations of the neutral exosphere, we conclude that the planetary ions discussed here are being created by the ionization of exospheric neutral species. Moreover, MESSENGER observations reveal that Mercurys cusps act as natural collection points and conduits for solar wind and planetary ions, such as the O group, Na group, and other ions,

that are likely to contribute to the neutral exosphere as they precipitate onto the planetary surface. There are orbit-to-orbit differences in observations, which reflect the magnetospheric response to a highly variable solar wind (*Baker et al.* 2011) and which are anticipated from the temporal variability of Earth-based observations of the Na exosphere (*Killen et al.* 2007). Given the high heavy-ion thermal pressures relative to proton pressures at Mercury reported here, the role of heavy ions must be included to properly understand the dynamics of Mercurys magnetosphere.

CHAPTER IX

Solar wind alpha particles and heavy ions in the inner heliosphere

This chapter is taken from: Gershman, D. J., T. H. Zurbuchen, L. A. Fisk, J. A. Gilbert, J. M. Raines, B. J. Anderson, C. W. Smith, H. Korth, and S. C. Solomon (2012), Solar wind alpha particles and heavy ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

Abstract

The Fast Imaging Plasma Spectrometer (FIPS) on the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) mission has made the first in situ measurements of solar wind plasma in the inner heliosphere since the Helios 1 and 2 spacecraft in the 1980s. Although a portion of the solar wind velocity distribution is obstructed by the spacecraft sunshade, a data analysis technique has been developed and tested with synthetic data that recovers both bulk and thermal speeds to 10% accuracy for a variety of ion species, providing the first measurements of solar wind heavy ions (mass per charge > 2 amu/e) at heliocentric distances within 0.5 AU. Solar wind alpha particles and heavy ions appear to have similar mean flow speeds at values greater than that of the protons by approximately 70% of the Alfvén speed. A statistical survey of proton and heavy ion bulk and thermal speeds mea-

sured by FIPS between 0.3 and 0.5 AU, indicating that heavy ions are, in general, both hotter and faster than protons. From an examination of the thermal properties of alpha particles and heavier solar wind ions, we find a ratio of the temperature of alpha particles to that of protons nearly twice that of previously reported Helios observations, though still within the limits of excessive heating of heavy ions observed remotely by the Ultraviolet Coronograph Spectrometer on the Solar and Heliospheric Observatory close to the Sun. Furthermore, examination of typical magnetic power spectra at the orbit of MESSENGER as well as at 1 AU reveals the lack of a strong signature of local resonant ion heating, implying that a majority of the observed heavy ion heating could occur close to the Sun. These results demonstrate that the solar wind at ∼0.3 AU is a blend of the effects of waveparticle interactions occurring in both the solar corona and the heliosphere, whereas the solar wind at 1 AU is more dominated by heliospheric processes.

9.1 Introduction

The solar wind is the manifestation of the ubiquitous expansion of the hot solar corona into space. Although the origin, heating, and acceleration of the solar wind are still much debated, there is consensus on a number of key issues. Initial models of the solar wind were based on the assumption of an isothermal plasma, with electrons and ions having equal temperatures (*Parker* 1958). However, we know from ultraviolet observations of neutral hydrogen coupled to solar wind plasmas near the Sun through charge exchange and from direct observations of heavy ions (mass $>$ 4 amu) that the ions are preferentially heated, and that this heating is necessary to attain the observed solar wind speeds (*Kohl et al.* 1997, 1998). At about 2 solar radii, solar wind ions associated with coronal holes have relative temperatures that strongly depend on the

ratio of their mass to charge (m/q); for instance, ratios of the temperatures of Mg^{9+} and O^{5+} to that of neutral hydrogen, $T_{\mathrm{Mg9+}}/T_{\mathrm{H}}$ and $T_{\mathrm{O5+}}/T_{\mathrm{H}}$, are approximately equal to 64 at 1.7 and 2.2 solar radii, respectively (*Kohl et al.* 1997). Similarly, the derived flow speeds near 2 solar radii are also higher for heavy ions, with $v_{\mathrm{O5+}}/v_{\mathrm{H}} \sim 2$ (*Kohl et al.* 1998).

These remote sensing data elucidate the important role of waveparticle interactions in most parts of the closed and open corona, even though the precise physical processes involved in these interactions is still debated (e.g., *Fisk et al.* (1999); *Verdini and Velli* (2007); *Cranmer et al.* (2007)). Most importantly, these data show that the collisional coupling of heavy ions to solar wind protons is not as dominant as initially assumed in the first comprehensive model of heavy ions by *Geiss et al.* (1970). In these collision-dominated models, heavy ions are coupled to the accelerating protons through Coulomb collisions, and thus their velocity is typically smaller than that of protons. Similarly, due to the dominance of these collisions, heavy ions were predicted to have thermal speeds smaller than those of protons, again in contradiction with key results from the Ultraviolet Coronograph Spectrometer (UVCS) on the Solar and Heliospheric Observatory (SOHO).

In situ observations of solar wind heavy ions to date have been limited to heliocentric distances of 1 AU and beyond. These measurements were performed mostly by solar wind composition instruments such as those on the Advanced Composition Explorer (ACE), Ulysses, and Solar TErrestrial RElations Observatory (STEREO) spacecraft (*Gloeckler et al.* 1992, 1998; *Galvin et al.* 2008). Through their filtering of ions by energy per charge (E/q) and the measurement, after post-acceleration, of time of flight and total energy, the velocity (v), mass (m), and charge (q) of heavy ions are determined. There are some observational limitations that are worth mentioning: the Solar Wind Ion Composition Spectrometer (SWICS) instruments on ACE and Ulysses measure only the speed of each ion and do not have sufficient angular

resolution to recover the full three-dimensional orientation of the velocity vector, prohibiting the analysis of full distribution functions (see *von Steiger and Zurbuchen* (2006)).The Plasma and Suprathermal Ion Composition (PLASTIC) instrument on STEREO, however, has the ability to resolve incident direction (*Bochsler et al.* 2010).

Generally, these observations near 1 AU show a streaming of heavy ions faster than protons along the magnetic field at differential velocities that are of the same order, but typically less than the Alfvén speed (*Asbridge et al.* 1976; *Zurbuchen et al.* 2000; *Berger et al.* 2011). The actual amount of differential streaming depends on the mass and charge of ions in a way that is not fully understood (*Hefti et al.* 1998; *von Steiger and Zurbuchen* 2006), although there seems to be an ordering of the data by q^2/m, indicating a role of collisional effects. Heavy ions observed near 1 AU also tend to have thermal speeds that are approximately equal, indicating a mass-proportional temperature. From a comparison of data gathered near 1 AU with data from Ulysses that span up to 5.4 AU, it is evident that, with increasing heliocentric distance, the solar wind becomes increasingly less collisional.

Simple electrostatic analysis can be used for the measurement of the kinetic properties of alpha particles relative to protons during time periods when the solar wind is not too hot (*Neugebauer* 1976; *Marsch et al.* 1982a). Such analyses are used as part of the Solar Wind Electron, Proton and Alpha Monitor (SWEPAM) (*McComas et al.* 1998a) or Faraday cup (e.g., *Ogilvie et al.* (1995)) experiments. These works provide an intriguing picture of the importance of collisional processes, waveparticle interactions, and plasma instabilities. In interpreting these data, it is critical to recognize that the alpha particles carry $\sim$20% of the internal pressure of a solar wind plasma (*McKenzie et al.* 1995) and are therefore a major contributor to plasma dynamics, such as the dispersion properties of waves and instabilities (e.g., *Gary et al.* (2001)). Following a long-duration analysis of alpha particle and proton data near 1 AU, *Kasper et al.* (2008) demonstrated the importance of collisional processes that

shape the relative speed, temperature, and temperature anisotropies of alpha particles and protons. They also argued that there is local heating through the local dissipation of kinetic Alfvén waves.

The differential heating of heavy ions with respect to the protons is expected to occur near the transition of two subintervals of the magnetic field spectra: the inertial range and the dissipation range (*Frisch* 1996). The inertial range as it is expected to exist in the plasma frame occurs at frequencies substantially lower than the gyro-frequency of heavy ions, obscured by Doppler shifts that are a consequence of measuring the magnetic field $\mathbf{B}$ in the spacecraft frame. Here, the energy transport occurs in an approximately scale-invariant cascade process, carrying energy to successively smaller spatial scales and higher values of frequency (f) without any dissipation. The spectral slope of this transport varies as $f^{-5/3}$ or $f^{-3/2}$ for dominant fluid dynamics (Kolmogorov) or magnetic transport effects (Kraichnan), respectively. At sufficiently high frequencies, dissipation sets in (*Leamon et al.* 1998). This dissipation strongly depends on the kinetic properties of the plasma. The importance of various dissipation processes is a subject of current debate (*Leamon et al.* 1998, 1999; *Bale et al.* 2005; *Hamilton et al.* 2008; *Smith et al.* 2012). In order to test the predictions of such turbulence models, a combined dataset of heavy ions and magnetic field measurements is required at multiple heliocentric distances.

The only comprehensive analysis of solar wind ions in the inner heliosphere was made by the Helios spacecraft in the late 1970s and early 1980s. Helios 1 and Helios 2 had highly eccentric heliospheric orbits with perihelia of 0.29 AU and 0.3 AU, respectively. The mission, payload, and plasma instrument were described by *Schwenn et al.* (1975) and *Rosenbauer et al.* (1977). The distribution functions of protons (*Marsch et al.* 1982b) and alpha particles (*Marsch et al.* 1982a) were characterized in detail, including examinations of the temperatures and differential flow speeds between alpha particles and protons as functions of heliocentric distance and solar wind

speed. However, the Helios payload included no heavy ion instruments, and there have been no new in situ measurements in the inner heliosphere for the past 30 years.

Here, we focus on the dynamic properties of heavy ions measured with a mass-resolving instrument that is part of the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft that is currently in orbit about Mercury (*Solomon et al.* 2001). We provide the first combined measurements of solar wind protons, alpha particles, and a group of heavy ions, and we place these data into context with previous observations by Helios and those more recently obtained near 1 AU. Specifically, we address the nonthermal behavior of alpha particles and heavy ions, including a discussion of the typical solar wind turbulent environment inferred from high-resolution magnetic field measurements from both MESSENGER and ACE. In Section 9.2, we provide an introduction of the methodology for deriving the relevant kinetic parameters from partially obscured measurements and show the recovered solar wind parameters for the inner heliosphere as measured by MESSENGER. In Section 9.3, we detail observations related to the thermal and kinetic properties of alpha particles and heavy ions with respect to the protons. Finally, a concluding discussion of our results appears in Section 9.4.

9.2 Solar wind observations by MESSENGER

The Fast Imaging Plasma Spectrometer (FIPS), one of two charged particle sensors on MESSENGERs Energetic Particle and Plasma Spectrometer (EPPS) instrument (*Andrews et al.* (2007)), was designed to measure magnetospheric ions around Mercury with m/q ratios up to 40 amu/e and E/q ratios ranging from 40 eV/e to 10 keV/e. Typical solar wind plasmas fall well within this range, such that FIPS is fully capable of resolving the solar wind, including several of its less abundant ion species and pickup ions. However, as a consequence of the missions thermal design, the direct line of sight of FIPS in the solar direction is obstructed by the MESSEN-

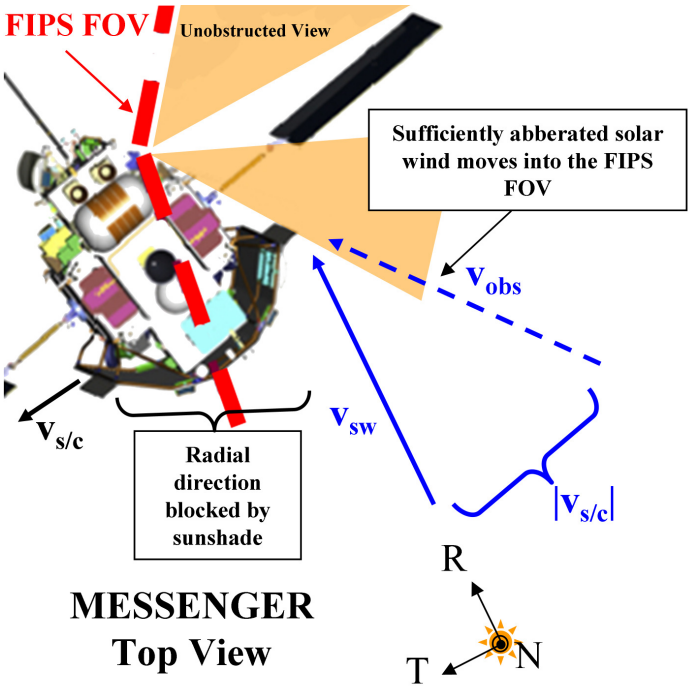

Figure 9.1: Top view of a portion of the MESSENGER spacecraft depicting the obstruction of the solar direction by the spacecraft sunshade and the FIPS field of view (FOV). Because of substantial aberration of $\mathbf{v}_{sw}$ from the spacecraft velocity $\mathbf{v}_{s/c}$, radially traveling particles are measured with an effective $\mathbf{v}_{obs}$ by the spacecraft, allowing, under some conditions, a fraction of the solar wind distribution to be analyzed by FIPS.

GER spacecraft sunshade, as illustrated in Figure 9.1. This obstruction nominally prevents the instrument from sampling the majority of the solar wind distribution.

Fortunately, due to nonradial particle velocities from magnetic turbulence, increased plasma temperatures close to the Sun, and aberration from a substantial spacecraft velocity in the tangential direction, a portion of the solar wind distribution function can often be sampled by the instrument and analyzed. After deriving the moments of a partial Maxwell-Boltzmann distribution function (Appendix D.1), we find that whereas it is difficult to calculate accurately the plasma density without additional constraints on the three-dimensional solar wind bulk velocity, the solar wind speed and thermal velocity should be readily recoverable. This section describes a methodology for recovering estimates of solar wind speed and thermal velocities of

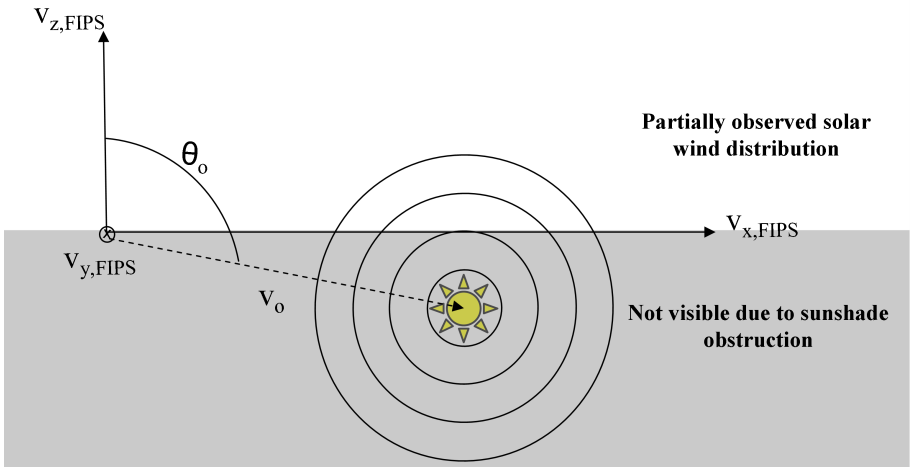

Figure 9.2: Coordinate system centered in the MESSENGER FIPS frame. Here, the z-axis is perpendicular to the sunshade obstruction and the solar wind core is centered at $[v_o sin(\theta_o), 0, v_o cos(\theta_o)]$. Only the portion of the solar wind distribution with $\theta < \pi/2$ is measured by FIPS.

diverse ion species to within 10% accuracy from such partial distributions.

9.2.1 Distribution functions for instrument observations

A coordinate system for FIPS with respect to a partially observed solar wind plasma is defined in Figure 9.2. This coordinate system will enable the derivation of a distribution function for FIPS observations that can be used to recover solar wind bulk and thermal speeds. The x-y plane is the plane of the sunshade, and the z-axis is defined such that it completes a right-handed coordinate system. The average velocity of the solar wind is denoted v_o, and its direction is defined such that the center of the distribution lies in the x-z plane with coordinates $[v_o sin(\theta_o), 0, v_o cos(\theta_o)]$, i.e. $\phi_o = 0$. The angle $|\pi/2 - \theta_o|$ represents the minimum angular distance from the distribution center that lies within the instruments field of view (FOV). Following standard spherical coordinate system definitions, θ is the angle from the z-axis and ϕ is the clock angle in the x-y plane.

We assume that any other obstructions to the instrument FOV are negligible, so that the solar wind distribution is visible over the ranges $\theta = [0, \pi/2]$ and $\phi =$

$[0, 2\pi]$. The highly supersonic solar wind is confined to a small angular width in the instrument FOV, so other obstructions, such as solar array panels or small protrusions of the spacecraft body into the FOV, should not markedly affect the fraction of the distribution observed.

With this newly defined coordinate system, a model for the solar wind distribution function takes the form of an isotropic drifting Maxwellian distribution in spherical coordinates:

$$f(v,\theta,\phi) = n_{\mathrm{o}} \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{v_{\mathrm{th}}^3} exp\left(-\frac{v^2 + v_{\mathrm{o}}^2 - 2vv_{\mathrm{o}}(sin\theta sin\theta_{\mathrm{o}} cos\phi + cos\theta cos\theta_{\mathrm{o}})}{2v_{\mathrm{th}}^2}\right), \tag{9.1}$$

where n_{o} is the plasma density and v_{th} is related to the temperature T by $v_{\mathrm{th}} \equiv \sqrt{kT/m}$ and k is the Boltzmann constant.

A FIPS energy scan implicitly integrates over the instrument FOV to give an observed distribution as a function of measured speed, $f_{\mathrm{obs}}(v)$. This integration can be explicitly written as

$$f_{\mathrm{obs}}(v) = \int_{\mathrm{o}}^{\pi/2} sin\theta d\theta \int_{\mathrm{o}}^{2\pi} d\phi \cdot f(v,\theta,\phi). \tag{9.2}$$

Equation (9.2) can be integrated analytically under the assumption of supersonic flow and a center of the distribution near the edge of the sunshade (see Appendix D.2). From this solution, the measured quantities F_0, F_1, and F_2 may be calculated and manipulated to produce the observed density $n_{\mathrm{obs}} < n_{\mathrm{o}}$ and estimates of v_{o} and v_{th} directly from FIPS measurements:

$$F_0 = \int_{\mathrm{o}}^{\infty} f_{\mathrm{obs}}(v) \cdot v^2 dv \approx n_{\mathrm{obs}},$$

$$F_1 = \left\langle \frac{1}{v} \right\rangle = \int_{\mathrm{o}}^{\infty} \frac{1}{v} \cdot f_{\mathrm{obs}} \cdot v^2 dv \approx \frac{n_{\mathrm{obs}}}{v_{\mathrm{o}}}, \text{ and}$$

$$F_2 = \left\langle \frac{(v - v_{\mathrm{o}})^2}{v} \right\rangle = \int_{\mathrm{o}}^{\infty} \frac{(v - v_{\mathrm{o}})^2}{v} \cdot f_{\mathrm{obs}} \cdot v^2 dv \approx \frac{n_{\mathrm{obs}} \cdot v_{\mathrm{th}}^2}{v_{\mathrm{o}}}. \tag{9.3}$$

n_{obs}, v_{o}, and v_{th} can be recovered through manipulations of these moments,

$$n_{\mathrm{obs}} = F_0,$$

$$v_{\mathrm{o}} = \frac{F_0}{F_1}, \text{ and}$$

$$v_{\mathrm{th}} = \sqrt{\frac{F_2}{F_1}}. \tag{9.4}$$

These relations apply to any measured ion species, providing that the proper conversion from instrument energy to ion velocity is determined. Similar moments are used to derive solar wind parameters for observations from ACE/SWICS and Ulysses/SWICS (*von Steiger et al.* 2000).

9.2.1.1 Kappa distributions

To first order, an isotropic Maxwell-Boltzmann distribution is a good description of the solar wind. However, a more general distribution function for solar wind plasmas is a kappa distribution (*Vasyliunas* 1968):

$$f_\kappa(v) \sim \left(1 + \frac{1}{\kappa} \frac{v^2}{2v_{\mathrm{th}}^2}\right)^{-(\kappa+1)} \tag{9.5}$$

The parameter κ, defined for $1.5 < \kappa < \infty$, describes the shape of the velocity

distribution away from a quasi-Maxwellian core. A value $\kappa \approx 1.5$ gives a v^{-5} power-law tail, i.e., a strong suprathermal particle population commonly observed in the solar wind. This particular distribution will place an upper bound on that population, as the measured tail typically begins at several times the solar wind speed (*Fisk and Gloeckler* 2006; *Fisk et al.* 2010).

Higher κ values describe distributions with weaker tails, and $\kappa \to \infty$ reduces the distribution to a tailless Maxwellian. For a low-κ distribution, the results from Section 9.2.1 may not necessarily apply. The effect of these tails on the recovery of solar wind parameters will be to decrease velocity estimates and increase temperature estimates.

9.2.2 Recovery error estimates

To quantify the recovery errors in v_o and v_th, data from synthetic kappa distributions have been produced and analyzed using the methodology of the previous section. Two sets of kappa distributions were used: $\kappa = 1.6$ and $\kappa \to \infty$. Each set consists of distributions with varying $M \equiv v_\mathrm{o}/v_\mathrm{th}$ and θ_o values. A $v_\mathrm{o} = 450$ km/s is selected with M varying from 3 to 12 and θ_o varying from 90° to 120°.

For a given M and θ_o, a kappa distribution was sampled over $\theta = [0, \pi/2]$ and $\phi = [0, 2\pi]$ at 1° resolution and accumulated to form $f_\mathrm{obs}(v)$. This $f_\mathrm{obs}(v)$ was integrated following equation (9.3) to find F_0, F_1, F_2 and subsequently estimates for $v_\mathrm{o,recovered}$ and $v_\mathrm{th,recovered}$. These estimates were compared with the known input values of v_o and v_th, and the relative error between them was calculated from

$$E_\mathrm{vo}(M, \theta_\mathrm{o}) = \frac{|v_\mathrm{o} - v_\mathrm{o,recovered}(M, \theta_\mathrm{o})|}{v_\mathrm{o}} \tag{9.6}$$

and

$$E_\mathrm{vth}(M, \theta_\mathrm{o}) = \frac{|v_\mathrm{o}/M - v_\mathrm{th,recovered}(M, \theta_\mathrm{o})|}{v_\mathrm{o}/M}. \tag{9.7}$$

The recoveries tested here are applicable to multiple ion species in the same plasma. Although different species may have different bulk and/or thermal velocities, only motion of the distribution core with respect to the plane of the sunshade results in changes to θ_o, i.e., only one dimension of ion motion can affect θ_o. Ion species with different solar wind speeds, for example, can still have the same θ_o values, and consequently their errors are expected to scale with one another, allowing for detailed comparison between them.

9.2.2.1 Bulk velocities

For almost all distributions visible to within 20° of the core, regardless of the suprathermal population, the error in recovered velocity is less than 10%, as demonstrated in Figure 9.3, which shows the relative error in recovered bulk velocities for both sets of kappa distributions. Small-scale structure in these images is because multiple pairs of M and θ_o values produce similar distributions. Because of the finite sensitivity of FIPS and typical solar wind plasma densities in the inner heliosphere (*Marsch et al.* 1982b), it is unlikely that any distribution centered more than 20° away from the sunshade obstruction will generate any noticeable measured events within the instrument. It is therefore expected that, under the stated assumptions, FIPS can produce accurate estimates of solar wind speeds for any measured solar wind distribution.

9.2.2.2 Thermal velocities

As expected from Appendix D.1, and shown in Figure 9.4, the error in recovered thermal velocity is very small for the Maxwell-Boltzmann distribution and is relatively independent of the visible fraction of the distribution function. A strong suprathermal tail, however, creates substantial errors in recovered thermal velocity except for either very hot ($M \approx 3$) distributions or when almost half of the distribution function is

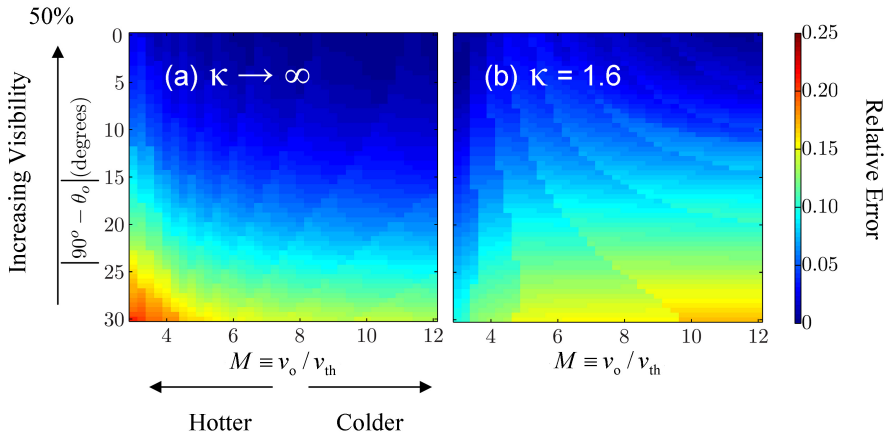

Figure 9.3: Relative error in recovered velocity for (a) $\kappa \to \infty$ (Maxwellian) and (b) $\kappa = 1.6$ velocity distributions. A decreasing $M \equiv v_o/v_{th}$ indicates hotter, less supersonic plasma. $(90^\circ - \theta_o) = 0$ indicates that 50% of the distribution is visible, with a lower fraction visible at increasing angular distance from the core. For both sets of distributions, when the core is within 20° of the sunshade obstruction, the velocity is recovered to within 10% accuracy.

visible.

Although the errors for the $\kappa = 1.6$ case may be discouraging, they can nonetheless be used to aid in identification of periods of recoverable thermal velocities. Consider the turbulent solar wind moving in and out of the FIPS FOV. As this movement occurs, the instantaneous θ_o value changes. Following the results in Figure 9.4b, such movement in the presence of a strong suprathermal population will create sharp changes in the recovered v_{th} value as a function of time. Therefore, despite not knowing the precise values of θ_o or κ, if the recovery of thermal velocity is restricted to time periods where there is a stably recovered v_{th} with time, it is likely that either FIPS is observing a near-Maxwellian distribution or that it is close to the solar wind core such that the effects of the suprathermal particle population will not affect the recovery. During these time periods, and with the given assumptions, it is estimated that the recovery of thermal velocity will be accurate to within 10%.

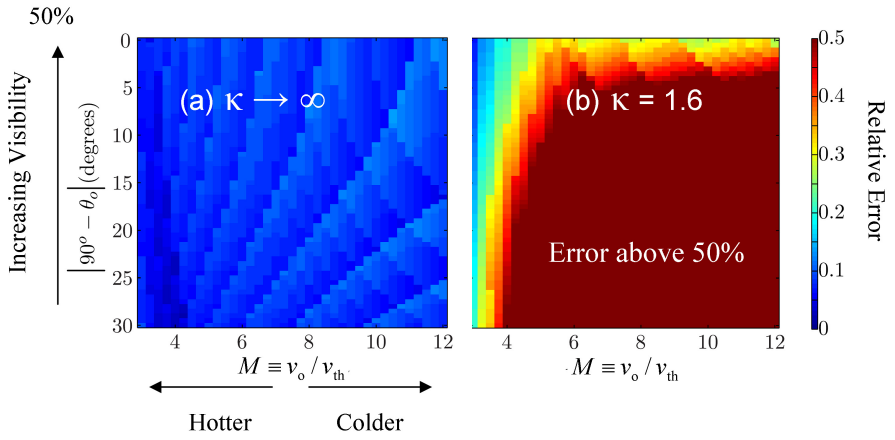

Figure 9.4: Relative error in recovered thermal velocities for (a) $\kappa \to \infty$ (Maxwellian) and (b) $\kappa = 1.6$ velocity distributions. For the Maxwell-Boltzmann case, the recovered thermal velocity is within 10% independent of the fraction of the distribution function that is visible to FIPS. For the $\kappa = 1.6$ case, however, the strong suprathermal tail creates large changes in recovered $v_{\rm th}$ with angular distance from the core.

9.2.3 Recovery criteria

From Section 9.2.2, bulk and thermal velocities of solar wind ions are expected to be recoverable to within 10% accuracy. However, the solar wind is not the only plasma population measured by FIPS as MESSENGER orbits Mercury, passing through the planets magnetosphere multiple times per day. Not only are criteria needed to determine whether or not we can expect accurately recovered solar wind parameters for a particular ion species, but the plasma population that is being measured by the instrument must be determined.

The highest count rates in the instrument correspond to proton events. Therefore, a set of filters that examine the plasma properties of the measured protons were used to determine whether the solar wind distribution was being observed. Less abundant species such as He^{2+} and other heavy ions were recovered only during these times. H^+ velocities were recovered for each instrument energy scan, averaged over approximately 1 min or 10 s depending on instrument operating mode. Longer time-averaged

velocities of H^+ are averages of these individual scans. Heavy ion measurements, due to limited statistics, were summed over several scans before recovering their velocities. The thresholds below which velocities were not recovered due to insufficient statistics were selected as $n_{\rm obs} < 5 \times 10^{-3}$ cm^{-3} for bulk velocity and $n_{\rm obs} < 1.5 \times 10^{-2}$ cm^{-3} for thermal velocities.

Because of the limitations of mass resolution for FIPS, heavy ion charge states such as O^{6+} and C^{5+} cannot be easily separated. Consequently, events from all ions with m/q between $\sim$2 and 3 amu/e were accumulated together. These events, however, will typically be dominated by O^{6+} and C^{5+}, the most abundant heavy ion species in the solar wind (*von Steiger et al.* 2000). For this reason, the average heavy ions will be denoted as $\langle OC \rangle$. Since the conversion from energy-per-charge to velocity depends on m/q, an average solar wind m/q of 2.5 is used on the basis of abundances reported by *von Steiger et al.* (2000). This factor may introduce additional errors on the order of about 5% to the determination of the average heavy ion velocities.

Bulk velocities have been recovered from FIPS measurements if: (1) sufficient numbers of H+ events are measured to obtain 10% peak widths, (2) MESSENGER was outside of Mercurys bow shock as determined by data from magnetic field measurements (*Winslow et al.* 2012), (3) the measured H^+ energy distribution is singly peaked, (4) the mean incident location of events is calculated to be within 30 of the radial direction, and (5) more than two consecutive energy scans meet criteria (1-4). Thermal velocities are recovered if criteria (1-5) are met with the added restriction that there must be recoverable thermal velocities from at least five consecutive energy scans and that the standard deviation of those velocities stays below 5 km/s, indicating a stable recovery.

As an example, Figure 5 shows a period of recovered bulk and thermal velocities for H^+, He^{2+}, and $\langle OC \rangle$ for the period from 10 June to 15 July 2011. The stream structure of the solar wind (*McComas et al.* 1998b) as well as plasma temperatures

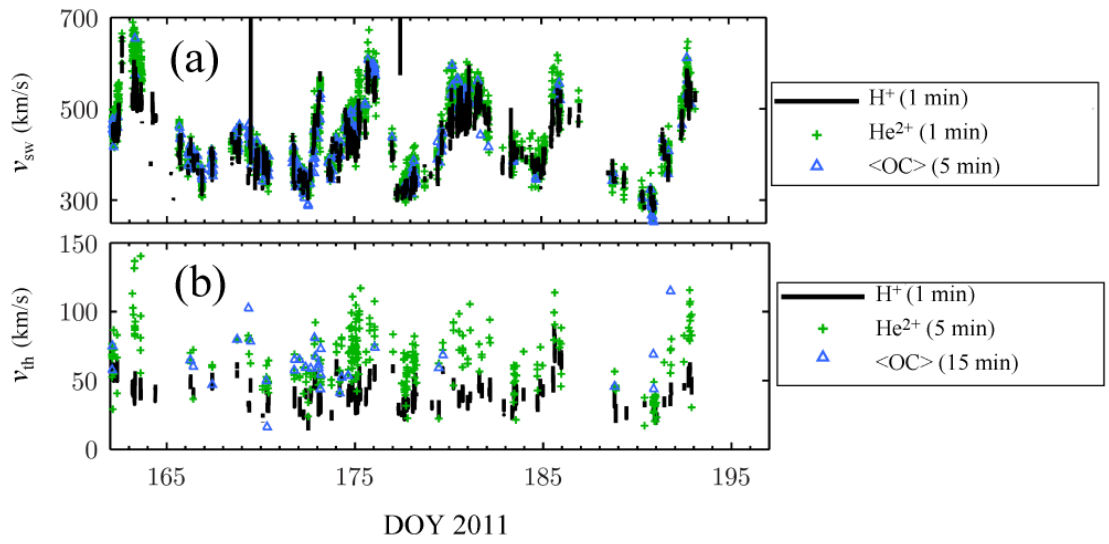

Figure 9.5: Recovered (a) bulk velocities and (b) thermal velocities of H^+, He^{2+}, and $\langle OC \rangle$ for the period from 10 June to 15 July 2011. These filtered datasets show typical solar wind structure, with speeds and temperatures characteristic of inner heliospheric plasma. Time accumulations for each data product are indicated.

consistent with previous observations of inner heliospheric solar wind from Helios 1 and 2 (*Freeman* 1988; *Marsch et al.* 1982b) are recovered from the filtered dataset.

9.3 Heavy ion measurements in the inner heliosphere

With bulk and thermal velocities recoverable for both solar wind protons and heavy ion species, the effects of propagating Alfvénic fluctuations and average heavy ion kinetic properties can be characterized in the inner heliosphere (solar distance R < 0.5 AU) with data from both the FIPS and Magnetometer (MAG) (*Anderson et al.* 2007) instruments on MESSENGER.

9.3.1 Identification of Alfvénic turbulence

Alfvénic fluctuations are ubiquitous structures in the solar wind (*Coleman* 1968; *Belcher and Davis* 1971), particularly in fast streams associated with coronal holes

(*Smith et al.* 1995). They are identified as periods of nearly constant magnetic field magnitude $|\mathbf{B}|$ and plasma density n (mass density ρ), with correlations between changes in the measured solar wind velocity and magnetic field vector, such that

$$\delta\mathbf{v} \equiv \mathbf{v} - < \mathbf{v} > \tag{9.8}$$

and

$$\delta\mathbf{B} \equiv \mathbf{B} - < \mathbf{B} > . \tag{9.9}$$

with $\delta\mathbf{v} \propto \delta\mathbf{B}$, where $\mathbf{v}$ is the solar wind velocity vector and $\mathbf{B}$ is the magnetic field vector. The proportionality constant ($\delta v/\delta B$) is related to the wave velocity $\mathbf{v}_{\rm w}$ of the propagating fluctuation by the relation

$$\mathbf{v}_{\rm w} = \frac{\delta v}{\delta B}\mathbf{B} \tag{9.10}$$

(*B. E. Goldstein et al.* 1995).

For an ideal Alfvén wave, $\delta\mathbf{v} = \pm\delta\mathbf{B}/\sqrt{4\pi\rho}$, leading to a wave propagating with the Alfvén speed along the direction of $\mathbf{B}$. Because of the large uncertainties in ρ from FIPS measurements, the local Alfvén speed cannot be reliably determined. However, from the correlation between $\delta\mathbf{v}$ and $\delta\mathbf{B}$, the effective wave speed from (9.10) is readily calculable. For large-amplitude nonlinear Alfvénic fluctuations, correlations in $\mathbf{v}$ and $\mathbf{B}$ are present in all vector components, not just those transverse to the direction of propagation (*Belcher and Davis* 1971). However, any analysis of FIPS observations is restricted to the direction radial to the Sun, as the recovered speed is approximately equal to the radial component of the solar wind velocity. Therefore, Alfvénic fluctuation events will be identified as time periods during which the measured $v_{\rm H+}$ correlates with the radial (with respect to the Sun) component of the magnetic field, $B_{\rm R}$, and the magnetic field magnitude $|\mathbf{B}|$ is nearly constant. The

constant of proportionality between $\delta v_{\mathrm{H}+}$ and δB_{R} will be used to derive the radial propagation speed of the fluctuation from equation (9.10).

9.3.2 Differential streaming of heavy ions

Two periods of propagating Alfvénic fluctuations identified in the FIPS dataset on 19 April 2011, from 00:00:00 to 02:40:00 UTC, and 28 August 2011, from 14:00:00 to 15:00:00 UTC, are shown in Figures 6 and 7, respectively, with proton and heavy ion velocities from FIPS and corresponding magnetic field vector measurements from MAG. In each case, there is a strong correlation between $v_{\mathrm{H}+}$ and B_{R} and a nearly constant $|\mathbf{B}|$. The time resolution of the data from each event varies because of different FIPS instrument operating modes. For each case, the maximum available time resolution was used to derive $v_{\mathrm{H}+}$ in order to resolve the fine structure of the fluctuations. The thermal velocities for the 19 April event did not meet the stability criteria for a reliable recovery and are, as a consequence, not included in this analysis. However, for the 28 August event, the H^{+}, He^{2+}, and $\langle OC \rangle$ thermal velocities are available and are shown in Figure 9.7.

From previous observations (*von Steiger and Zurbuchen* 2006), it is expected that the timescale of interactions between solar wind ions and propagating waves scales as $(q/m)^{\alpha}$, where α is related to the spectral slope of the magnetic field fluctuations. Heavier ions should therefore scatter quickly into the frame of the waves or Alfvénic turbulence, leading to a motion relative to protons, and also excessive heating. In addition, these ions are heated such that their thermal velocities approach that of the wave speed. This process results in differential streaming of heavy ions with respect to the protons in the direction of the magnetic field (*Asbridge et al.* 1976; *Marsch et al.* 1982a; *Zurbuchen et al.* 2000; *Berger et al.* 2011). Since waves tend to propagate outward from the Sun (*Roberts et al.* 1987; *Gosling et al.* 2009), such streaming results in an increased measured solar wind speed of heavy ions.

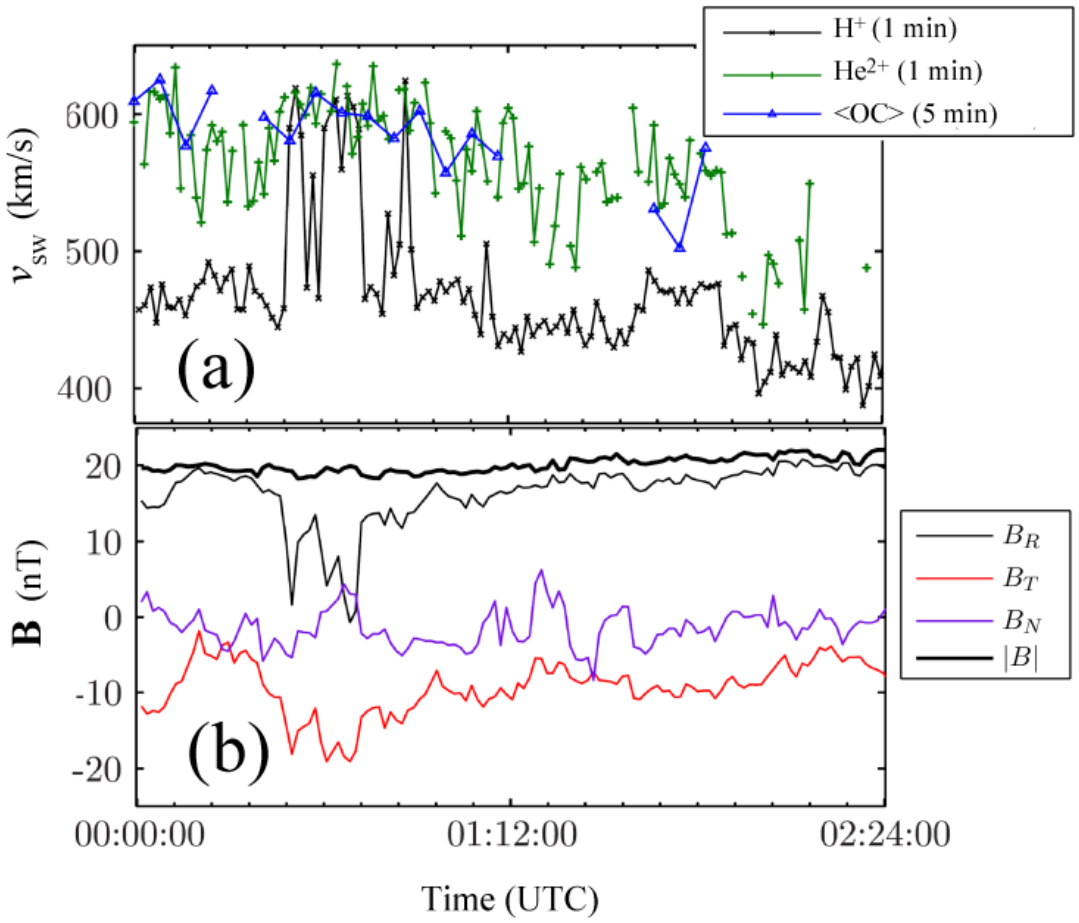

Figure 9.6: Alfvénic fluctuation event for 19 April 2011, from 00:00:00 to 02:40:00 UTC, in instrument survey mode. (a) FIPS-measured bulk speeds of H^+, He^{2+}, and $\langle OC \rangle$ show differential streaming of heavy ions with respect to solar wind protons. Time accumulations for each data product are indicated. (b) Magnetic field measurements from MAG (1-min averages) indicate strong correlations of B_R and V_{H+} with a nearly constant $|\mathbf{B}|$.

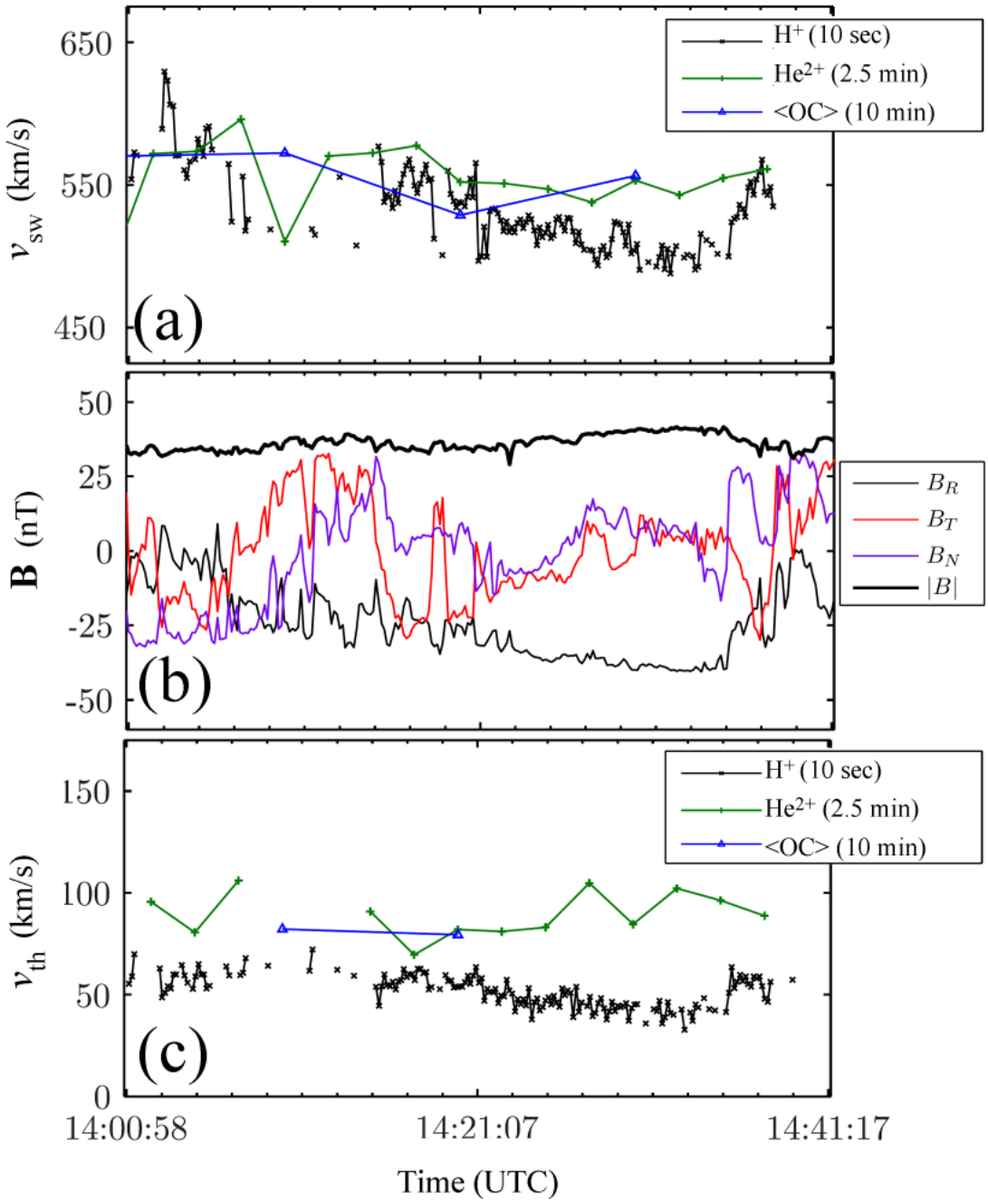

Figure 9.7: Same as Figure 9.6 but for 28 August 2011, from 14:00:00 to 15:00:00 UTC, and with (c) corresponding thermal velocities. During this time period, FIPS was in its burst scanning mode. Time accumulations for each data product are indicated. The magnetic field data have been averaged over 10-s intervals in order to observe the fine structure of the B_R -v_{H+} correlation.

This wave-particle scattering is incomplete, however, and particles have been observed to stream with only a fraction of the propagating wave speed. Some weak correlation between the He^{2+} speed and B_R in Figures 9.6 and 9.7 is further evidence of this incomplete scattering process, as only particles not traveling in the frame of the wave could be susceptible to such interactions. The constant of proportionalities between δv_{H+} and δB_R for the events in Figures 9.6 and 9.7 are -9.2 km/s·nT and -2 km/s·nT, respectively, leading to maximum calculated wave speeds of 175 km/s and 80 km/s. The He^{2+} bulk radial velocities appear to stream ahead of the protons at a fraction of this speed, approximately 70%, for both the 19 April and 28 August events, consistent with the results of *Marsch et al.* (1982a). The $\langle OC \rangle$ bulk velocities also appear to lead the protons by similar factors, though with its longer required accumulation time and consequently fewer available data points, this relationship is less clear.

For the 28 August event, the heavy ion thermal velocities also lead that of the proton velocity, indicating substantial heating for each species when compared with H^+. The average thermal velocity of heavy ions during the event is approximately 90 km/s, consistent with particles scattered into the frame of the propagating wave. Ions with markedly different thermal speeds could result in biases in their recovered bulk velocities, following the derivation in Appendix D.1. However, the effect of this bias would be to decrease the recovered speed of the heavy ions, such that their relative streaming speed would be slightly underestimated.

Although the full time interval of each event appears to have a strong $v_{H+} - B_R$ correlation, there are short periods within each period with no observed differential streaming, i.e., $v_{H+} = v_{heavy}$. These periods can be understood as marking changes in the propagation direction of the fluctuations. Nominally, Alfvén waves are transverse perturbations that travel parallel to the mean field. However, for nonlinear high-amplitude waves, the propagation direction may be altered by large fluctuations.

During these events, although the wave speed $|\mathbf{v}_\mathrm{w}|$ must stay constant, the individual components can change to accommodate an evolving direction of propagation. The periods in Figures 9.6 and 9.7 where $v_{\mathrm{H}+} = v_\mathrm{heavy}$ correspond to times when the magnetic field becomes notably nonradial, i.e., $|B_\mathrm{T}|, |B_\mathrm{N}| > |B_\mathrm{R}|$, in agreement with the analysis of *Berger et al.* (2011) at 1 AU. With the direction of propagation no longer in the radial direction, the differential speed of heavy ions and H^+ measured by FIPS should be small. These observations indicate that Alfvénic fluctuations travel along the local magnetic field direction. High time-cadence measurements are vital for resolving this phenomenon, as average heavy ion velocities over the timescale of the fluctuations can wash out this effect.

9.3.3 Average heavy ion kinetic properties

Because of the limited solar wind dataset of FIPS, it is not possible to identify all potential Alfvénic fluctuation events from instrument observations. Consequently, wave speeds can be identified only for a handful of events for which there are large segments of continuous measurements. However, it is possible to investigate average trends of heavy ion dynamics. Such data for the inner heliosphere (R $<$ 0.5 AU) from FIPS can be compared with previous observations at 1 AU. For the full time range available in 2011, average bulk and thermal velocities were recovered from FIPS measurements.

9.3.3.1 Bulk flow

The average bulk velocities of He^{2+} compared with that of H^+ and $\langle OC \rangle$ are shown in Figure 9.8. At lower solar wind speeds ($v < 450$ km/s), $v_{\mathrm{H}+} \approx v_{\mathrm{He2}+}$, indicating the absence of substantial differential streaming. However, for higher speeds ($v >$ 450 km/s), the helium ions consistently stream ahead of the protons, a trend in good agreement with increased Alfvénic wave activity in coronal-hole winds. For all mea-

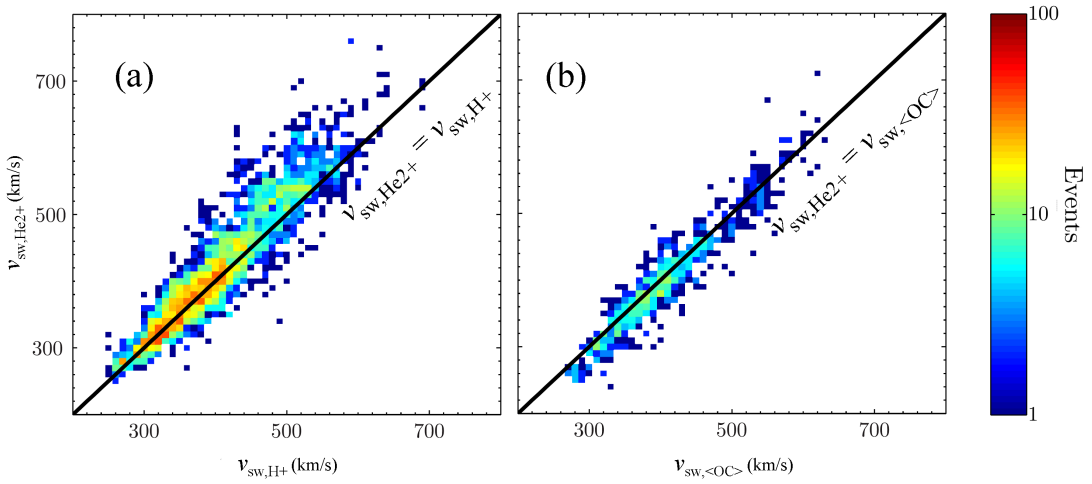

Figure 9.8: $v_{\mathrm{He2+}}$ as a function of (a) $v_{\mathrm{H}}+$ bulk velocities and (b) $\langle OC \rangle$ bulk velocities measured by FIPS. For slower solar wind speeds ($v < 450$ km/s), helium and hydrogen travel together. For higher speed winds ($v > 450$ km/s), at which Alfvénic turbulence becomes more prevalent, the helium ions stream ahead of the protons. Helium and $\langle OC \rangle$ flow together for all measured solar wind speeds.

sured solar wind speeds, however, there does not appear to be any average streaming between alpha particles ($m = 4$ amu) and heavier ions ($m =$ 12-16 amu). Such behavior is consistent with an m/q-dependent physical processes, in which higher-m/q ions all flow with the protons in the slower solar wind and all stream ahead of the protons with similar velocities during times of increased Alfvénic wave activity in the fast wind.

9.3.3.2 Thermal properties

The thermal velocities are compared between He^{2+} and H^{+} and between He^{2+} and $\langle OC \rangle$ in Figure 9.9. Lines indicating thermal equilibrium, equal thermal speeds, and the maximum differential thermal speeds ($v_{\mathrm{th,O5+}}/v_{\mathrm{th,H}} \approx 2.33$ at 3.5 solar radii (*Kohl et al.* 1998) observed close to the Sun by UVCS are included in the figure. Although UVCS observed temperatures only of neutral hydrogen, similar trends are expected to hold for that of H^{+}. Heavy ions here are observed to be substantially

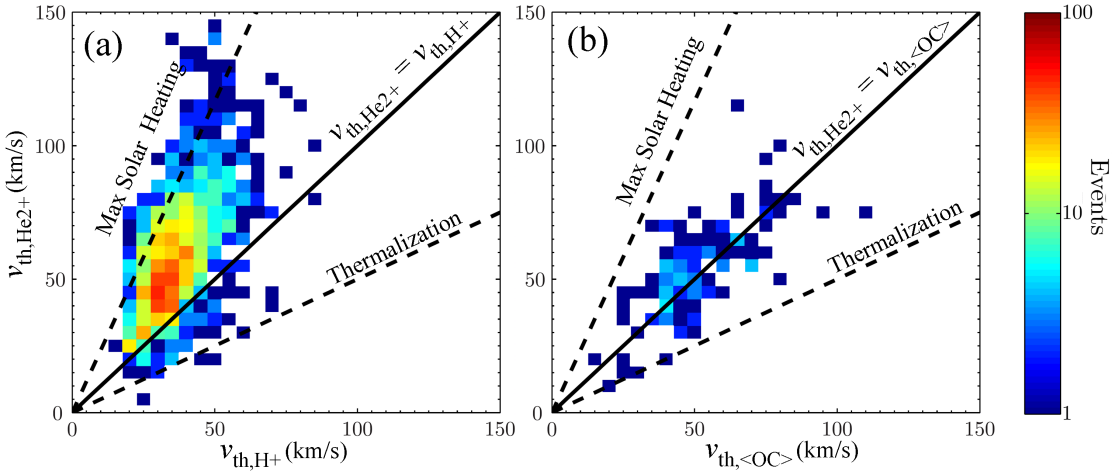

Figure 9.9: $v_{\mathrm{th,He2+}}$ as a function of (a) $v_{\mathrm{th,H+}}$ bulk velocities and (b) $\langle OC \rangle$ bulk velocities measured by FIPS. Helium is substantially heated with respect to the protons; on average $T_{\mathrm{He2+}}/T_{\mathrm{H+}} \approx 9$. Both He^{2+} and $\langle OC \rangle$ have similar thermal speeds, indicating substantial heating of the heavier ions as well.

hotter than protons, with practically no observed alpha particles in thermal equilibrium with the protons as observed at 1 AU. The maximum $v_{\mathrm{th,He2+}}/v_{\mathrm{th,H+}}$ from FIPS, however, stayed within the bounds of UVCS observations, indicating that a substantial additional heating source outside the solar wind acceleration region may not be required to explain FIPS observations. Further, Figure 9.9b indicates that the thermal velocities for He^{2+} and $\langle OC \rangle$, like their bulk velocities, are nearly identical.

For Parker spiral geometry and a solar wind density that scales as $1/R^2$, the Alfvén speed (v_{A}) increases by about a factor of 2 from 1 AU to 0.4 AU, the average distance of MESSENGER from the Sun during this time period. The proton thermal speed, however, for a measured $R^{-0.9}$ scaling for temperature in the fast solar wind in the inner heliosphere (*Totten et al.* 1995), increases only by a factor of approximately 1.5 from 1 AU to 0.4 AU, such that the overall ratio $v_{\mathrm{A}}/v_{\mathrm{th,H+}}$ will increase by a factor of 1.3, i.e., the proton thermal velocity does not scale with the Alfvén speed. As discussed in Section 9.3.2, He^{2+} ions can be heated by propagating Alfvénic fluctuations such that their thermal speed approaches that of the Alfvén speed. It follows that

the helium thermal velocity during these times should scale with the Alfvén velocity. At 1 AU, $T_{\mathrm{He2+}}/T_{\mathrm{H+}} = 4$, i.e., equal proton and alpha thermal speeds correspond to time periods of increased magnetic wave activity and differential streaming (*Kasper et al.* 2008). Inside 1 AU, therefore, if He^{2+} particles have their thermal speeds ordered by the Alfvén speed, a peak in the ratio $T_{\mathrm{He2+}}/T_{\mathrm{H+}} \approx 8$ is expected. From Helios, $T_{\mathrm{He2+}}/T_{\mathrm{H+}}$ in the inner heliosphere was 5-6 (*Marsch et al.* 1982a). The average $T_{\mathrm{He2+}}/T_{\mathrm{H+}}$ from FIPS observations, ≈11, is higher than this scaled ratio and nearly twice that of the previous Helios observations.

Temperature enhancements measured by FIPS as a function of $v_{\mathrm{H+}}$ for He^{2+} and $\langle OC \rangle$ are shown in Figure 9.10 for an average $\langle OC \rangle$ mass of 14.4 amu (*von Steiger et al.* 2000). In the fast wind, where Alfvénic turbulence is most prevalent, the temperature ratio between the heavy ions and protons is largest, with the maximum ratio approaching the maximum observed heavy ion heating from UVCS measurements (*Kohl et al.* 1998). The minimum ratios occur for the slowest measured solar wind speeds, at which Coulomb collisions are expected to play an increasingly important role in heavy ion kinematics. The temperature ratio $T_{\langle \mathrm{OC} \rangle}/T_{\mathrm{He2+}}$ is approximately equal to $m_{\langle \mathrm{OC} \rangle}/m_{\mathrm{He2+}}$, consistent with results shown in Figure 9.9b, indicating that heavy ions move with similar thermal velocities. From this analysis, measurements of $T_{\mathrm{ion}}/T_{\mathrm{H+}}$ in the inner heliosphere appear to be ordered by the solar wind regime, consistent with similar analysis at 1 AU by *Kasper et al.* (2008), but with substantially higher observed alpha particle temperatures.

As discussed in Appendix D.1, the recovery of the thermal velocities of ion species relies on the assumption of an isotropic plasma. If substantial anisotropy is present in the He^{2+} particles with respect to those of H^+, the ratio of the recovered thermal velocities could be biased. Furthermore, there may also be a larger or variable suprathermal population for heavier ions than for H^+, enhancing the estimates of alpha particle temperatures. Finally, it is possible that the solar wind populations

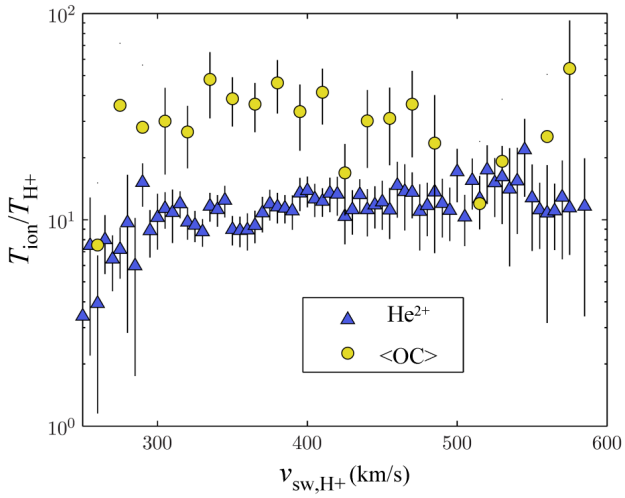

Figure 9.10: $T_{\rm ion}/T_{\rm H+}$ as a function of solar wind speed for He^{2+} and $\langle OC \rangle$. The error bars for each point indicate counting uncertainty, with the percent error scaling as the inverse square root of the number of measurements, $1/\sqrt{N}$. With decreasing solar wind speed, i.e., less Alfvénic turbulence, the plasma tends toward a lower temperature ratio. The temperature ratio $T_{\langle \rm OC \rangle}/T_{\rm He2+}$ is approximately equal to $m_{\langle \rm OC \rangle}/m_{\rm He2+}$, implying similar thermal speeds for heavy ions, as shown in Figure 9.9b.

most easily measured by FIPS statistically tend to have higher He^{2+} temperatures. As evidenced by data gaps in the observations of *Marsch et al.* 1982a close to the Sun, available Helios observations may exhibit an opposite bias. Alpha particles and protons become increasingly more difficult to separate with increasing $T_{\rm He2+}/T_{\rm H+}$ ratios using only electrostatic deflection, leading to a possible explanation for the increased temperatures measured by MESSENGER when compared with Helios observations. A more detailed analysis will be necessary to characterize all potential biases in the recovery. Nevertheless, from FIPS observations, there are no recovered thermal velocities that indicate that He^{2+} and H^+ are in thermal equilibrium, in striking contrast to data collected at 1 AU.

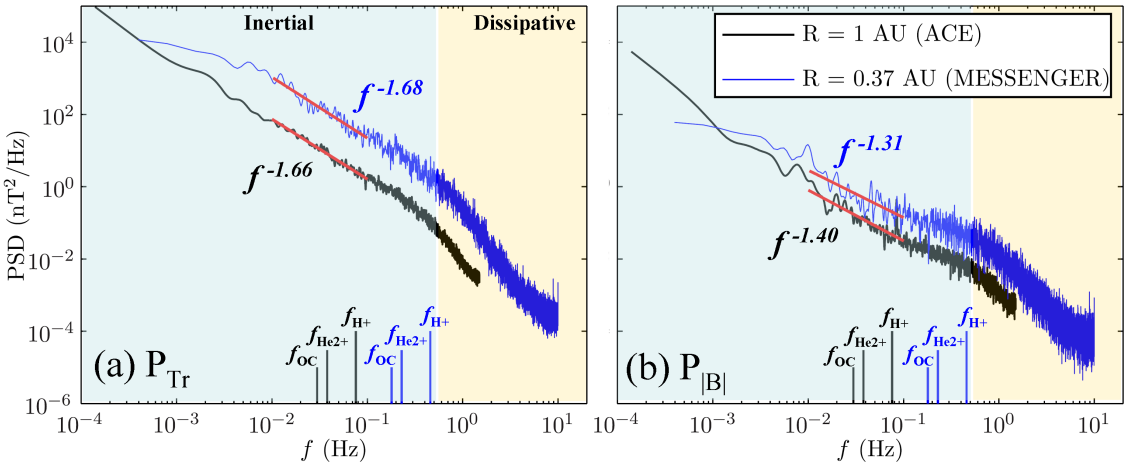

Figure 9.11: Representative high-resolution spectra for MESSENGER (in blue) and ACE (in black) during a time of coronal-hole-associated wind (ACE 6 March 2011, 04:00:00 07:59:59 UTC; MESSENGER 29 June 2011, 01:26:00 02:08:00 UTC). Power spectra are shown for (a) the trace of three-dimensional spectral density distributions of $\mathbf{B}$, i.e. P_{Tr} and (b) fluctuations in the magnitude of $\mathbf{B}$, i.e., $P_{|\mathbf{B}|}$. The local gyro-frequencies of key ions (H^+, He^{2+}, and $\langle OC \rangle$) are computed from average $|\mathbf{B}|$ values during this time period, ≈5 nT and ≈30 nT for ACE and MESSENGER, respectively.

9.3.3.3 Magnetic turbulent heating

To interpret the non-thermal heating and differential acceleration of heavy ions as a function of heliocentric distance, a comparison of two representative spectra of locally measured magnetic field data in the fast solar wind ($v_{\mathrm{H+}} \approx 500$ km/s) at 0.37 and 1 AU are shown in Figure 9.11. The ACE data were measured by the magnetometer on ACE (*Smith et al.* 1998) with a maximum time-resolution of 3 vectors/s. MESSENGER data were measured with a maximum time-resolution of 20 vectors/s. Spectral densities for the trace of the total power of fluctuations in the magnetic field components (P_{Tr}) and the total power of fluctuations in the magnetic field magnitude ($P_{|\mathbf{B}|}$) are shown. For each spectrum, there is a power-law fit in the inertial frequency range 10^{-2} Hz$< f < 10^{-1}$Hz. Figure 9.11 also marks the gyro-frequencies of H^+, He^{2+} and $\langle OC \rangle$.

In the low-frequency inertial regime, the spectral shape of P_{Tr} is almost independent of heliocentric distance, consistent with Helios observations (*Marsch* 1991). The power-law fits in this region show a slope consistent with Kolmogorov scaling, i.e., $f^{-5/3}$ for both spectra. However, the spectral power at 0.37 AU is elevated by a factor of 20 in the frequency range 10^{-3} Hz$< f <$ 100 Hz. Following the Wentzel-Kramers-Brillouin (WKB) approximation and under steady plasma conditions, $P_{|\mathbf{B}|}$ is expected to increase follow r^{-3} dependence in the inner heliosphere (*Isenberg and Hollweg* 1982), consistent with the observed change in $P_{|\mathbf{B}|}$.

As discussed in Section 9.3.1, one view of the fluctuations in $\mathbf{B}$ can be expressed as a combination of inward- and outward-propagating Alfvén waves that interact in a non-linear fashion (*Marsch* 1991; *M. L. Goldstein et al.* 1995). Interactions between these large-amplitude waves and particles are expected to be the physical reason for the differential streaming of solar wind heavy ions. The spectral density of $P_{|\mathbf{B}|}$ is smaller than P_{Tr} by over two orders of magnitude, indicating that these inertial range fluctuations are approximately incompressible. This incompressibility is consistent with the observed properties of Alfvénic fluctuations for which $|\mathbf{B}|$ remains nearly constant but with large fluctuations in all individual components. However, the spectral slope of $P_{|\mathbf{B}|}$ is substantially flatter than that of P_{Tr}, indicating increasingly compressible turbulence toward the transition to the dissipative range.

At both ACE and MESSENGER, the transition from the incompressible inertial range to the compressible dissipation range manifests as a change in spectral slope near $f = 1$ Hz. Although the 1 AU heavy ion gyro-frequencies are well within the inertial range, the corresponding gyro-frequencies at R = 0.37 AU occur near this point. Despite important Doppler shifts not having been included in the labeling of ion gyro-frequencies in Figure 9.11, a few important conclusions can be drawn from this analysis: (1) the transition from inertial to dissipation ranges and the kinetic timescale of heavy ions and protons in the inner heliosphere evolve differently with

heliocentric distance but span a similar band of frequencies; (2) solar wind heavy ions will necessarily interact with solar wind turbulence at lower frequencies, and consequently will always experience a substantially larger power level than that of the protons, e.g., $P_{\rm Tr}(f_{\rm OC}) \approx 5P_{\rm Tr}(f_{\rm H+})$ in the inertial range; (3) there appears to be a distinct increase in the compressive nature of fluctuations in the frequency range near the ion gyro-frequencies in the inner heliosphere; and (4) for local heavy ion resonant interaction, one would expect spectral breaks at the locations of heavy ion gyro-frequencies. The absence of these markers suggests that either the majority of the heavy ion wave-heating process occurs close to the Sun, or that the turbulent cascade masks any signature of a gradual heliospheric process.

9.4 Concluding remarks

Observations by the FIPS sensor on MESSENGER constitute the first measurements of the properties of heavier ions in the solar wind at ∼0.3 AU, and the first measurements of any solar wind properties at this heliocentric distance since observations by the Helios spacecraft 30 years ago. Our results, as in the case of the solar wind at 1 AU, are divided by the dominant source of turbulence, i.e., by fast solar wind, which has substantial Alfvénic turbulence, versus slow solar wind, which has more compressive turbulence. In the case of the slow wind, the protons and heavier ions tend to have the same flow speed and approximately equal temperatures, which were presumably set in collision-dominated processes in the solar corona. In the case of the fast wind, the presence of Alfvénic turbulence causes the heavier ions to flow faster than protons by a sizable fraction of the Alfvén speed. Although this fraction is similar to that observed at 1 AU, the Alfvén speed at 0.3 AU is approximately a factor 3 larger than at Earth. Similarly, the heavier ion temperatures are markedly larger in the fast solar wind than those of the protons.

The fast solar wind appears to be a blend of thermal speeds that are determined in

the heliosphere and those determined in the solar corona. The ratios of ion-to-proton temperatures at ∼0.3 AU are larger than would result if the ions simply acquired a fraction of the Alfvén speed as their thermal speed, as is observed at 1 AU, providing evidence of m/q-dependent heating that has been observed spectroscopically at the Sun. An examination of the magnetic environment at both MESSENGER and ACE reinforces this observation, as although the turbulent power associated with Alfvénic fluctuations is an order of magnitude higher at R $\sim$ 0.3 AU than at Earth, there is no strong spectral signature of local heating at either heliocentric distance.

Despite the fact that the FIPS sensor on MESSENGER, with its view obstructed by the spacecraft sunshade, is not ideally suited for observing the solar wind, it has been possible to draw conclusions about the properties of heavier ions in the solar wind at 0.3 AU, and thus place constraints on the mechanisms of the acceleration and evolution of the solar wind as it propagates outward from the Sun. Moreover, we should expect that the accuracy of the results presented here will improve with additional observations and sharpened statistics as MESSENGER continues in its extended mission.

CHAPTER X

Observations of interstellar helium pickup ions in the inner heliosphere

This chapter is taken from: Gershman, D. J., G. Gloeckler, J. A. Gilbert, J. M. Raines, L. A. Fisk, S. C. Solomon, E. C. Stone, and T. H. Zurbuchen (2012), Observations of interstellar helium pickup ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

10.1 Introduction

As the Sun travels through the local interstellar medium (LISM), neutral gas flows into the heliosphere. These particles, consisting primarily of hydrogen and helium, follow ballistic trajectories through the solar system, guided only by the combined gravitational and radiation pressure forces of the Sun (*Fahr* 1968; *Blum and Fahr* 1970). Near the Sun's vicinity, this neutral population is lost via ionization reactions, whereby neutral particles of the interstellar gas become ions that are embedded in and picked up by the constantly expanding solar atmosphere, the solar wind. Initial analytical studies of this process (*Blum and Fahr* 1970; *Fahr* 1971) focused on

trajectories of interstellar neutral hydrogen. Because of substantial Lyman-α radiation pressure, a cavity devoid of interstellar hydrogen forms around the Sun (*Thomas* 1978; *Fahr* 1978). For interstellar helium, however, which is largely unaffected by radiation pressure, the Sun acts as a gravitational lens, forming a neutral focusing cone downwind from the direction of interstellar flow, as illustrated in Figure 10.1. This cone was first observed remotely through backscattered He 58.4 nm observations by an instrument on the Solrad 11 satellite (*Weller and Meier* 1981). The structure (i.e. width and peak density) of the cone is controlled only by the kinetic properties of the LISM neutral helium gas and the ionization rates of helium in the heliosphere (*Feldman et al.* 1972; *Blum et al.* 1975; *Meier* 1977; *Thomas* 1978; *Wu and Judge* 1979).

Ionized helium particles, although lost to the neutral population, are measured by spaceborne ion mass spectrometers as so-called "pickup ions", named after their cyclic motion about a magnetic field that is embedded in the expanding solar wind. As first measured by *Möbius et al.* (1985a), such observations are important for not only verifying the accuracy of the deduced properties and flow direction of the interstellar medium (*Möbius et al.* 1995; *Möbius* 1996; *Noda et al.* 2001; *Gloeckler and Geiss* 2001; *Gloeckler et al.* 2004), but also as a useful diagnostic for understanding the evolution of physics of the ion pickup process (*Gloeckler et al.* 1995a; *Gloeckler and Geiss* 1998a; *Saul et al.* 2007), a phenomenon common in planetary magnetospheres (e.g., *Hartle et al.* 1982; *Zurbuchen et al.* 2008) or around comets (*Ipavich et al.* 1986; *Neugebauer et al.* 1989; *Gloeckler et al.* 2000b).

The properties of the LISM helium have been constrained from in situ neutral (*Witte et al.* 2004; *Möbius et al.* 2012) and pickup ion (*Gloeckler et al.* 2004) measurements and from remote sensing observations of helium glow (*Vallerga et al.* 2004; *Lallement et al.* 2004) and solar wind charge exchange emission (*Koutroumpa et al.* 2009) using instrument forward models to fit observations. *Möbius et al.* (2004)

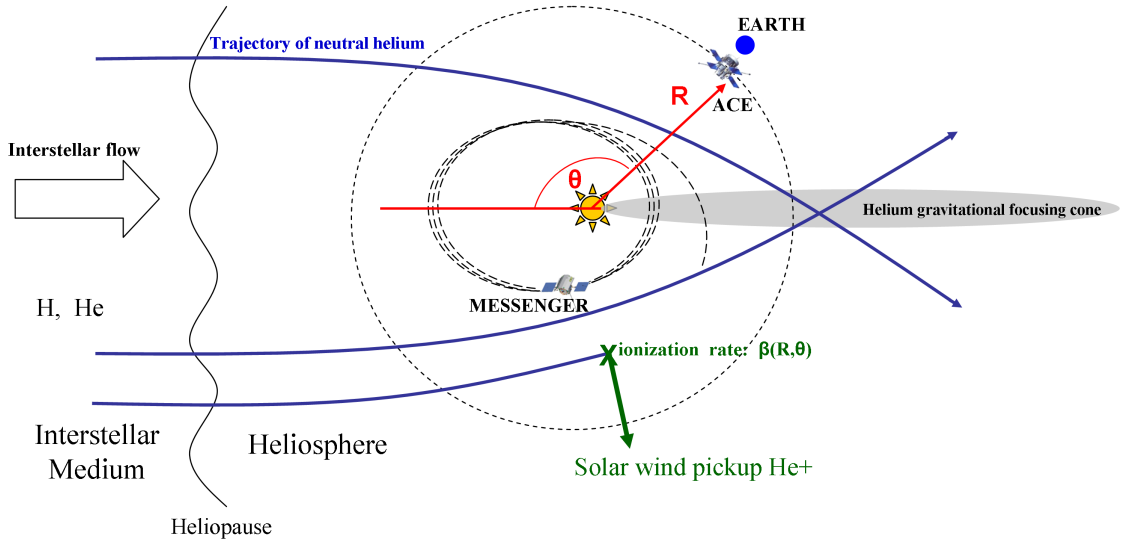

Figure 10.1: Illustration of the trajectories of neutral particles that converge to form the helium gravitational focusing cone directly downwind from the direction of interstellar flow. Neutral helium throughout the heliosphere can be ionized and injected into the solar wind as pickup ions that can be measured in situ by SWICS on ACE, which orbits the Sun at L1 (1AU) and FIPS on MESSENGER, which passed through the focusing cone several times during its inner heliospheric passage to Mercury. The coordinates (R,θ) are defined as the heliocentric distance and the angle from the interstellar flow direction, respectively.

summarized a coordinated effort to combine these different diagnostics to form a set of consistent parameters for the interstellar gas flow velocity vector and temperature in J2000 ecliptic coordinates, arriving at longitude $\lambda_\infty = 75.38° \pm 0.56°$, latitude $\beta_\infty = -5.31° \pm 0.28°$, velocity $v_\infty = 26.24 \pm 0.45$ km/s, and temperature $T_\infty = 6306 \pm 390$K. However, new results from the Interstellar Boundary Explorer (IBEX) (*McComas et al.* 2004a), a sensor for energetic neutrals, has produced an alternate set of parameters of $\lambda_\infty = 79° \pm 3°$, $\beta_\infty = -4.9° \pm 0.2°$, velocity $v_\infty = 23.5 \pm 3$ km/s, (*Möbius et al.* 2012) and $T_\infty \approx 6200$K (*Bzowski et al.* 2012).

To date, the only available He^+ pickup ion measurements have been for heliocentric distances ≥ 1 AU. Although these measurements do span a variety of heliographic latitudes, the only studies of the longitudinal structure of the cone use data collected at 1 AU in the ecliptic and from a single near-downwind pass near Jupiter (*McComas et al.* 2004b) with the Cassini Plasma Spectrometer (CAPS) on the Cassini spacecraft (*Young et al.* 2004). The Fast Imaging Plasma Spectrometer (FIPS), one of two charged particle sensors on the Energetic Particle and Plasma Spectrometer (EPPS) instrument (*Andrews et al.* 2007) on the MErcury Surface, Space ENvironment, GEochemistry, and Ranging (MESSENGER) spacecraft (*Solomon et al.* 2001) has yielded the first in situ measurements of He^+ inside of 1 AU and the first observations that span the full longitudinal structure of the gravitational focusing cone at any other heliocentric distance. Although also confined to the ecliptic plane, the spacecraft's cruise trajectory, as shown in Figure 10.1, led to several transits of the gravitational focusing cone inside 1 AU. This paper reports on the unprecedented dataset provided by those observing opportunities. When combined with data from the Solar Wind Ion Composition Spectrometer (SWICS) instrument (*Gloeckler et al.* 1998) on the Advanced Composition Explorer (ACE) spacecraft (*Stone et al.* 1998), constraints on the transport and large scale diffusion of He^+ ions within 1 AU can be developed to test the effectiveness of using near-Earth measurements to describe the longitudinal

structure of neutral helium population in the heliosphere.

It can be difficult to use pickup ion measurements to tightly constrain the direction of interstellar flow, because of the possibility of ion transport from the initial pickup location (*Möbius et al.* 1995) and from the fact that the measured pickup ion density is a strong function of ambient solar wind conditions (*Saul et al.* 2003; *Möbius et al.* 2010), potentially creating biases in the determination of cone center from any given downwind pass. We will argue that the cone structure can be recovered by averaging over multiple passes during times where the pickup ion production rate is expected to be steady-state, i.e. solar minimum. This technique provides improved accuracy of the determination of the downwind direction from pickup ion measurements and also provides information about ionization processes not available from 1 AU data.

In Section 10.2 of this paper, we discuss the interpretation of 1 AU pickup ion measurements, and in Section 10.3 we describe the most recent unusually low solar minimum and why it is uniquely suited for using pickup ion measurements to map the neutral density distribution in the heliosphere. In Section 10.4 , we combine data from MESSENGER/FIPS and ACE/SWICS to analyze the gravitational focusing cone structure at $R = 0.3$ AU and 1 AU between 2007 and 2009. Finally, in Section 10.5 , we discuss the implications of these measurements for both electron-impact ionization rates in the inner heliosphere and possible large-scale transport effects of pickup helium.

10.2 Limitations of Pickup Ion Observations

The abundance and dynamic state of pickup ion spectra at any heliocentric distance depend on: (1) the neutral density distribution throughout the heliosphere, (2) the ionization profile(s) for the production of new pickup ions, and (3) the evolution and dynamics of pickup ion velocity distributions in space. Models of each of these processes have been developed and successfully used to match measurements,

constraining important parameters of the underlying physical processes.

10.2.1 Neutral Density Distribution in the Heliosphere

A LISM neutral helium model requires knowledge of not only the direction and properties of interstellar flow, but also of the three-dimensional, time-dependent ionization rates. Far from the Sun (>1 AU), photoionization by extreme ultraviolet (EUV) photons is the dominant loss mechanism for helium. Losses associated with resonant charge exchange reactions are lower by an order of magnitude (*McMullin et al.* 2004) and are not considered here. As a simplification, photoionization is sometimes modeled as being spherically symmetric, but during solar minimum there can be strong latitudinal variations in EUV coronal emissions, namely a ≈20% reduction in the polar regions associated with coronal holes (*Auchère et al.* 2005).

Electron-impact ionization, although much weaker than photoionization at 1 AU, does not scale as $1/R^2$, and therefore can become important and possibly dominant close to the Sun (<1 AU), where few detailed measurements have been made of solar wind electron distribution functions (*Rucinski and Fahr* 1989; *McMullin et al.* 2004). Consequently, its radial dependence is much less known than heliospheric, optically thin photoionization, which can be readily scaled from 1 AU observations to other distances as $1/R^2$. The electron distributions found in the steady, high latitude ($> 30°$) winds have been found to be markedly different than those observed in the lower latitudes (*Maksimovic et al.* 2000), leading to lower than expected ionization at high latitudes from the reduced densities and temperatures typically associated with the coronal hole winds.

10.2.2 He^+ Production Rates

Similar uncertainties exist for the production rates of pickup ions that change on a much smaller time scale. Large variations in solar activity produce changes in the

photoionization rate on the time scales of years. However, a sudden coronal mass ejection or solar flare can have an immediate, albeit transient, impact on ionization rates near the Sun (*Rucinski et al.* 2003). These transients cause important uncertainties in the determination of the local production of He^+ and have the potential to produce errors that propagate when combined with models of pickup transport.

10.2.3 Pickup Ion Velocity Distributions

Numerous analytical and numerical models have been proposed to predict pickup ion transport, and a number of these models have been compared with pickup ion data (e.g., *Vasyliunas and Siscoe* 1976; *Isenberg* 1997; *Schwadron* 1998). Typically, these models include four key physical process: (1) neutrals are relatively slow moving compared to the solar wind and, upon ionization, are instantaneously injected into a magnetic gyration; (2) newly formed ions gyrate around the frozen-in magnetic field and are convected outward with the solar wind speed ($v = v_{\rm sw}$), quickly forming a ring distribution; (3) ions preferentially pitch angle scatter around into a shell; and (4) ions cool due to the expanding solar wind and undergo energy scattering, thus populating inner shells (i.e.,$v < v_{\rm sw}$ in the solar wind frame).

He^+ measurements combined with forward models have been used to show that ions appear to cool near adiabatically, at least in the outward hemisphere most often observed ($v > v_{\rm sw}$ in the spacecraft frame) (*Saul et al.* 2009). However, this angular scattering, especially during time periods of radial magnetic field, is not isotropic, and pickup ions tend to populate the sunward sectors moreso than the anti-sunward sectors (*Gloeckler and Geiss* 1998a; *Saul et al.* 2007). The detailed structure of the velocity distributions in each sector has not yet been determined. Such velocity anisotropies are predicted to strongly influence large-scale diffusion and ion transport in the heliosphere (*Möbius et al.* 1995). With only measurements of the downwind direction at one heliocentric distance (i.e. 1 AU), potential effects of large-scale

diffusion and ion transport cannot be easily characterized.

10.3 Observations at Solar Minimum

The most recent solar minimum (2007-2009) is a period of unusually quiet and constant solar activity, with overall reduced levels of solar wind mass flux (*McComas et al.* 2008; *Fisk and Zhao* 2008) and radiation emission from the Sun (*Woods* 2010). The photoionization and electron-impact ionization rates are reported to be uncharacteristically low and nearly constant when compared with those of previous minima during the past 50 years during which space data exists. Furthermore, the magnetic flux configuration on the Sun is dominated by low-latitude coronal holes (*Abramenko et al.* 2010), creating a more steady-state, spherically symmetric ionization source. Thus, this time-period presents a unique opportunity for the quantitative study of heliospheric pickup ions.

10.3.1 Helium Ionization Rates

The derived EUV photoionization rate (*McMullin et al.* 2002) during the 2007-2009 period is constant to within approximately 25% of the mean value as shown in Figure 10.2a. These rates were calculated using measured EUV flux from the Solar Extreme-ultraviolet Monitor (SEM) sensor on the SOlar and Heliospheric Observatory (SOHO), photoionization cross sections from *Samson et al.* (1994), and the solar spectral irradiance from the Whole Heliosphere Interval 2008 Version 2 described by *Woods et al.* (2009). When compared with the analogous rates used by *Bzowski et al.* (2012), calculated with a different set of measurements and irradiance spectra, there is agreement to within 10%. An approximately steady-state ionization rate for 2-3 years should result in a nearly steady state neutral helium distribution within the entire heliosphere.

The *Rucinski and Fahr* (1989) model is used to describe electron-impact ioniza-

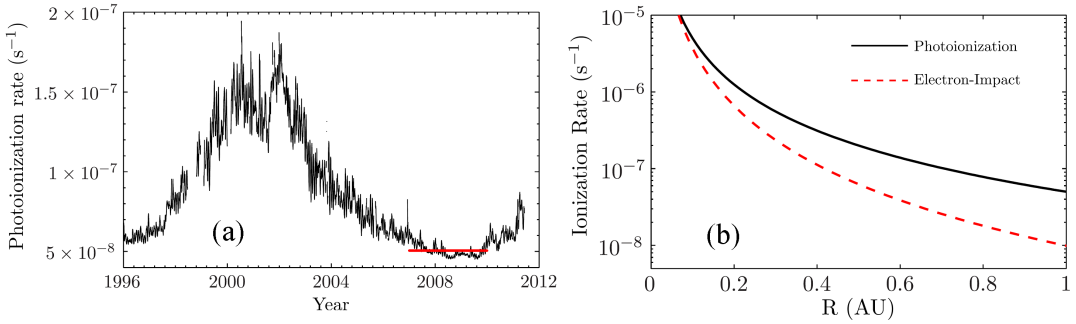

Figure 10.2: (a) Derived helium photoionization rateat 1 AU following the procedure given by *McMullin et al.* (2002). During the most recent solar minimum in 2007-2009, the photoionization rate at 1 AU stayed constant at approximately 5×10^{-8} s^{-1}. (b) Radial profiles of photoionization and electron-impact ionization rate following *Rucinski and Fahr* (1989) for the most recent solar minimum. At 1 AU the electron-impact ionization rate is about 20% of the photoionization rate but becomes increasingly important inside $R = 0.5$ AU.

tion using a Maxwell-Boltzmann description of the solar wind electrons with density and temperature scaling for both core and halo populations. Here, the model input electron densities and temperatures at 1 AU were all reduced by 20% and 10% respectively, in line with recent observations by Ulysses (*McComas et al.* 2008; *Fisk and Zhao* 2008). Electron temperature radial dependencies are also available from Ulysses observations of polar coronal holes (*Issautier et al.* 1998; *Maksimovic et al.* 2000). However, since our in situ measurements are made close to the Sun, we use the original radial dependencies from *Rucinski and Fahr* (1989) that were derived from observations of electron distribution functions made inside of 1 AU. The radial profile of the calculated electron-impact ionization rate is compared with that of the photoionization rate in Figure 10.2b. As expected, electron-impact ionization becomes important closest to the Sun, within 0.5 AU. At 1 AU, the electron-impact value during this time period is approximately 1×10^{-8} s^{-1}, about 20% of the corresponding photoionization rate.

10.3.2 Neutral Helium Distribution from 2007-2009

Given the large scale properties of the LISM penetrating into the heliosphere, the *Thomas* (1978) hot interstellar gas model computes the neutral helium density throughout the heliosphere for a steady-state, spherically symmetric ionization that varies as $1/R^2$. As discussed above, the assumptions of steady-state, spherically symmetric ionization are most appropriately applied to this most recent solar minimum. The equations from *Thomas* (1978) were modified with an additional integral as discussed in *Rucinski and Fahr* (1989), to accommodate an arbitrary ionization radial dependence, thereby enabling the inclusion of electron-impact ionization effects.

The neutral density, n, depends on two variables, R and θ, where θ is the angle from the upwind direction, such that $\theta = 180^\circ$ lies along the cone center, as shown in Figure 10.1. Figure 10.3a shows the calculated neutral density profiles upwind ($\theta = 0^\circ$) and downwind ($\theta = 180^\circ$) using interstellar parameters from *Möbius et al.* (2004) (Model A) and a combination of parameters from *Möbius et al.* (2012) and *Bzowski et al.* (2012) (Model B) with and without the inclusion of electron-impact ionization. The corresponding downwind/upwind enhancements are shown in Figure 10.3b. Both sets of interstellar parameters give qualitatively similar density profiles and maximum downwind enhancements. The effect of electron-impact ionization is to push the neutral helium farther from the Sun by about 0.1 AU and reduce the maximum downwind enhancement inside $R = 0.3$ AU with the 1 AU enhancement remaining relatively unchanged. The cone structure at $R = 0.3$ AU is therefore expected to be much more sensitive to the radial dependence of electron-impact ionization.

Similar downwind enhancements over a given set of radial distances implies that the longitudinal structure is nearly independent of heliocentric distance, allowing for the separation of $n(R, \theta)$ into $n(R)n(\theta)$ for that range. Furthermore, since measurements with $\theta < 90^\circ$ should also depend only on R, such a separation will also be possible for the crosswind/upwind longitudes regardless of cone structure.

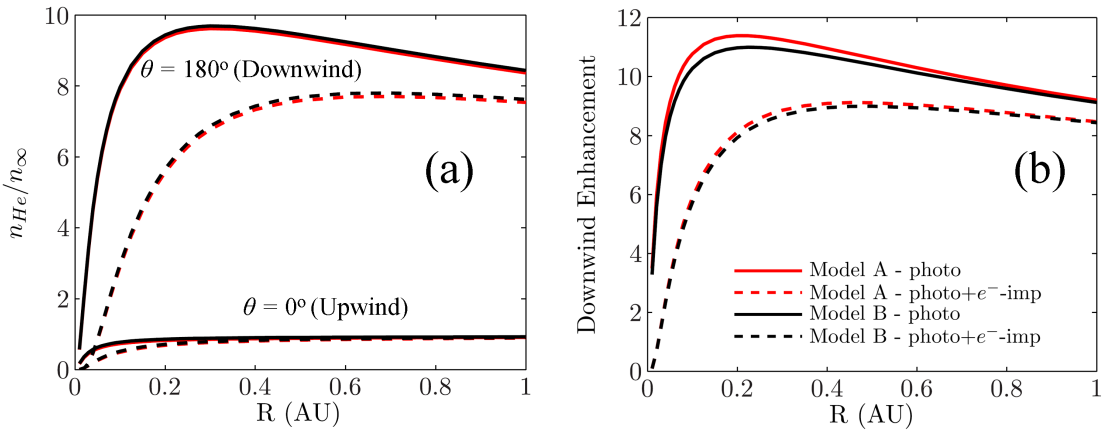

Figure 10.3: (a) Neutral density distribution of helium as a function of heliocentric distance using the LISM parameters from Model A (*Möbius et al.* 2004) and Model B (*Möbius et al.* 2012; *Bzowski et al.* 2012) for upwind ($\theta = 0°$) and downwind ($\theta = 180°$) profiles with and without the inclusion of the effects of electron-impact ionization. The strongest enhancement in the neutral helium occurs at about $R = 0.3$ AU. The densities here are normalized by the LISM value, n_∞. (b) The downwind/upwind ratios from part (a) as a function of heliocentric distance.

10.3.3 Pickup Ions as a Proxy for Neutrals

As detailed in Appendices E.1 and E.2, a separability of the neutral density $n(R,\theta) \approx n(R)n(\theta)$ directly translates into a separability in measured pickup ion velocity distributions for MESSENGER/FIPS and ACE/SWICS, i.e., $n_{\text{obs,ion}}(R,\theta) \approx n(\theta) \cdot R^{-\alpha}$. Under the assumption that this separation is valid, measured pickup ion densities can be used as proxy for the neutral density distribution in the heliosphere. However, there can be situations where this separation is not valid. Such cases must be considered before scaling and interpreting instrument observations.

Depending on the speed of pitch angle scattering, there is the possibility of large-scale diffusion and transport of pickup ions leading to a possible offset and/or spreading of the focusing cone location that increases with heliocentric distance. For such a case, $n_{\text{ion,obs}}$ is no longer separable. However, due to MESSENGER's orbit, the cone transits (i.e. $30° < \lambda < 120°$) are confined between $R = 0.3$ AU and 0.4 AU, with ob-

servations at $R > 0.4$ AU corresponding to upwind longitudes that are independent of θ regardless of any large-scale diffusion that may be present. Therefore, it is unlikely that any transport effects will have sufficient time to evolve such that they could bias any radial scaling of FIPS observations.

As shown by *Saul et al.* (2003, 2007) and *Möbius et al.* (2010), the measured pickup ion density can be distorted through solar wind stream interaction regions and the orientation of the interplanetary magnetic field. These distortions are statistically independent of ecliptic longitude because they are caused by a rotating Sun. Their effects should average to zero if the time scale of their variation is faster than the time scale of changes in the neutral density distribution in the heliosphere, as should be the case for this most recent solar minimum.

With a sufficient number of MESSENGER spacecraft rotations, the variations in the FIPS field of view (FOV) should also not contribute to any apparent longitudinal structure in pickup ion measurements. Therefore, under these assumptions, and if the radial dependence, assumed to be of the form $R^{-\alpha}$, of $n_{\mathrm{obs,ion}}$ at any set of longitudes is measured (for example, in the upwind/crosswind directions), then all $n_{\mathrm{obs,ion}}$ can be scaled to remove any dependence on R. This scaling enables the recovery of the relative neutral cone structure, $n(\theta)$, as a function of only θ, from a set of pickup helium observations at different heliocentric distances.

10.4 He^+ Measurements

The data analysis here consists of double-coincidence measurements from two spaceborne time-of-flight mass spectrometers sensors, MESSENGER/FIPS ($0.3 < R < 0.7$ AU), and ACE/SWICS ($R \sim 1$ AU). Observations from ACE/SWICS are derived using a recently developed double-coincidence data product introduced by *Gilbert et al.* (2012), as He^+ can be easily resolved from other species without the need for a solid state detector that is nominally required to determine charge state

composition of the solar wind. FIPS, due to its imposed mass and power constraints, does not have triple-coincidence capabilities. Although the absence of energy measurements limits analyses with respect to solar wind ions [*Gershman et al.* 2012d], as with SWICS, He^+ is well-resolved using its double coincidence time of flight measurements.

10.4.1 Data Overview

During 2007-2009, MESSENGER passed through the gravitational focusing cone once at $R \approx 0.6$ AU, and at least three times at $R \approx 0.3$ AU. A time series of He^+ measurements from MESSENGER/FIPS and ACE/SWICS during this time period is shown in Figure 10.4. FIPS and SWICS data have been time-averaged in 4-h and daily intervals, respectively. In each pass, a clear enhancement in observed He^+ is apparent near the downwind direction, i.e. $\lambda \approx 75°$. Similar enhancements occur when ACE passes through the cone at 1 AU in late November-early December of 2007 and 2008. The He^+ densities observed by FIPS also changed with MESSENGER's radial variation in its heliocentric orbit. However, enhancements at both $R \approx$ 0.6AU and $R \approx 0.3$AU indicate that increased He^+ flux measured by FIPS is a result of the spacecraft passing through the cone rather than a change in heliocentric distance. Data collected by MESSENGER/FIPS before September 21, 2007 (DOY 264) relied on a version of the instrument flight software that limits our analysis and its comparisons to the data collected later. Therefore, we restrict our detailed analysis of MESSENGER/FIPS data to those collected after September 21, 2007.

10.4.2 Determination of λ_∞ from Pickup Ion Observations

All MESSENGER/FIPS He^+ observations have been scaled to $R = 0.3$ AU by the factor R^α. By combining MESSENGER observations with corresponding data from ACE/SWICS at 1 AU, the longitudinal structure of the gravitational focusing

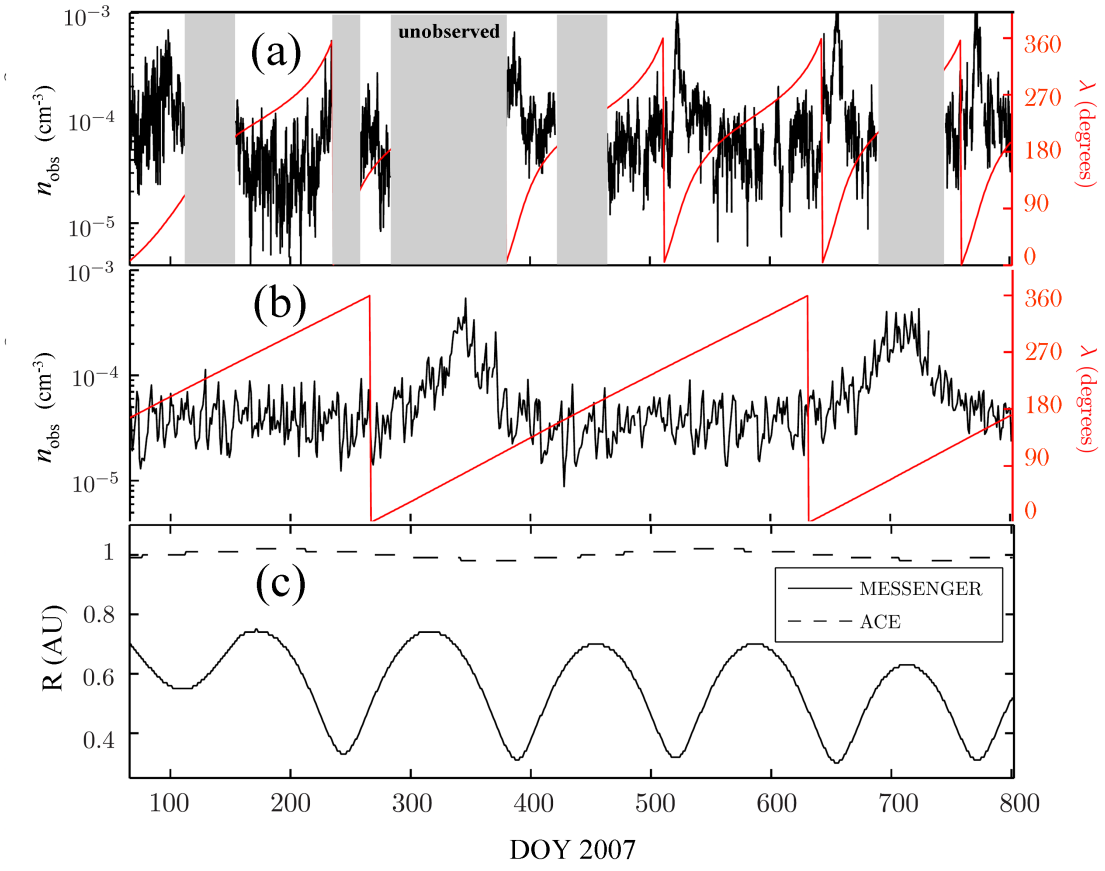

Figure 10.4: Measurements of He^+ during 2007-2009 for (a) MESSENGER/FIPS and (b) ACE/SWICS. In each panel the variation of the ecliptic longitude (λ) with time is also shown as is (c) the radial distance from the Sun of each spacecraft. An enhancement was clearly observed near the predicted downwind direction ($\approx 75°$) at all distances, with MESSENGER having passed through the cone once near $R = 0.6$ AU, and then several times at $R = 0.3$ AU, and ACE passing through the cone yearly at $R = 1$ AU.

cone can be examined and compared with estimates from the models introduced in Section 10.3 to determine the parameters of the LISM that provide the best fit to the data. For this analysis, observed densities were binned in 2.5° intervals with respect to ecliptic longitude λ.

In order to scale MESSENGER/FIPS measurements by a factor R^{α}, the exponent α had to be estimated. Observations of He^{+} in the upwind directions ($165° < \lambda < 345°$) should only vary with heliocentric distance, and therefore provide a basis for determining α. For a given α, σ_{uw}/μ_{uw} represents the relative variability of the scaled upwind measurements, where (σ_{uw}) is the standard deviation of the upwind values and μ_{uw} is their mean value. Because the upwind densities should be nearly independent of λ, the value of σ_{uw}/μ_{uw} that minimizes σ_{uw}/μ_{uw} is selected as the exponent for the radial scaling factor. For the FIPS measurements presented here, $\alpha = 0.57 \pm 0.5$ minimizes the quantity σ_{uw}/μ_{uw}, with the uncertainty defined by the values of α that give σ_{uw}/μ_{uw} values within 10% of the minimum. The large uncertainty in this exponent implies that α may be taken as a free parameter when comparing measurements with neutral density models.

As mentioned above, we used the LISM He gas parameters from *Möbius et al.* (2004), Model A, and from the recent IBEX results (*Möbius et al.* 2012; *Bzowski et al.* 2012), Model B. The velocity and temperature of the LISM will affect the downwind enhancement and cone width. However, these parameters alone do not determine these properties, as the ionization rates in the heliosphere will also affect cone structure. Therefore errors in either of these properties have the potential to reduce the accuracy of any fit.

The two sets of LISM He gas parameters give the same estimate of $\approx -5°$ for the ecliptic latitude of the downwind flow, they differ in their determination of λ_{∞}, the downwind ecliptic longitude. Therefore, we took λ_{∞} as a free parameter for our fitting such that we could examine the mean direction of the recovered cones at $R =$

0.3AU from FIPS and $R = 1$ AU from SWICS. With this procedure, we could also evaluate the presence of any large-scale ion transport or diffusion that would manifest itself as an apparent shift of the cone center.

Both sets of LISM v_∞ , T_∞ parameters were used as input to the neutral model from Section 10.3 with an assumed ecliptic latitude of the downwind direction of $-5°$. Neither spacecraft orbit passed through the true downwind direction; the maximum θ values were $\theta = 172.5°$ and $\theta = 175°$ for MESSENGER and ACE, respectively. In increments of 0.25°, λ_∞ was varied from 65° to 85° and the model from Section 10.3 is used to generate neutral densities at $R = 0.3$ AU and 1 AU, including the effects of electron-impact ionization. These densities were normalized by the average of their upwind values ($\mu_{\rm uw}$) and compared with the spacecraft derived cone structures. For MESSENGER/FIPS, the radial scaling exponent, α, was varied from 0 to 1.25 in increments of 0.025, spanning the uncertainty determined above, with increasing values of α resulting in a smaller observed downwind enhancements.

The Pearson's χ^2 error (*DeGroot and Schervish* 2011) of the fit for $30° < \lambda < 120°$ was computed for each (α, λ_∞) combination for MESSENGER/FIPS and for each λ_∞ value for ACE/SWICS. For MESSENGER/FIPS, the χ^2 error as a function of (α, λ_∞) is shown in Figure 10.5 for both sets of LISM parameters. The χ^2 errors are nearly identical for both sets of parameters; best fits were ($\alpha = 0.60 \pm 0.1, \lambda_\infty = 76° \pm 1.5°$) for Model A, and ($\alpha = 0.525 \pm 0.1, \lambda_\infty = 76° \pm 1.5°$) for Model B with $\chi^2_{\rm min} = 3.7$ and 4.6, respectively. The uncertainties in these values correspond to χ^2 errors less than $1.1\chi^2_{\rm min}$. The best-fit value of α is close to that which minimized the quantity $\sigma_{\rm uw}/\mu_{\rm uw}$, indicating robustness of this recovery. An average $\alpha = 0.5625$ was selected for the scaling of FIPS data. For the ACE/SWICS data, a best fit $\lambda_\infty = 78.5°$ was obtained with $\chi^2_{\rm min} = 4.5$.

Distance-scaled FIPS observations for $\alpha = 0.5625$ as well as SWICS measurements as a function of ecliptic longitude with corresponding best-fit model curves using both

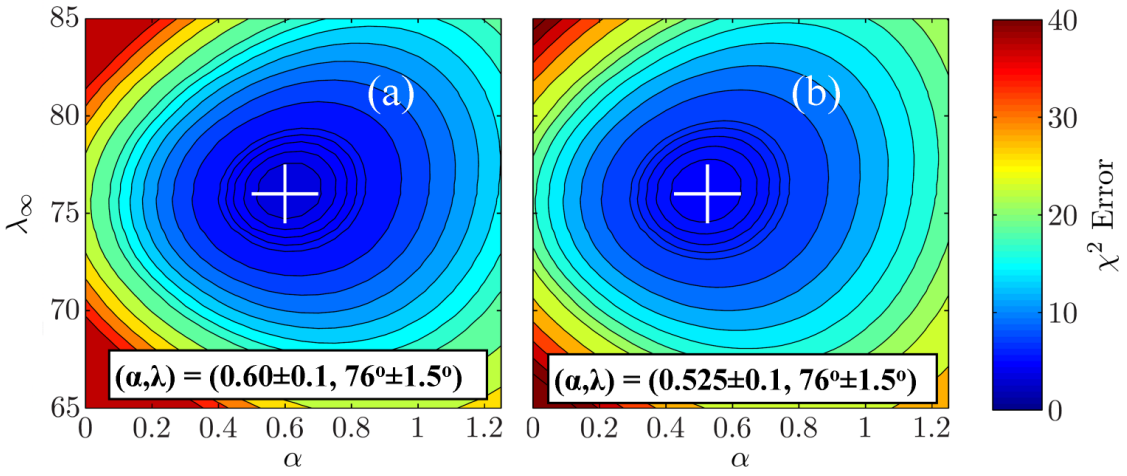

Figure 10.5: Pearson χ^2 error as a function of (α, λ_∞) for MESSENGER/FIPS for (a) Model A (*Möbius et al.* 2004), and (b) Model B (*Möbius et al.* 2012; *Bzowski et al.* 2012). The errors are nearly identical for both sets of parameters, indicating that they produce similar neutral helium distributions. Best fit ($\alpha = 0.60 \pm 0.1, \lambda_\infty = 76° \pm 1.5°$) and ($\alpha = 0.525 \pm 0.1, \lambda_\infty = 76° \pm 1.5°$) were obtained for Models A and B, respectively.

sets of LISM parameters are shown in Figure 10.6. The models at 1 AU provide excellent matches to the SWICS data in terms of both downwind enhancement and cone width, with no significant distinction observable between the two sets of LISM parameters. The scaled FIPS data case provides an optimal fit of the crosswind longitudes, but has a large average downwind enhancement (∼8), consistent with a neutral model that includes almost no electron-impact ionization. However, the smaller enhancement (∼6) predicted by the neutral model that includes electron-impact ionization is still within the uncertainty of the downwind FIPS measurements where there are fewer observations. For reference, an $\alpha \approx 1.1$ scaling would best match the average downwind enhancement of ~ 8 but results in a larger minimum χ^2_{min} error for the crosswind data. Solar wind speed measurements from FIPS (*Gershman et al.* 2012d) indicate that the downwind He^+ observations for at least two of the cone transits accumulated here took place shortly after a high speed solar wind stream.

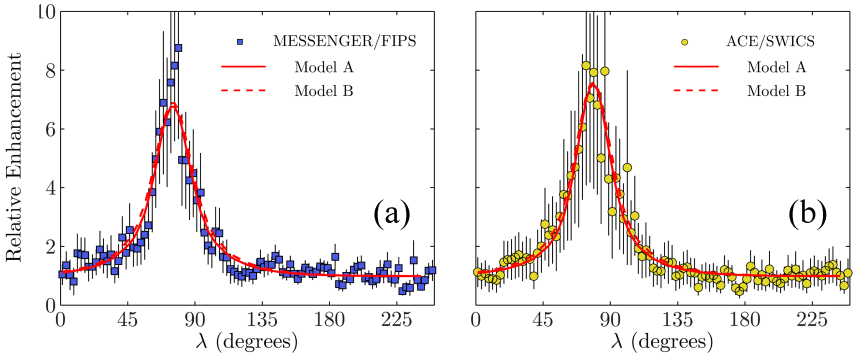

Figure 10.6: Recovered helium gravitational focusing cone structure from MESSENGER/FIPS ($R = 0.3$ AU) and ACE/SWICS (1 AU). Error bars indicate the $1/\sqrt{N}$ counting error where N is the number of observations in each 2.5° interval. Best-fit neutral models using LISM parameters from Model A (*Möbius et al.* 2004) and Model B (*Möbius et al.* 2012; *Bzowski et al.* 2012) are shown for (a) $R = 0.3$ AU using $\alpha = 0.5625$, $\lambda_\infty = 76°$, and (b) $R = 1$ AU, $\lambda_\infty = 78.5°$. Electron-impact ionization effects are included. Similar derived cone centers at $R = 0.3$ AU and $R = 1$ AU indicate that there is no substantial ($> 1 - 2°$) ion transport of diffusion between the two radial distances.

Such solar wind conditions may have led to enhanced measured He^+ densities (*Möbius et al.* 2010) and could explain the systematic underestimation of the neutral model at these points.

10.5 Discussion

From pickup ion data alone it is difficult to tightly constrain the interstellar flow direction, even under the simplest conditions at solar minimum. This difficulty is due to a number of processes taking place that affect neutral helium particles entering the heliosphere and the dynamics and transport of pickup ions. The time dependence of these processes becomes even more complex at solar maximum. However, a relatively wide cone enables a robust fit using many data points. From these fits, independent of the radial scaling of FIPS data, the MESSENGER/FIPS and ACE/SWICS data indicate $\lambda_\infty = 76° \pm 1.5°$ and $78.5° \pm 1.5°$, respectively. These values are closer to the

recent IBEX results than previous determinations from pickup ion data during solar maximum (*Gloeckler et al.* 2004). However, both values are within the errors given by *Möbius et al.* (2004).

Within the uncertainty of the pickup ion observations reported on here, the subtle differences between LISM parameters from *Möbius et al.* (2004) and IBEX results cannot be distinguished. Given the tight constraints on v_∞ and T_∞, errors in the modeled neutral cone structure are primarily a function of the uncertainty in the helium ionization rates. Large variations in the predicted downwind enhancement at $R = 0.3$ AU with different neutral models from Section 10.3 indicate that the cone structure in the inner heliosphere is extremely sensitive to the effects of electron-impact ionization. Conversely, it is difficult to use the pickup ion observations at 1 AU with ACE/SWICS to resolve such effects. Further analysis of MESSENGER/FIPS measured distribution functions may help to constrain these ionization profiles.

The good agreement between the derived λ_∞ between the MESSENGER/FIPS data scaled to $R = 0.3$ AU and the ACE/SWICS data at 1 AU indicates that any large-scale ion transport or diffusion between these two radial distances is constrained to be within $1-2°$. While a cone center offset of $+2°$ is consistent with previous calculations in terms of both magnitude and longitudinal direction (*Möbius et al.* 1995), any offset due to transport should be accompanied by an increase in cone width and reduction in downwind enhancement when compared with the predicted neutral density, and neither effect is observed. These constraints imply that the pitch-angle scattering of He^+ pickup ions does not create any marked longitudinal streaming.

10.6 Concluding Remarks

This most recent solar minimum has provided a unique opportunity for the study of the interstellar helium distribution in the heliosphere. The expected ionization rates of neutral particles were not only at an all-time low, but also reached a near

steady-state. This unusually quiet period of solar activity enabled the averaging of spacecraft data over long time periods, providing a more robust delineation of cone structure than a single downwind pass.

With data averaged over this most recent solar minimum, the structure of the helium gravitational focusing cone has been mapped at multiple heliocentric distances for the first time from in situ measurements of pickup ions. Radial distance-scaled MESSENGER/FIPS observations combined with data from ACE/SWICS during the same time period, led to the delineation of the cone structure at $R = 0.3$ AU and 1 AU. Comparisons with models of the neutral density yielded good matches with established LISM parameters and $\lambda_\infty = 76^\circ \pm 1.5^\circ$ and $78.5^\circ \pm 1.5^\circ$ for $R = 0.3$ AU and 1 AU, respectively. The lack of any significant ($> 1 - 2^\circ$) differences in the cone structure between those solar distances during this time period implies that measurements at 1 AU are a surprisingly sensitive and useful tool for mapping the distribution of neutral helium in the ecliptic, though measurements inside of 0.5 AU may be required to best understand the effects of electron-impact ionization. These fits do not require the presence of any secondary neutral helium population in the heliosphere. Through its nominal and extended mission in orbit around Mercury, MESSENGER will continue to provide observations of He^+ in the inner heliosphere, to be compared with data collected near-Earth.

CHAPTER XI

Conclusions

This work has presented novel methodology and engineering solutions to significantly improve upon spaceborne mass spectrometer instrumentation. While devices can easily be designed in theory to filter particles by their mass-to-charge ratios, the practical implementation of such ideas yields significant design challenges. Not only must these sensors be robust enough to function within the harsh and dynamic space environment, but they must be sensitive and precise enough to make measurements of species spanning multiple orders of magnitude in abundance. In order to enable breakthrough scientific discoveries, the next generation of spaceborne mass spectrometers will require order of magnitude increases in all aspects of their performance, and yet the high risks involved with launching a space instrument demand the use of heritage technology and measurement techniques that have been proven to be successful.

Two approaches to tackle this seemingly paradoxical challenge are to either develop innovative modifications to heritage designs to obtain the required performance, or to invent new analysis methods that reduce the risk of trying new technologies. The former was addressed by the development of a practical scheme in which higher order auxiliary excitation can be applied to an existing quadrupole mass spectrometer sensor. The latter was addressed by the development of a framework for the complete forward modeling and removal of noise events in a time-of-flght mass spec-

trometer using in-flight data, such that even unforeseen issues that arise in-flight as a consequence of a non-heritage design can be mitigated using data post-processing techniques.

11.1 Summary

Higher order auxiliary excitation in conjunction with a multi resonant tank circuit presents a practical way to enhance the mass resolution and peak shapes of quadrupole mass spectrometers. These improvements were demonstrated using both circular rod and hyperbolic rod sensors, indicating that both commercial and spaceborne instruments can benefit from auxiliary excitation. The performance increases measured in the laboratory matched well with those predicted by numerical simulations of ions in a quadrupole field. Consequently, these simulations can be used to inform decisions pertaining to the design of the next-generation of sensors.

Noise events in time-of-flight mass spectrometers can hinder our ability to scientifically analyze measured instrument data. Here, a number of the most common noise sources that plague these sensors were introduced, characterized, and forward modeling, enabling a direct estimation of the signal-to-noise ratio of a given dataset. Further, through the use of kernel density estimation and a Monte Carlo event processing algorithm, measured events were separated into sets of 'real' and 'noise', with less than 25% misidentification expected when combined with accurate noise modeling. These techniques, although applicable to a variety of sensors, were used here to enable scientific analysis of an unprecedented dataset: heavy ion species measured in the inner heliosphere in both the solar wind and in Mercury's magnetosphere by FIPS on MESSENGER.

11.2 Future Work

The QMS technologies developed here have broader impacts that reach well beyond the scope of this dissertation. The commercial applications are evident, where more easily manufactured circular rod sensors can be endowed with near hyperbolic rod-like performance using auxiliary excitation. Furthermore, these techniques enable instrument miniaturization, where the shortcomings of using smaller rods can be overcome with applied excitation. The multi-resonant tank circuit technology introduced here also has applications for ion traps or multipole arrays, where multiple frequency components are required for successful instrument operation. Finally, digital driving of these signals using the presented FPGA-based control system provides a simple and stable solution for the management of their amplitudes and frequencies.

The noise modeling and removal algorithms developed using MESSENGER/FIPS are readily applicable to any such instrument, but also to any sensor that produces discrete sets of events. Although noise sources in TOF-MS can be used to inform the design of future sensors, the true power of such techniques lies in their ability to resurrect older datasets that may have been classified as too noisy to process. For example, instrument measurements taken decades ago can be reanalyzed in the hopes of obtaining previously unavailable scientific analysis.

Continued scientific analysis of MESSENGER/FIPS data will result in an increased understanding of both heliospheric and magnetospheric plasma processes. For the duration of the MESSENGER mission, increased on-orbit data will extend the statistical studies of solar wind alpha particles and heavy ion kinetic properties, as well as result in an increased sampling of interstellar helium pickup ions over the solar cycle. These data will not only provide a set of boundary conditions upstream from Mercury's bow shock, but also will provide the only available set of in situ solar wind measurements in the inner heliosphere for most of the next decade.

The solar wind recovery algorithms presented also may apply to other spacecraft

with obscured field of views, such as other planetary missions that, like MESSENGER, often desire upstream solar wind conditions but lack direction observations of the solar direction. These methods can be further extended to study the effects of obscured observations on the recovery of suprathermal particle distributions with MESSENGER/FIPS, an important feature in the solar wind that remains unmeasured at Mercury.

11.3 Concluding Remarks

The challenges of developing robust and precise space instrumentation will only grow as the desire for further exploration throughout and even beyond the Solar System continues to blossom. It can be both difficult and disheartening to formulate and implement practical solutions for these challenges. For all of the effort, however, the payoff can be immense, whether it be a circuit design that can enable existing quadrupole mass spectrometer rod sets to measure the building blocks of life, or for data processing techniques to take a brand new instrument topology suffering from unique noise sources, and uncover the signature of interstellar helium particles measured closer to the Sun than ever before. Such innovations are vital for the development of the next-generation of spaceborne sensors and for the future of space exploration.

APPENDICES

APPENDIX A

Higher order parametric excitation modes for spaceborne quadrupole mass spectrometers

This appendix is taken from: Gershman, D. J., B. P. Block, M. Rubin, M. Benna, P. R. Mahaffy, and T. H. Zurbuchen (2011), Higher order parametric excitation modes for spaceborne quadrupole mass spectrometers, *Rev. Sci. Instrum., 82*, 125109.

A.1 Multiresonant Tank Circuit Design Considerations

This section provides more detailed technical requirements for tank circuit components as well as an in depth derivation of tank circuit behavior.

A.1.1 Component Values

Consider the multiresonant circuit introduced in Section 2.3.2. The poles of this circuit occur when the denominator of Z_{IN} from equation (2.15) goes to zero such that

$$\omega_{\pm}^2 = \frac{\omega_1^2}{2}\left(1 + \frac{\omega_o^2}{\omega_1^2}\left[1 + \frac{L_o}{L_1}\right] \pm \sqrt{\left(1 + \frac{\omega_o^2}{\omega_1^2}\left[1 + \frac{L_o}{L_1}\right]\right)^2 - \frac{4\omega_o^2}{\omega_1^2}}\right), \tag{A.1}$$

where

$$\omega_+ = \omega_{\text{aux}} \text{ and } \omega_- = \Omega_{\text{RF}}. \tag{A.2}$$

Even with specified frequencies of operation ω_+ and ω_-, there are multiple solutions for the values of the individual components L_0, C_0, L_1, and C_1. Fortunately, the parallel capacitance C_o is bounded from below by the known capacitance of the quadrupole rods. Equations (2.14)-(2.16) can then be manipulated to solve for L_0, L_1, and C_1 in terms of the desired operating frequency Ω_{RF}, the parameters f and g, and the parallel capacitance C_o, yielding,

$$L_o = \frac{(gf + 1 - g)^2}{f^2 \Omega_{\text{RF}}^2 C_o}, \tag{A.3}$$

$$C_1 = \frac{g(g-1)(g - gf - 2)(gf + f + 1 - g)(f-1)^2}{(gf + 1 - g)^4} C_o, \tag{A.4}$$

and

$$L_1 = \frac{(gf + 1 - g)^2}{g(g-1)(g - gf - 2)(gf + f + 1 - g)(f-1)^2 \Omega_{\text{RF}}^2 C_o}. \tag{A.5}$$

A.1.2 Inductor Design Considerations

A non-ideal inductor can be modeled as an inductance with parallel and series resistances (R_s and R_p, respectively) and a self-capacitance (C_s) , as shown in the schematic in Figure A.1. The losses associate with R_p and R_s dominate the determination of Q of the circuit according to equation (A.6), which is defined as the ratio

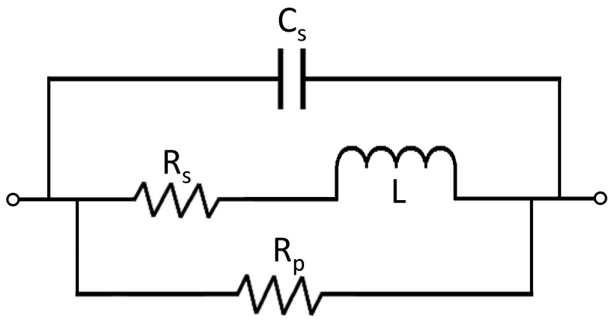

Figure A.1: Lumped element model of a non-ideal inductor with self-capacitance C_{s}, and series and parallel resistances R_{s} and R_{p}, respectively.

of reactance of the inductor to its real resistance,

$$Q \approx \frac{\omega L R_{\mathrm{p}}}{R_{\mathrm{s}} R_{\mathrm{p}} + \omega^2 L^2}. \tag{A.6}$$

A high Q (> 100) is required to create resonant peaks sharp enough to allow for efficient driving of the tank circuit, encouraging the use of a ferromagnetic core for L_0 and L_1. However, due to the high currents in the tank circuit, amplitude permeability and magnetic saturation effects become problematic, changing the operating point of the circuit (i.e., detuning) as a function of magnetic flux inside the cores. To synthesize an appropriate inductor, an air-gap ferrite or powdered iron core can be used that may have a reduced Q with respect to a ferrite toroid, but provide increased temperature and amplitude stability. Litz wire can be used to wind the inductors to minimize eddy current losses. With this technique, an inductor with $L \sim 10^{-4} H$, $R_{\mathrm{p}} \sim 10^{5} \Omega$, $R_{\mathrm{s}} \sim 10^{0} \Omega$, and $C_{\mathrm{s}} \sim 10^{-12} F$ can be readily synthesized.

C_{s}, along with L, defines the self-resonant frequency of the core. The self capacitance of the inductor will also affect the location of the poles of the circuit. When viewed as a lump circuit, the effective inductance of the core appears as

$$L_{\mathrm{eff}} = \frac{L}{|1 - \omega^2 L C_{\mathrm{s}}|}. \tag{A.7}$$

For proper operation, the self-capacitance should be well above the operating frequencies, leading to $L_{\text{eff}} > L$, creating an effective boost in inductance for a given wound core.

A.1.3 Maximum Current Through Components

The driving current through a tank circuit is determined from the input impedance of the synthesized network. At resonance, this driving current is minimal. However, the peak current through each of the individual components of the tank circuit can be larger than the driving current. The currents through each of the components for a given quadrupole voltage are dominated by the ideal reactive impedances, such that for operation near $f = 2$, and $g = 1/2$, the maximum currents through the tank circuit components are

$$I_{\text{max,L}_1,\text{C}_1} \approx \frac{\Omega_{\text{RF}}^2 C_1}{|1 - \Omega_{\text{RF}}^2 L_1 C_1|} V_{\text{max}} \approx 0.77 \Omega_{\text{RF}} C_{\text{o}} V_{\text{max}}, \tag{A.8}$$

$$I_{\text{max,L}_{\text{o}}} \approx \frac{V_{\text{max}}}{\Omega_{\text{RF}} L_{\text{o}}} \approx 1.78 \Omega_{\text{RF}} C_{\text{o}} V_{\text{max}}, \tag{A.9}$$

and

$$I_{\text{max,C}_{\text{o}}} \approx \Omega_{\text{RF}} C_{\text{o}} V_{\text{max}}. \tag{A.10}$$

Fortunately, for $f \approx 2$ operating points, the currents in the new components L_1 and C_1 are of the same order, and in fact smaller than, those of L_0 and C_0. From equations (A.3)-(A.5), the values of L_1 and C_1 are also expected to be smaller than those of L_0 and C_0, respectively. Therefore, similar requirements on the construction of such components should apply.

A.1.4 Relative Impedance At Resonance

Rewriting the input impedance of equation (2.11) in terms of factored pole frequencies, ω_+ and ω_- yields,

$$Z_{\mathrm{IN}}(\omega) = \frac{-j\omega L_{\mathrm{o}}(\omega^2 - \omega_1^2)\omega_{\mathrm{o}}^2}{(\omega^2 - \omega_+^2)(\omega^2 - \omega_-^2)}. \tag{A.11}$$

The relative amplitudes of Z_{IN} at the resonant frequencies depend on the parameters f and g. Assuming Z_{IN} diverges similarly for $\omega \to \omega_+$ and $\omega \to \omega_-$, the ratio of Z_{IN} at the poles becomes,

$$\frac{Z_{\mathrm{IN}}(\omega_-)}{Z_{\mathrm{IN}}(\omega_+)} \sim \frac{\omega_-(\omega_1^2 - \omega_-^2)}{\omega_+(\omega_+^2 - \omega_1^2)} = \frac{(gf + 2 - g)gf^2}{(1-g)(gf - g + 1 + f)}. \tag{A.12}$$

For $f \approx 2$, the parameter g can be written in terms of the desired relative amplitudes of Z_{IN} at the poles such that,

$$g \approx 2\sqrt{\frac{\frac{Z_{\mathrm{IN}}(\omega_-)}{Z_{\mathrm{IN}}(\omega_+)} + 1}{\frac{Z_{\mathrm{IN}}(\omega_-)}{Z_{\mathrm{IN}}(\omega_+)} + 4}} - 1. \tag{A.13}$$

For $Z_{\mathrm{IN}}(\omega_-)/Z_{\mathrm{IN}}(\omega_+) = 1$, i.e., two resonances having similar amplitudes, equation (A.13) gives $g \approx 1/4$.

In a similar manner to equation (A.12), the relative amplitudes of the single-pole tank circuit impedance $Z_{\mathrm{IN,o}}$ and multiresonant tank circuit impedance Z_{IN} can be computed in terms of parameters f and g to yield,

$$\frac{Z_{\mathrm{IN,o}}(\Omega_{\mathrm{RF}})}{Z_{\mathrm{IN}}(\Omega_{\mathrm{RF}})} \sim \frac{(gf + 1 - g)^2(f+1)}{(gf + 2 - g)gf^2}. \tag{A.14}$$

Note that equations (A.12) and (A.14) are only approximations since they do not include the effects of finite Q inductors on input impedance. These equations do, however, to first order, illustrate the relative behavior of these circuits with changes in f and g.

A.1.5 Driving Current Requirements

For $f \approx 2$ there can be a significant reduction in the driving point impedance at the fundamental RF resonant frequency as a consequence of adding new poles and zeros into the impedance network. This reduction directly translates into the current required to drive the rods with a particular voltage. For higher frequency separation between the two poles (i.e., $f > 2$), this reduction is eliminated, but the amount of excitation needed to produce island a given island splitting significantly increases.

In order to mitigate this effect, the circuit zero can be shifted toward higher frequencies at the expense of the auxiliary signal resonant peak amplitude. Note that when adding a Foster-derived network to an existing resonant circuit, the location of the fundamental RF frequency shifts to lower frequencies, so in comparing driving current, one must compare the driving point impedance of a circuit whose operating frequencies are the same i.e., a circuit with a Foster-derived network must be compared to a single-pole circuit with additional parallel capacitance sufficient to push the pole to a lower frequency.

The reduced driving point impedance at the fundamental RF frequency can be balanced by the tremendous increase in driving point impedance at the auxiliary frequency. For example, for a single-pole high-Q circuit, the auxiliary frequency is relatively far from the resonant peak, resulting in a driving point impedance reduction by over 90%. For a high-Q multiresonant tank circuit, the driving point impedance at both the fundamental and auxiliary frequencies are comparable in magnitude.

The driving point impedance of the tank circuit dictates the driving current requirements to produce a particular potential difference across the quadrupole rods. For a traditional quadrupole operating at a given frequency $\Omega_{\mathrm{RF}}^2 = 1/[L_0 C_0]$, the current required, I_{o}, is inversely proportional to Z_{Dro}, the driving point impedance of a single-pole tank circuit. The driving point impedance is close to, but not exactly the same as Z_{IN}. The driving circuitry sees the tank circuit impedance reflected across

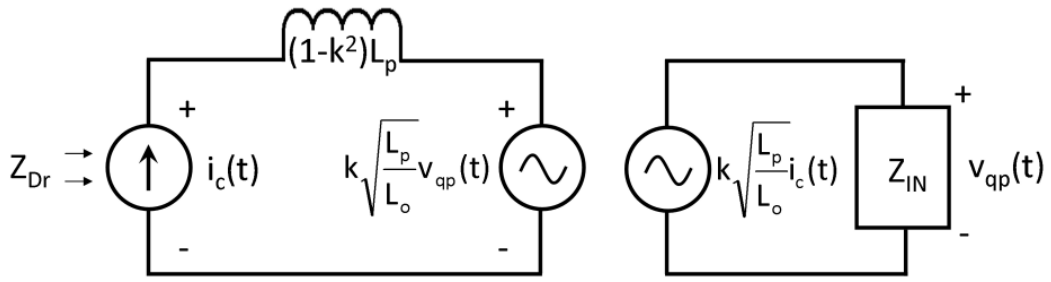

Figure A.2: Model of an inductor with coupling coefficient k, primary inductance L_p, and secondary inductance L_o. Z_Dr is the measured driving point impedance of the circuit. Z_IN is the impedance of the tank circuit network, including the secondary inductance L_o, quadrupole rod capacitance C_o, and any Foster-derived impedance networks.

the primary inductance, L_p, with coupling coefficient k. Using the circuit model of a transformer shown in Figure A.2, the driving point impedance is,

$$Z_\mathrm{Dr} = j\omega(1-k^2)L_\mathrm{p} + k^2 L_\mathrm{p} Z_\mathrm{IN}. \tag{A.15}$$

Note that the locations of the poles of Z_Dr will be identical to those of Z_IN, but the location of the zeros will be shifted based on the coupling coefficient k and the inductance L_p. Near resonances, the second term in equation (A.15) dominates, so $Z_\mathrm{Dr} \propto Z_\mathrm{IN}$, enabling the use of the ratios from equations (A.12) and (A.14) when calculating I. When driving the quadrupole at two frequencies with excitation strength q', the new driving current I, normalized by I_o, can be written as

$$\left(\frac{I}{I_\mathrm{o}}\right) = \frac{Z_\mathrm{Dro}(\Omega_\mathrm{RF})}{Z_\mathrm{Dr}(\Omega_\mathrm{RF})}\left[1 + q' \times \frac{Z_\mathrm{Dr}(\omega_-)}{Z_\mathrm{Dr}(\omega_+)}\right] \approx \frac{Z_\mathrm{INo}(\Omega_\mathrm{RF})}{Z_\mathrm{IN}(\Omega_\mathrm{RF})}\left[1 + q' \times \frac{Z_\mathrm{IN}(\omega_-)}{Z_\mathrm{IN}(\omega_+)}\right]. \tag{A.16}$$

The relative current I/I_o becomes a function of the pole-zero spacing parameters f and g, and the excitation strength q'. For $f = 2$, Figure A.3 illustrates how $g$can be selected in order to minimize I/I_o for a given value of q'. For $q' \approx 0.1$, $g \approx 1/2$ is

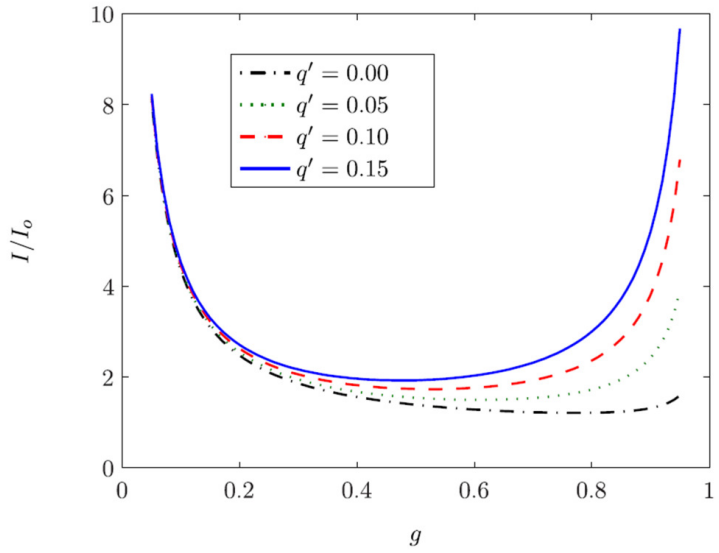

Figure A.3: Calculated relative driving current as a function of modulation strength q' and pole-zero spacing parameter g for $f = 2$. As the excitation strength q' is increased, the value of g for which the driving current is a minimum decreases.

obtained.

For comparison, the transfer function $H_o(s)$ of a single-pole high-Q tank circuit (adapted from the general transfer function of a 2nd order bandpass filter*Shenoi* (2005) varies as

$$H_o(s) = \frac{H_o s \Delta\omega}{s^2 + s\Delta\omega + \Omega_{RF}^2}, \tag{A.17}$$

where $s \equiv j\omega$ and H_o is the maximum amplitude of the transfer function. Taking the magnitude of equation (A.17) and using the definition of Q from equation (A.16) yields the relative input impedance of a single-pole tank circuit at Ω_{RF} and ω_{aux},

$$\frac{|H_o(s = j\omega_{aux})|}{|H_o(s = j\Omega_{RF})|} = \frac{Z_{IN,o}(\omega_{aux})}{Z_{IN,o}(\Omega_{RF})} = \frac{1}{\sqrt{1 + Q^2(1 - \frac{1}{f^2})^2}}. \tag{A.18}$$

Equations (A.16) and (A.18) can then be combined to give the relative driving current required for a single-pole tank circuit with excitation strength q',

$$\left(\frac{I}{I_{\mathrm{o}}}\right) = 1 + q'\sqrt{1 + Q^2(1 - \frac{1}{f^2})^2}. \tag{A.19}$$

A.1.6 Tank Circuit Implementation and Tuning

Tuning a physically realized multiresonant tank circuit can be nontrivial due to stray reactance in the circuit. Even if circuit components are synthesized with precise values from equations (A.3)-(A.5), the tank circuit will not be properly tuned. Stray inductances and capacitances in the circuit will slightly shift pole and zero locations. It is therefore important to measure the operating point of the circuit in such a way that the act of measuring the locations of the resonances does not detune the network. The developed system interface already has a feedback mechanism that allows the FPGA to measure the voltage across the quadrupole rods. By sweeping the frequency applied to the rods and using this feedback signal, the locations of the poles can be measured and adjusted in the network's flight configuration.

A transformer was constructed and integrated into the impedance network, including the quadrupole rods, creating a single-pole resonant circuit. By adding known amounts of parallel capacitance to C_0 and measuring the location of the pole frequency (Ω_{o}), the effective inductance $L_{\mathrm{o,eff}}$ can be determined, as well as the total C_0. Following equation (2.10), there should be a linear relationship between $1/\Omega_{\mathrm{o}}^2$ and $C_{0,\mathrm{known}}$, such that

$$\frac{1}{\Omega_{\mathrm{o}}^2} = L_{\mathrm{o,eff}}(C_{\mathrm{o,known}} + C_{\mathrm{o,unknown}}). \tag{A.20}$$

This relationship is shown to hold in Figure A.4 for a measured network. Originally, an operating point of f=1.9167, g = 0.5, and $\Omega_{\mathrm{RF}} = 2\pi\times$1MHz was desired for 80-pF rods. Following equations (A.3)-(A.5) (and allowing for some additional parallel capacitance), a transformer with secondary inductance of $153\mu H$ was created. However, from Figure A.4 , it is clear that the effective inductance of the transformer

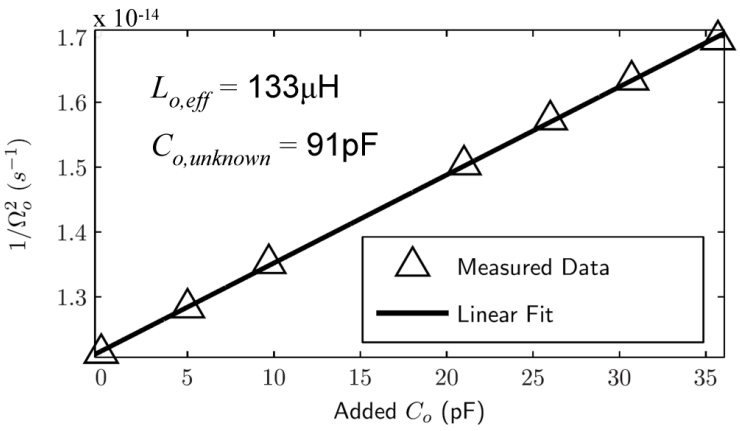

Figure A.4: $1/\Omega_o^2$ vs. added parallel capacitance C_o for a measured single-pole tank circuit.

Table A.1: Measured properties of multiresonant tank circuit inductors.

	L_0	L_1, a	L_1, b
Core Type	PI-157-7	PI-157-10	PI-157-10
Number of windings	53-53	87	87
Wire	40-44 Litz	40-44 Litz	40-44 Litz
L	153uH	136	135uH
C_p	4.8pF	1.1pF	1.1pF
R_p	$874k\Omega$	$330k\Omega$	$276k\Omega$
R_s	1.4Ω	1.2Ω	1.5Ω
Q (1MHz)	408	241	203

in the full tank circuit is about 20% less than that of the measured secondary at $153\mu H$. Instead of rewinding the secondary coil to add more turns, this transformer can be used to achieve the desired operating point, but at a slightly higher frequency.

The L_1 inductors were constructed close to the desired L_1/L_0 ratio, taking into account the measured value of $L_{o,eff}$. L_1, a and L_1, b were constructed with values of $135\mu H$ and $136\mu H$ respectively. The fully measured characteristics of each inductor can be found in Table A.1.

After integrating the L_1 inductors into the tank circuit, the capacitance C_1 can be adjusted to sweep through operating points. The measured locations of the poles can be readily used to compute the value of f. However, the determination of parameter g is not as trivial. The location of the measured Z_{Dr} zeros will be shifted from

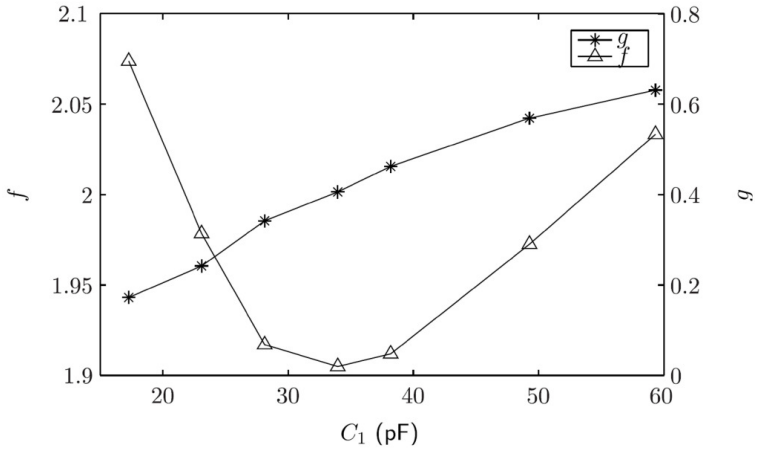

Figure A.5: Operating parameters f and g vs. Foster-derived network capacitance for a multiresonant tank circuit.

that of Z_{IN} due to a non-unity coupling coefficient between the transformer primary and secondary windings. Furthermore, by measuring the circuit response through a voltage feedback signal, the precise location of the circuit zero is difficult to measure directly. However, the measurement of the change in pole locations is sufficient to recover the operating parameter g. Consider the term $\omega_1^2\omega_{\mathrm{o}}^2 \equiv 1/[L_{\mathrm{o}}C_{\mathrm{o}}L_1C_1]$. Using equations (A.3)-(A.5), rearranging this product and solving for the zero frequency yields:

$$\omega_1 = \frac{\omega_{\mathrm{aux}}\Omega_{\mathrm{RF}}}{\omega_{\mathrm{o}}}, \tag{A.21}$$

where ω_{o} is measured during the first part of the tuning process before the Foster-derived network was added to the tank circuit, and ω_{aux} and Ω_{RF} are the measured pole locations with changing C_1. Figure A.5 shows measured parameters f and g for the implemented tank circuit as a function of a known added C_1. Note that the total C_1capacitance and effective L_1 inductance can be recovered in a similar fashion to $L_{\mathrm{o,eff}}$ and total C_{o}.

For a total C_0 of approximately 100pF and a total C_1 value of approximately 38pF,

a desired operating point is achieved with frequencies of 1.04 MHz and 1.99 MHz, a frequency spacing parameter of $f \approx 1.9167$, and a zero location parameter of $g \approx 0.5$. As demonstrated, while it can be difficult to precisely match all component values, it is certainly possible to achieve an acceptable operating point. Once tuned, small, slow drifts in operating point over time can be tolerated as long as the location of the resonant peaks are known and the strength of the auxiliary signal can be adjusted accordingly.

A.1.7 Driving Circuitry Considerations

A simple and efficient method of driving the tank circuit is a voltage-controlled current source, for example a bipolar transistor biased to operate as a class C-like amplifier. The current waveform is shown in Figure A.6 , along with its corresponding frequency spectrum. This rectification, as well as any other distortion in the signal (from D/A quantization, for example), creates harmonics of the fundamental RF frequency. The high-Q tank circuit normally filters out these harmonics to create a pure, clean signal across the rods. However, with an additional pole near $2\Omega_{\mathrm{RF}}$, there is a risk that some of this distortion will appear across the transformer secondary. Therefore, care should be taken in selection of a system operating point to avoid significant signal distortion at harmonics of the fundamental frequency affecting the stability island formation and consequently, quadrupole performance.

A.2 FPGA Design Considerations

This section provides more detailed information about FPGA design and algorithm development.

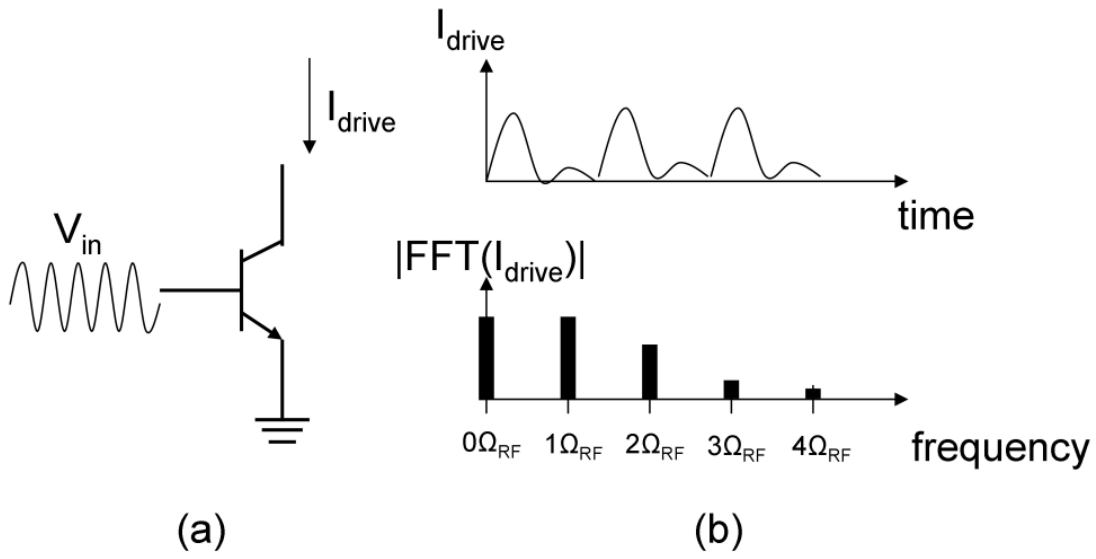

Figure A.6: (a) Sample schematic of voltage-controlled current source biased for class-C-like operation. (b) Clipped current waveform at the output of the amplifier with corresponding frequency spectrum.

A.2.1 Synthesis of Frequency Components

An FPGA has a single frequency reference clock, typically on the order of 100 MHz. Direct digital synthesis (DDS) is a method of creating an arbitrary waveform from this clock, as opposed to the utilization of one or more crystal oscillators or tank circuits. For a given frequency, each clock cycle corresponds to an accumulation of phase. At each cycle, a digital word is generated from a sine wave phase-amplitude look-up table (*Vankka and Halonen* 2001). The necessary frequencies for quadrupole applications are on the order of a few MHz, providing at least a few dozen samples per cycle. A DDS block for each frequency component can be used and the output digital words can be scaled, combined, and output to a digital-to-analog converter (DAC) .

A.2.2 Detection of Frequency Components

The auxiliary waveform serves to modulate the envelope of the fundamental RF signal. The frequency of this modulation can be written in terms of the number of corresponding RF cycles according to equation (A.22) (for $1.5 < f < 2$),

$$\#\text{ RF cycles} = \frac{1}{2-f}. \tag{A.22}$$

For example, for $f = 1.9167$, 12 RF cycles are required to modulate the envelope of the signal one period (i.e., $\approx 0.012ms$ for a $1MHz$ fundamental frequency). There are a number of different methods available that can be used to sense the amplitudes of driving frequency components, in both analog and digital realms. For example, Fast Fourier transforms (FFT) or general filter-based amplitude demodulation techniques can be applied. However, to maximize detection speed and to take advantage of the known operating frequencies and low-modulation indices, a simple envelope detection technique has been implemented.

For $\frac{1}{2-f} \gg 1$, the envelope of the quadrupole signal changes slowly with respect to subsequent zero crossings. For relatively low levels of modulation ($q' \sim 0.1$), the time between zero crossings are relatively unchanged from the unmodulated signal. Let t_{n} and $t_{\mathrm{n+1}}$ be the times of two adjacent zero crossings of the measured signal. Integrating over the absolute value of the waveform between these two time steps should yield a number proportional to the signal envelope, i.e.,

$$\int_{t_{\mathrm{n}}}^{t_{\mathrm{n+1}}} |Vcos(\Omega_{\mathrm{RF}}t) + q'Vcos(\omega_{\mathrm{aux}}t)|dt \approx \frac{2}{\pi}(V + q'Vcos(\Omega_{\mathrm{RF}}(2-f)\frac{t_{\mathrm{n}} + t_{\mathrm{n+1}}}{2}))\Delta t_{\mathrm{n,n+1}}, \tag{A.23}$$

where $\Delta t_{\mathrm{n,n+1}} \equiv t_{\mathrm{n+1}} - t_{\mathrm{n}}$.

The integral in equation (A.23) generates one number every FPGA clock cycles.

Values corresponding to intermediate clock cycles can be computed through linear interpolation, which serves to reconstruct the envelope signal, $V+q'Vcos(\Omega_{\mathrm{RF}}(2-f)t)$ for all cycles. Integrating this signal over a full envelope period, accumulating positive and negative signal sectors separately, gives a number proportional to V. Integrating the derivative of the envelope signal gives a number proportional to $q'V$. These proportionality constants will be frequency dependent. However, for fixed frequency operation, these constants can be determined through calibration, resulting in the measured amplitudes of the fundamental RF frequency (V) and the auxiliary signal ($q'V$).

This envelope detection algorithm can be implemented using only addition and fixed point multiplication operations, ideal for FPGA applications. No phase-locked loops or additional digital filters are needed, only a zero-crossing detector. For 10 measured periods of the envelope signal and $f = 1.9167$, the total data acquisition time would be approximately $0.12ms$.

A.2.3 PID Control Loops

A proportional-integral-derivative (PID) controller is a closed-loop feedback control scheme (*Bennett* 1993). An error signal between the measured (through a feedback signal) and desired rod voltage is calculated. The error signal (e(t)) is scaled (P), integrated (I), differentiated (D), and then applied to the rods as the amplitude V(t) in equation (A.24). Over several time steps, if the controller is properly tuned (i.e. appropriate coefficients K_{p}, K_{i}, and K_{d} are selected), the measured set-point approaches the desired one,

$$V(t) = K_{\mathrm{p}}e(t) + K_{\mathrm{i}}\int_0^t e(\tau)d\tau + K_{\mathrm{d}}\frac{d}{dt}e(t). \tag{A.24}$$

The total settling time of the closed loop will depend on several factors. First, a

sufficient sample of the feedback signal must be digitally acquired and the amplitudes of frequency components must be determined. The PID control outputs are updated with the latest error signal and a new output waveform is applied to the rods. A characteristic settling time will be required before the signal should be resampled. That settling time will be controlled by the bandwidth of the tank circuit (Q $\approx$ 200 @ 1MHz) such that

$$\Delta t_{\text{settle}} = \frac{2\pi Q}{\Omega_{\text{RF}}} \approx 0.2ms. \tag{A.25}$$

Combined with an acquisition time of $0.12ms$, the total open loop settling time will be less than $1ms$. The closed loop settling time will be determined by the number of steps the controller takes to reach its nominal value. For this simple application, the basic PID loop settling time was tuned to about 20 steps. However, more sophisticated feedback techniques, such as adding a feed-forward component to the existing PID controller, could potentially serve to significantly reduce the required settling time. Since the amplitude of the excitation signal is proportional to that of the fundamental RF waveform, the auxiliary PID loop will necessarily converge after the settling of the main PID loop.

APPENDIX B

Noise characterization and design considerations of in space-based time-of-flight sensors

This chapter is taken from: Gilbert, J. A., D. J. Gershman, G. Gloeckler, R. A. Lundgren, T. M. Orlando, J. McLain, and T. H. Zurbuchen (2012), Background noise in space-based time-of-flight sensors - Part I: Characterization and design consideration, *Space Sci. Rev.*, in prep. In addition to assisting with the development and revision of the manuscript text, I was responsible for the electron stimulated desorption contribution to the paper, including the laboratory testing and SIMION simulations.

B.1 ESD SIMION Simulation

A three-dimensional ion optics model of the FIPS TOF chamber was created in SIMION (*Dahl* 2000) and used to simulate the trajectories of ions desorbed by electrons emitted from the wire grids. The voltage potentials in the model were set to the same values used in flight, and the effective potentials of the field control grids

in front of the start and stop MCP assemblies were modeled according to work by *Read et al.* (1998) on field penetration through wire grids. Ions were initialized at rest uniformly over the front surface of the start MCP. These ions were quickly accelerated up to ∼5keV/e due to electric fields in the instrument that are dependent on the value of the post-acceleration voltage. When they struck the top of the TOF housing, a secondary electron was emitted. The instrument-measured TOF corresponds to the time for an ion to travel from the front of the start MCP to the top of the TOF housing, and the time for the secondary electron to be measured by the stop MCP assembly.

The mass spectra from the lab measurements of Figure 6.5 were discretized and used to determine the relative ratios of desorbed ion masses, with the exception of H^+, whose abundance was scaled down by a factor of 30 in order to enable better comparison with the peak shapes of other simulated masses. The true relative desorbed mass spectra are expected to vary with electron energy. For example, a reduction in the H^+ peak at such comparatively high incident electron energies to those of the laboratory tests is not unreasonable. Higher energy electrons will deposit increasing amounts of energy deeper into the material. Desorbed H^+ ions must therefore diffuse to the surface without getting neutralized, a process that becomes increasingly difficult with increasing depth in the material.

Figure B.1 compares the desorbed-ion simulations with FIPS in-flight measurements made at two different post-acceleration voltages. Note only are the the simulated times of flight in good agreement with those of the in-flight data in terms of the TOF value, peak shape, and relative counts, but the spectra shifts, as expected, with changing post-acceleration voltage. Errors in the field penetration model will contribute to a constant offset error in TOF value as the time for a secondary electron to reach the stop MCP should be independent of the mass of its corresponding desorbed ion. The overall lower flight times of the simulation may also be due to an

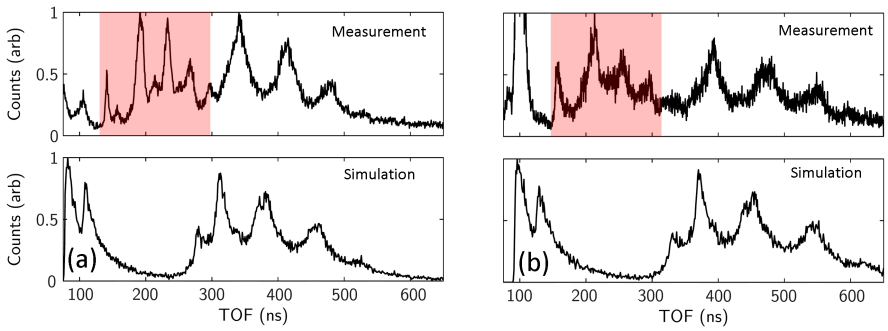

Figure B.1: Comparison of FIPS flight data and SIMION simulations of desorbed hydrocarbons from the start MCP (bottom row) for post-acceleration voltages of (a) -10.5 kV and (b) -8.5 kV. The simulated TOF values of ions that travel from the start MCP to the top of the TOF chamber match closely with peaks from the flight data. The shaded region in the flight data contains peaks from ions that follow trajectories that collide with the side wall of the TOF chamber instead of the top, creating a set of peaks at a constant fraction of the higher-TOF distribution.

underestimation of the flight path of the ESD ions by a few mm, a TOF error that scales in proportion to the square root of the ion mass. The shaded region in Figure B.1 indicates ESD noise events caused by electron emission onto a local area near the edge of the MCP, as discussed previously. These events were not simulated due to the difficulties in accurately modeling fringing field effects.

APPENDIX C

Noise forward modeling and removal through post-processing in space-based time-of-flight sensors

This appendix is taken from: Gershman, D. J., J. A. Gilbert, J. M. Raines, G. Gloeckler, S. C. Solomon, and T. H. Zurbuchen (2012), Background noise in space-based time-of-flight sensors - Part II: Noise event forward modeling and removal through post-processing, *Space Sci. Rev.*, in prep.

C.1 Event processing errors

Before being applied to in-flight data, the performance of the removal process must be evaluated using synthetic datasets. A direct calculation of the change in SNR after processing data is not necessarily an adequate metric in determining the accuracy of the noise removal. In fact, the results of such a computation may be misleading as, in addition to removing noise-based events, an algorithm may remove some real

events, biasing measurements of SNR. Therefore, the relative loss of signal, $|\Delta S/S|$, and relative reduction in noise, $|\Delta\eta/\eta|$, should be analyzed individually. Note that the total number of recovered real events will be set by the total amount of modeled noise, so errors in processing will likely be a misclassification of noise and real events rather than removal of the incorrect number of events. Therefore, it is expected that $|S| \approx \eta - |\Delta\eta|$.

Steps that contribute to data loss are those whose total remaining signal ($f^*_{\mathrm{m,real}}$) is less than that of the distribution of true real events ($f_{\mathrm{m,real}}$). Accumulating these differences and dividing by the total number of real events yields the fraction of events lost due to the removal process, $|\Delta S/S|$, as shown in equation (C.1),

$$\left|\frac{\Delta S}{S}\right| = \sum_{\mathrm{i}} \begin{cases} \dfrac{f_{\mathrm{m,real}}(x_{\mathrm{i}}) - f^*_{\mathrm{m,real}}(x_{\mathrm{i}})}{\sum_{\mathrm{j}} f_{\mathrm{m,real}}(x_{\mathrm{j}})}, f_{\mathrm{m,real}}(x_{\mathrm{i}}) \geq f^*_{\mathrm{m,real}}(x_{\mathrm{i}}) \\ 0, f_{\mathrm{m,real}}(x_{\mathrm{i}}) < f^*_{\mathrm{m,real}}(x_{\mathrm{i}}) \end{cases} . \tag{C.1}$$

On the contrary, steps that overestimate the real signal, i.e. $f^*_{\mathrm{m,real}} > f_{\mathrm{m,real}}$, can be used according to equation (C.2) to calculate the fraction of the noise removed, $|\Delta\eta/\eta|$,

$$\left|\frac{\Delta\eta}{\eta}\right| = 1 - \sum_{\mathrm{i}} \begin{cases} \dfrac{f^*_{\mathrm{m,real}}(x_{\mathrm{i}}) - f_{\mathrm{m,real}}(x_{\mathrm{i}})}{\sum_{\mathrm{j}} f_{\mathrm{m,noise}}(x_{\mathrm{j}})}, f^*_{\mathrm{m,real}}(x_{\mathrm{i}}) \geq f_{\mathrm{m,real}}(x_{\mathrm{i}}) \\ 0, f^*_{\mathrm{m,real}}(x_{\mathrm{i}}) < f_{\mathrm{m,real}}(x_{\mathrm{i}}) \end{cases} . \tag{C.2}$$

If $f_{\mathrm{m,real}}$ and $f_{\mathrm{m,noise}}$ are known, equations (C.1) and (C.2) can be used to evaluate the accuracy of a noise removal process. Note that these quantities are not directly calculable from data and can only be used in the analysis of synthetic datasets. Ideally, $|\Delta S/S| \to 0$ and $|\Delta\eta/\eta| \to 1$, i.e. all of the noise is removed without eliminating any of the true signal.

The overall misidentification of events ($|\Delta f/f|$) can be calculated by adding the number of removed signal events (ΔS) with the number of remaining noise events, ($\eta - \Delta\eta$), and normalizing by the total number of events, N,

$$\left|\frac{\Delta f}{f}\right| = \frac{S + (\eta - \Delta\eta)}{N}. \tag{C.3}$$

Using the known relationships between S, η, and SNR_{m},

$$\begin{aligned} S &= N \cdot \frac{SNR_{\mathrm{m}}}{1 + SNR_{\mathrm{m}}} \text{ and} \\ \eta &= N \cdot \frac{1}{1 + SNR_{\mathrm{m}}}, \end{aligned} \tag{C.4}$$

and combining them with equations (C.1) and (C.2), equation (C.3) can be written as,

$$\left|\frac{\Delta f}{f}\right| = \frac{1}{1 + SNR_{\mathrm{m}}}\left(1 - \left|\frac{\Delta\eta}{\eta}\right|\right) + \frac{SNR_{\mathrm{m}}}{1 + SNR_{\mathrm{m}}}\left(\left|\frac{\Delta S}{S}\right|\right). \tag{C.6}$$

As expected, for the ideal recovery where $|\Delta S/S| \to 0$ and $|\Delta\eta/\eta| \to 1$, the misclassification of events $|\Delta f/f| \to 0$.

C.1.1 Limitations of Monte Carlo event processing

As discussed in Section 7.4.2, P_{noise} is used instead of $P_{\mathrm{m,noise}}$ as part of an iterative Monte Carlo event processing technique. The error introduced by this approximation can be quantified through Monte Carlo processing of two synthetic datasets. Two test cases are examined: (1) real events uniformly distributed in time, and (2) real events all occurring at a single time step. Noise sampled from the distribution shown

in Figure 7.2 is superimposed on each signal.

In the case of a perfectly localized signal, $P_{\text{noise}} = P_{\text{m,noise}}$. Therefore, even for small N, a Monte Carlo processing technique will separate $f_{\text{m,real}}$ and $f_{\text{m,noise}}$ with almost 100% accuracy, with $|\Delta S/S| \approx 0$ and $|\Delta\eta/\eta| \approx 1$. For a uniformly distributed signal, however, $P_{\text{noise}} \neq P_{\text{m,noise}}$, especially for small N, resulting in errors in the noise removal process that are dependent on both N and SNR_{m}. For datasets with various SNR_{m}, $|\Delta S/S|$, $|\Delta\eta/\eta|$, and $|\Delta f/f|$ are computed for a varying number of noise-based events η = 10, 100, 1000, and 10,000 and are shown in Figure C.1.

Both fewer real and noise events are removed with increasing SNR, with approximately a 20% removal of real events and 80% removal of noise events at SNR $\approx$ 1 in Figures C.1a and C.1b. The overall event identification error, $|\Delta f/f|$, in Figure C.1c, however, is less than 20% for all SNR, peaking near SNR $\approx$ 1. Event misidentification is a consequence of using the large N limit noise probability P_{noise} instead of $P_{\text{m,noise}}$. These errors will be greatest for non-localized event distributions. As expected, improved results are obtained for large N, where $P_{\text{m,noise}} \rightarrow P_{\text{noise}}$. However, despite these misidentifications, it is important to remember that the algorithm will necessarily remove the correct amount of noise, i.e., the total amount of recovered signal will be accurate to within the errors of f_{noise}.

C.2 Event processing errors from KDE

With appropriately defined metrics $|\Delta S/S|$,$|\Delta\eta/\eta|$, and $|\Delta f/f|$, the performance of a KDE-implemented noise removal algorithm can be analyzed using synthetic data as a function of the total number of events and dataset SNR. The two test cases used in Appendix C.1. are re-examined here. Instead of a known P_{noise}, however, an approximate P_{noise} is obtained through KDE estimation of $f(x)$. $|\Delta S/S|$, $|\Delta\eta/\eta|$, and $|\Delta f/f|$ are shown as a function of SNR for both the uniform real event, and localized real event distributions in Figure C.2.

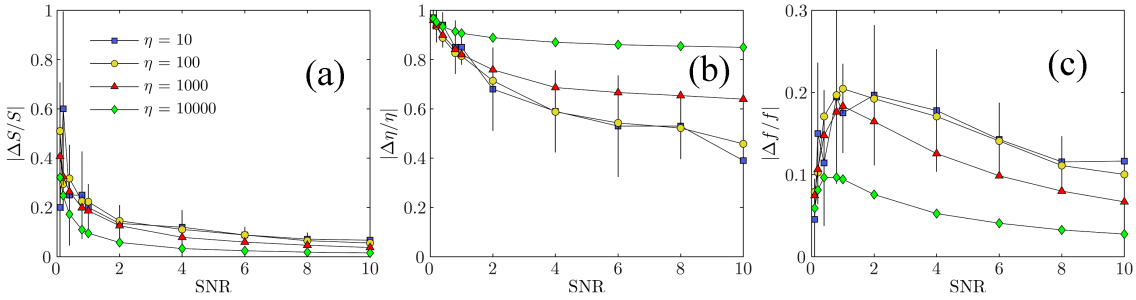

Figure C.1: (a) $|\Delta S/S|$, (b) $|\Delta\eta/\eta|$, and (c) $|\Delta f/f|$ as a function of SNR of a uniformly distributed real event signal for various numbers of noise based events η. An idealized Monte Carlo event processing algorithm was used with P_{noise} computed using the noise distribution from Figure 7.2. Each data point represents the average of 10 independent removals of the same dataset with the error bars indicating the standard deviation. The overall misidentification of events, $|\Delta f/f|$, is less than 20% for all SNR.

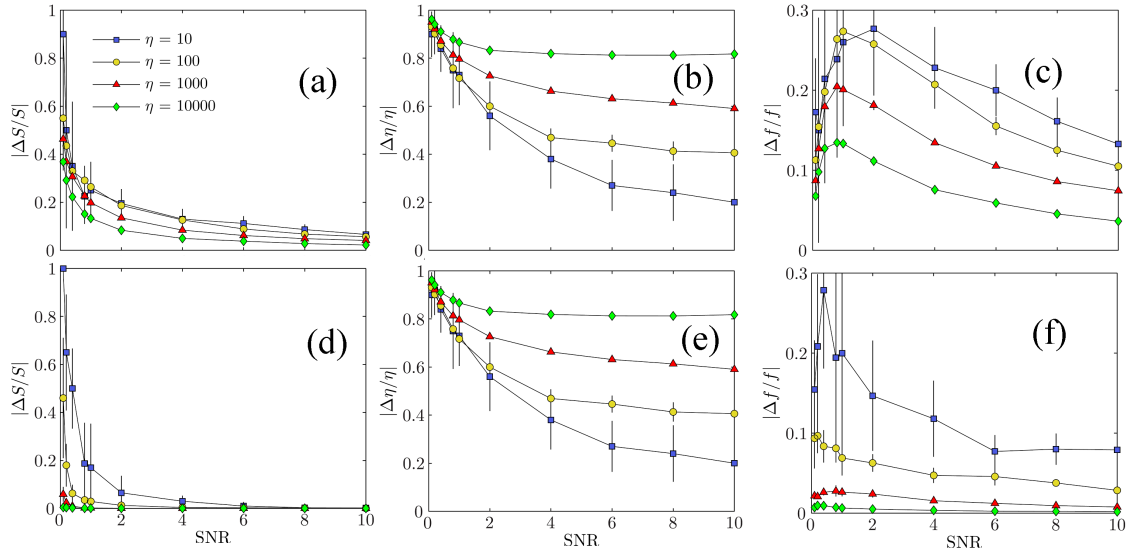

Figure C.2: (a), (b), (c) same as Figure 7.3 but using fm,KDE instead of known f for a uniform distribution of real events. (d),(e),(f) same as parts (a),(b), and (c) but for a localized event source at time step $t = 2800$. With KDE techniques, the overall misidentification error is expected to be less than 25%.

For the uniform real event distribution, the $|\Delta S/S|$,$|\Delta\eta/\eta|$, and $|\Delta f/f|$ values here are similar to the ideal Monte Carlo process results from Figure C.1, with slightly more real events being misidentified at SNR = 1 ($\sim$25%) and slightly less noise events removed at higher SNR values ($\sim$20%). The localized events also show some increased error when compared to the idealized case. Also, as expected, increased number of events N leads to more accurate removal. The KDE technique therefore appears to provide suitable estimates of $f(x)$, such that the end result of the Monte Carlo processing algorithm is similar to that of the ideal cases in all SNR regimes. From these tests, one can expect that for an accurate f_{noise}, the total number of expected noise events will be removed with less than 25% of the total events (noise and real) being misidentified for both localized and more uniform distributions of real events.

APPENDIX D

Solar wind alpha particles and heavy ions in the inner heliosphere

This chapter is taken from: Gershman, D. J., T. H. Zurbuchen, L. A. Fisk, J. A. Gilbert, J. M. Raines, B. J. Anderson, C. W. Smith, H. Korth, and S. C. Solomon (2012), Solar wind alpha particles and heavy ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

D.1 Moments of partial velocity distribution functions

To understand the effects of observing only a fraction of the solar wind plasma, we take the moments of the obstructed drifting isotropic Maxwell-Boltzmann velocity distribution defined by

$$f(v_x, v_y, v_z) = n_o \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{v_{th}^3} exp\left(-\frac{\left((v_x - v_o)^2 + v_y^2 + v_z^2\right)}{2v_{th}^2}\right). \quad \text{(D.1)}$$

Here, n_o is the proton density, v_o is the bulk speed, and v_{th} is the thermal speed that is related to the proton temperature by the definition

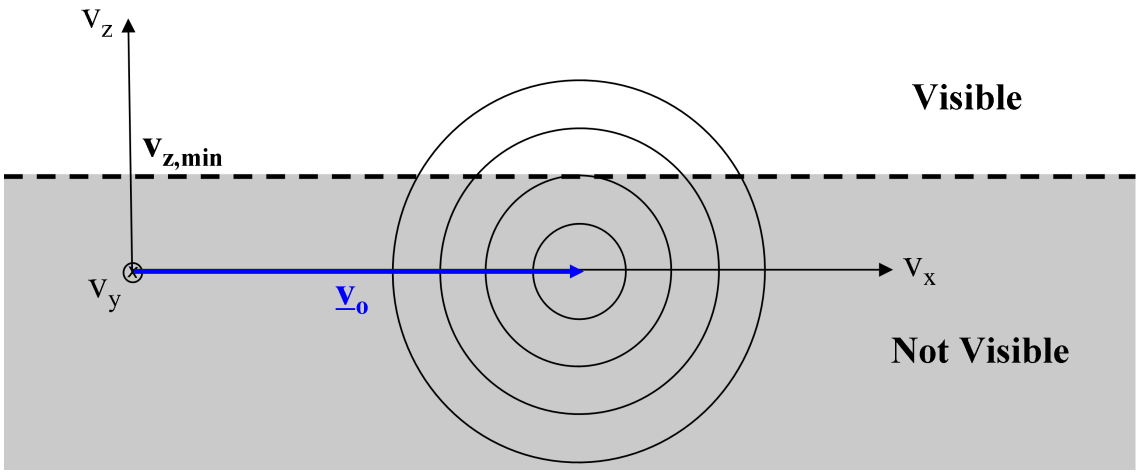

Figure D.1: Illustration of an obstructed isotropic drifting Maxwell-Boltzmann velocity distribution with center $[v_o, 0, 0]$. The distribution is visible for all values of v_x and v_y but blocked for $v_z < v_{z,min}$.

$$v_{th} \equiv \sqrt{\frac{kT}{m}}. \tag{D.2}$$

The plasma will be visible for all v_x and v_y but will be obstructed in the v_z direction for $v_z < v_{z,min}$, as shown in Figure D.1. It is assumed that no more than 50% of the distribution function will be in the observable $v_{z,min} > 0$ range.

The observed moments of interest are the 0th (density), 1st (velocity), and 2nd (temperature) moments of the distribution function, defined here as,

$$n_{obs} = \int\int\int dv_z dv_y dv_x \cdot f(v_x, v_y, v_z), \tag{D.3}$$

$$n_{obs} \langle \mathbf{v} \rangle = \int\int\int dv_z dv_y dv_x \cdot \mathbf{v} \cdot f(v_x, v_y, v_z), \text{ and} \tag{D.4}$$

$$n_{obs} \left\langle (v - v_o)^2 \right\rangle = \int\int\int dv_z dv_y dv_x \cdot (v - v_o)^2 \cdot f(v_x, v_y, v_z).. \tag{D.5}$$

To take into account the obstruction, the limits of integration in (D.3)-(D.5) are from $-\infty$ to $+\infty$ for $v_{\rm x}$ and $v_{\rm y}$, and for $v_{\rm z}$ are from $v_{\rm z,min}$ to $+\infty$.

D.1.1 Density

Integrating (D.3) yields,

$$n_{\rm obs} = \int_{\rm v_{z,min}}^{\infty} dv_{\rm z} \int_{-\infty}^{\infty} dv_{\rm y} \int_{-\infty}^{\infty} dv_{\rm x} \cdot f(v_{\rm x}, v_{\rm y}, v_{\rm z}) = \frac{n_{\rm o}}{2} \cdot erfc\left(\frac{v_{\rm z,min}}{v_{\rm th}}\right). \tag{D.6}$$

where $erfc$ is the complementary error function. Because only a partial distribution function is observed, only a fraction of the density is recovered, i.e., $n_{\rm obs} < n_{\rm o}/2$. Unless $v_{\rm z,min}/v_{\rm th}$ is well known, the total density $n_{\rm o}$ cannot be determined. Therefore, for an observing instrument, unless the location of the core of the distribution is well constrained, it will be difficult to recover the density from a partially observed distribution.

D.1.2 Velocity

Integrating (D.4) yields,

$$\langle v_{\mathrm{x}} \rangle = v_{\mathrm{o}},$$

(D.7)

$$\langle v_{\mathrm{y}} \rangle = v_{\mathrm{o}}, \text{ and}$$

(D.8)

$$\langle v_{\mathrm{z}} \rangle = \frac{\int\int\int v_{\mathrm{z}} \cdot dv_{\mathrm{z}} dv_{\mathrm{y}} dv_{\mathrm{x}} \cdot (v - v_{\mathrm{o}})^2 \cdot f(v_{\mathrm{x}}, v_{\mathrm{y}}, v_{\mathrm{z}})}{n_{\mathrm{obs}}} = \sqrt{\frac{2}{\pi}} \cdot v_{\mathrm{th}} \cdot \frac{exp\left(\frac{-v_{\mathrm{z,min}}^2}{2v_{\mathrm{th}}^2}\right)}{erfc\left(\frac{v_{\mathrm{z,min}}}{v_{\mathrm{th}}\sqrt{2}}\right)}.$$

$$\approx \begin{cases} \sqrt{\frac{2}{\pi}} \cdot v_{\mathrm{th}}, & \frac{v_{\mathrm{z,min}}}{v_{\mathrm{th}}} \ll 1 \\ v_{\mathrm{z,min}}, & \frac{v_{\mathrm{z,min}}}{v_{\mathrm{th}}} \gg 1 \end{cases}.$$

(D.9)

Unlike $\langle v_{\mathrm{x}} \rangle$ and $\langle v_{\mathrm{y}} \rangle$, $\langle v_{\mathrm{z}} \rangle$ will yield a non-zero value as a consequence of its non-symmetric limits of integration. Two limiting cases of $\langle v_{\mathrm{z}} \rangle$ can be considered here and are shown in (D.9). For $v_{\mathrm{th}}/v_{\mathrm{z,min}} \ll 1$, i.e., approximately 50% of the distribution is visible, $\langle v_{\mathrm{z}} \rangle$ is offset from the distribution center by approximately v_{th}. For $v_{\mathrm{th}}/v_{\mathrm{z,min}} \gg 1$, $\langle v_{\mathrm{z}} \rangle$ will be equal to $v_{\mathrm{z,min}}$, i.e., the location of the obstruction. For either case, for distribution with a bulk velocity $v_{\mathrm{o}}/v_{\mathrm{th}}, v_{\mathrm{o}}/v_{\mathrm{z,min}} \gg 1$, the error introduced into the determination of $\langle v \rangle$ from a $\langle v_{\mathrm{z}} \rangle$ offset will be relatively small.

D.1.3 Temperature

For $v_{\mathrm{o}}/v_{\mathrm{th}}$, $v_{\mathrm{o}}/v_{\mathrm{z,min}} \gg 1$, integrating (D.5) yields,

$$\left\langle (v - v_{\mathrm{o}})^2 \right\rangle \approx \left\langle (v_{\mathrm{x}} - v_{\mathrm{o}})^2 \right\rangle = \frac{\int\int\int dv_{\mathrm{z}} dv_{\mathrm{y}} dv_{\mathrm{x}} \cdot (v_{\mathrm{x}} - v_{\mathrm{o}})^2 \cdot f(v_{\mathrm{x}}, v_{\mathrm{y}}, v_{\mathrm{z}})}{n_{\mathrm{obs}}} = v_{\mathrm{th}}^2.$$

(D.10)

For a supersonic plasma, the thermal velocity can therefore be recovered independent of the fraction of the distribution that is visible. The recovered temperature from an isotropic Maxwellian velocity distribution is therefore expected to be accurate.

D.2 Derivation of observed distribution function

Equation (9.1) can be substituted into (9.2), and the Bessel function identity, $I_o(x) = (1/2\pi)\int_0^{2\pi} d\phi \cdot exp(xcos\phi)$ (*Arfken et al.* 2000), where I_o is modified Bessel function of the first kind, can be used to replace the integration over ϕ, yielding,

$$f_{obs}(v) = \int_0^{\pi/2} sin\theta d\theta \cdot n_o \left(\frac{1}{2\pi}\right)^{3/2} \frac{1}{v_{th}^3} exp\left(-\frac{v^2 + v_o^2 - 2vv_o cos\theta cos\theta_o)}{2v_{th}^2}\right) \times 2\pi \cdot I_o\left(\frac{vv_o sin\theta sin\theta_o}{v_{th}^2}\right). \quad \text{(D.11)}$$

For a supersonic distribution, i.e., $v_o/v_{th} \gg 1$, and using the asymptotic approximation, $I_o(x) \approx e^x/\sqrt{2\pi x}$, $x \gg 1/4$, (D.11) can be written as,

$$f_{obs}(v) = n_o \cdot \frac{1}{2\pi} \frac{1}{v_{th}^2\sqrt{vv_o}} \int_0^{\pi/2} d\theta \cdot \sqrt{\frac{sin\theta}{sin\theta_o}} \times exp\left(-\frac{v^2 + v_o^2 - 2vv_o cos(\theta - \theta_o)}{2v_{th}^2}\right). \quad \text{(D.12)}$$

Since $\theta \approx \theta_o \approx \pi/2$ for any expected solar wind distribution, equation (D.12) can be approximated as,

$$f_{\rm obs}(v) \approx n_{\rm o} \cdot \frac{1}{2\pi} \frac{1}{v_{\rm th}^2 \sqrt{v v_{\rm o}}} exp\left(-\frac{(v-v_{\rm o})^2}{2v_{\rm th}^2}\right) \cdot$$
$$\times \int_0^{\pi/2} d\theta \cdot \sqrt{\frac{sin\theta}{sin\theta_{\rm o}}} exp\left(-\frac{v v_{\rm o}}{2v_{\rm th}^2}(\theta-\theta_{\rm o})^2\right). \quad \text{(D.13)}$$

Assuming $\sqrt{\frac{sin\theta}{sin\theta_{\rm o}}} \approx 1$, equation (D.13) can be integrated analytically, yielding,

$$f_{\rm obs}(v) \approx n_{\rm o} \cdot D \cdot \frac{1}{\sqrt{2\pi} v v_{\rm o} v_{\rm th}} exp\left(-\frac{(v-v_{\rm o})^2}{2v_{\rm th}^2}\right), \quad \text{(D.14)}$$

where

$$D = \frac{1}{2} erfc\left(\sqrt{\frac{v v_{\rm o}}{2v_{\rm th}^2}}\left(\theta_{\rm o} - \frac{\pi}{2}\right)\right) \approx \frac{1}{2} erfc\left(\frac{v_{\rm o}}{v_{\rm th}\sqrt{2}}\left(\theta_{\rm o} - \frac{\pi}{2}\right)\right). \quad \text{(D.15)}$$

Equation (D.15) is in the form of the observed distribution function described by *von Steiger et al.* (2000) for Ulysses/SWICS, but with a different effective duty cycle factor, D. D depends on both v and $\theta_{\rm o}$. However, for a supersonic distribution, D is assumed to be constant in order to derive analytical moments of $f_{\rm obs}$.

APPENDIX E

Observations of interstellar helium pickup ions in the inner heliosphere

This chapter is taken from: Gershman, D. J., G. Gloeckler, J. A. Gilbert, J. M. Raines, L. A. Fisk, S. C. Solomon, E. C. Stone, and T. H. Zurbuchen (2012), Observations of interstellar helium pickup ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

E.1 Instrument Response Functions

In addition to observing pickup ions at different heliocentric distances, FIPS and SWICS have markedly different field of views, and consequently, analyze different sections of the pickup ion velocity distribution function, as illustrated in Figure E.1.

As discussed in Section 10.2.3, the pickup ion distribution is organized in the solar wind rest frame, not the spacecraft/instrument frame. Particles velocities relative to the center of this frame are typically normalized by the solar wind speed, $v_{\rm sw}$, such that they can be described by the vector $\mathbf{w} \equiv \mathbf{v}/v_{\rm sw}$. Previous pickup ion

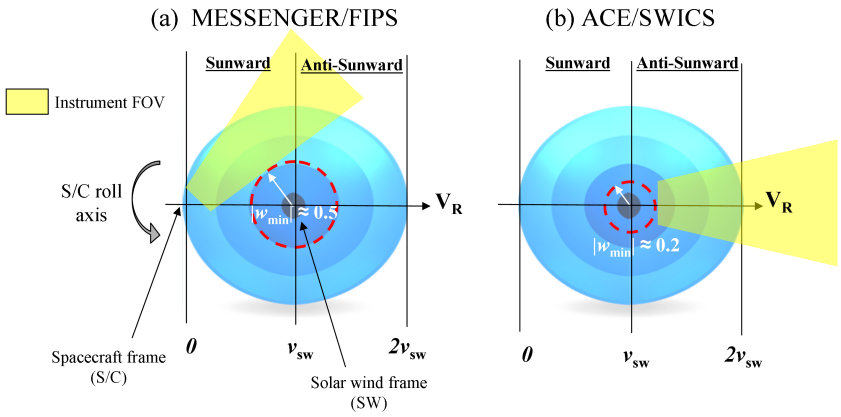

Figure E.1: Instrument field of view of a pickup ion distribution centered on the solar wind frame for (a) MESSENGER/FIPS and (b) ACE/SWICS. FIPS primarily views the sunward sector of the distribution; given the spacecraft roll-axis, other sectiosn of the distribution are sampled as the spacecraft rolls during flight. A smaller portion of the anti-sunward sector is visible to FIPS. ACE/SWICS, like most Sun-observing spacecraft, has a view exclusively in the anti-sunward sector. The dashed circle indicates the minimum observable velocity relative to the solar wind core that can be measured by each instrument.

observations indicate that the velocity distribution exhibits structure with respect to the anti-sunward/sunward directions, i.e. particles that move both faster and slow than the solar wind, respectively.

ACE, a solar wind monitor, has its instrument complement pointing directly towards the Sun, continuously measuring solar wind plasmas in the anti-sunward sector. In data processing, the lower energies are excluded to ensure that high fluxes of solar wind H^+ or He^{2+} particles do not contaminate the pickup He^+ time-of-flight track and that no He^+ of solar origin is included. These restrictions set a processing threshold of $|w| > 0.2$ for all ACE/SWICS data analysis of He^+ used here. Due to ACE's near circular orbit around the Sun (i.e. L1) and constant pointing, SWICS's view of the pickup ion distribution is approximately constant over time, though there will be some slight variations with solar wind speed.

Conversely, MESSENGER is intended to study the planet Mercury, with FIPS's direct view of the Sun masked by a large spacecraft sunshade. Due to this physi-

cal obstruction, FIPS cannot view particles with $|w| < 0.5$. Furthermore, particles measured with energy-per-charge ratios less than 500eV/e were excluded due to low signal-to-noise ratios at those values. As shown in Figure E.1, FIPS mostly views the sunward sector. While the instantaneous field of view is not symmetric about the radial direction, to first order the spacecraft roll axis is symmetric. As the spacecraft changes orientation in flight, different sections of the distribution function are sampled. The spacecraft rolled often during cruise, so that much of the sunward sector was measured at multiple heliocentric radii and heliographic longitudes.

These measurement constraints can be represented mathematically by an instrument response function, D, analogous to the factor used in *von Steiger et al.* (2000) and *Gershman et al.* (2012d). For SWICS, D will model visibility in the anti-sunward sector for $|w| > 0.2$. For FIPS, D will primarily be a function of $|w| > 0.5$ in the sunward-sector but also of spacecraft orientation. Due to the geometry of the FIPS field of view, $D(|w|)$ increases with increasing $|w|$, where $D(|w| = 1) > D(|w| \approx 0.5)$ and $D(|w| < 0.5) = 0$.

Newly picked up ions have $|w| \approx 1$ and cool to smaller $|w|$-shells as they are convected out into the heliosphere in the expanding solar wind. This cooling creates an effective mapping between the ions picked up at a particular heliocentric distance, R_{pickup}, and the $|w|$-value in the pickup ion velocity distribution scaled by the radial position of the instrument, R. For adiabatic cooling, this mapping is $R_{\text{pickup}} = R|w|^{3/2}$ (*Vasyliunas and Siscoe* 1976). With a $|w| > 0.5$ restriction for FIPS and a $|w| > 0.2$ restriction for SWICS, given their minimum heliocentric distances of 0.3 AU and 1 AU, respectively, neither instrument is expected to be sensitive to any ions produced inside R = 0.1 AU.

E.2 Pickup Ion Densities

The total pickup ion density as a function of the neutral density and ionization rate is,

$$n_{\rm ion}(R,\theta) = \frac{1}{R^2}\int\limits_{\rm R_{SUN}}^{R} dR' \cdot \frac{R'^2\beta(R')n(R',\theta)}{v_{\rm sw}}, \tag{E.1}$$

where $R_{\rm SUN}$ is the radius of the Sun, β is the local production (i.e. ionization) rate of He^+ and $v_{\rm sw}$ is the solar wind speed. The instrumentally observed density, $n_{\rm obs,ion}$, will be similar to the density from equation (E.1) but with the production rates convolved with the instrument response function D, such that,

$$n_{\rm obs,ion}(R,\theta) = \frac{1}{R^2}\int\limits_{\rm R_{SUN}}^{R} dR' \cdot D\left(|w| = \left(\frac{R}{R'}\right)^{2/3}\right) \cdot \frac{R'^2\beta(R')n(R',\theta)}{v_{\rm sw}}. \tag{E.2}$$

Following the discussion in Section 10.3.2, for radial distances with similar cone structure, i.e. constant downwind/upwind ratios, $n(R,\theta) \approx n(R)n(\theta)$. For this case,

$$n_{\rm obs,ion}(R,\theta) \approx \frac{1}{R^2}\int\limits_{\rm R_{SUN}}^{R} dR' \cdot D\left(|w|\right) \cdot \frac{R'^2\beta(R')n(R')n(\theta)}{v_{\rm sw}} \approx n(\theta)g(R), \tag{E.3}$$

where

$$g(R) \equiv \frac{1}{R^2}\int\limits_{\rm R_{SUN}}^{R} dR' \cdot D\left(|w|\right) \cdot \frac{R'^2\beta(R')n(R')}{v_{\rm sw}}. \tag{E.4}$$

Here, $g(R)$ is assumed to vary as $R^{-\alpha}$, a simple power law with heliocentric distance since n and β are expected to follow this form, and D should not be a strong function of R. With this approximation for $n_{\rm obs,ion}$, measured data from FIPS can be

scaled by R^{α} to normalize measurements to a single radial distance.

The largest errors in the approximation $n(R,\theta) \approx n(R)n(\theta)$ will occur for the range 0.1 AU $< R <$ 0.25 AU in equation (E.3), where the downwind enhancement varies by about 50% for the electron-impact ionization case in Section 10.3. However, consider the other contributions to the integral in equation (E.3). For all neutral density distributions and ionization profiles from Section 10.3.2, the scaled production rate $R^2\beta n$ increases with radial distance from the Sun inside of 0.25 AU. Furthermore, in the case of both SWICS, and FIPS, $D(|w|)$ increases with increasing $|w|$. Therefore, any contribution to the integral in equation (E.3) between 0.1 AU $< R <$ 0.25 AU should have minimal impact on the total instrument observed densities, even for MESSENGER/FIPS at $R = 0.3$ AU.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abramenko, V., V. Yurchyshyn, J. Linker, Z. Mikić, J. Luhmann, and C. O. Lee (2010), Low-latitude coronal holes at the minimum of the 23rd solar cycle, *Astrophys. J.*, *712*, 813–818.

Acuña, M. H., K. W. Ogilvie, D. N. Baker, S. A. Curtis, D. H. Fairfield, and W. H. Mish (1995), The global geospace science program and its investigations, *Space Sci. Rev.*, *71*, 5–21, doi:10.1007/BF00751323.

Alfred, R. L., F. A. Londry, and R. E. March (1993), Resonance excitation of ions stored in a quadrupole ion trap - Part IV: theory of quadrupolar excitation, *Int. J. Mass Spec. Ion Proc.*, *125*, 171–185, doi:10.1016/0168-1176(93)80040-L.

Allegrini, F., D. J. McComas, D. T. Young, J. J. Berthelier, J. Covinhes, J. M. Illiano, J. F. Riou, H. O. Funsten, and R. W. Harper (2006), Energy loss of 1–50 keV H, He, C, N, O, Ne, and Ar ions transmitted through thin carbon foils, *Rev. Sci. Instrum.*, *77*,501–044,501–7, doi:doi:10.1063/1.2185490.

Allegrini, F., R. F. Wimmer-Schweingruber, P. Wurz, and P. Bochsler (2003), Determination of low-energy ion-induced electron yields from thin carbon foils, *Nucl. Instrum. Meth. Phys. Res. B*, *211*, 487–494.

Anderson, B. J., et al. (2011), The global magnetic field of Mercury from MESSENGER orbital observations, *Science*, *333*(6051), 1859–1862, doi: 10.1126/science.1211001.

Anderson, B., M. Acuña, D. Lohr, J. Scheifele, A. Raval, H. Korth, and J. Slavin (2007), The Magnetometer instrument on MESSENGER, *Space Sci. Rev.*, *131*, 417–450.

Andrews, G., et al. (2007), The Energetic Particle and Plasma Spectrometer instrument on the MESSENGER spacecraft, *Space Sci. Rev.*, *131*, 523–556.

Arfken, G. B., H. J. Weber, and F. Harris (2000), *Mathematical Methods for Physicists*, 5th ed., Elsevier, Amsterdam, 1182pp.

Asbridge, J. R., S. J. Bame, W. C. Feldman, and M. D. Montgomery (1976), Helium and hydrogen velocity differences in the solar wind, *J. Geophys. Res.*, *81*, 2719–2727.

Auchère, F., D. R. McMullin, J. W. Cook, J. S. Newmark, R. von Steiger, and M. Witte (2005), A model for solar EUV flux helium photoionization throughout the 3-Dimensional heliosphere, in *Solar Wind 11/SOHO 16, "Connecting Sun and Heliosphere" Conference*, edited by B. Fleck, T. H. Zurbuchen, and H. Lacoste, Special Publication 592, European Space Agency, Noordwijk, The Netherlands, pp. 327-329.

Baker, D. N., et al. (2011), The space environment of Mercury at the times of the second and third MESSENGER flybys, *Planet. Space Sci.*, *59*(15), 2066–2074, doi: 10.1016/j.pss.2011.01.018.

Bale, S. D., P. J. Kellogg, F. S. Mozer, T. S. Horbury, and H. Reme (2005), Measurement of the electric fluctuation spectrum of magnetohydrodynamic turbulence, *Phys. Rev. Lett.*, *94*, id. 215002.

Balsiger, H., et al. (2007), Rosina-Rosetta orbiter spectrometer for ion and neutral analysis, *Space Sci. Rev.*, *128*, 745–801.

Bame, S. J., J. R. Asbridge, W. C. Feldman, M. D. Montgomery, and P. D. Kearney (1975), Solar wind heavy ion abundances, *Solar Phys.*, *43*, 463–473, doi: 10.1007/BF00152368.

Bauer, S., and H. Lammer (2004), *Planetary Aeronomy: Atmosphere Environments in Planetary Systems*, 1st ed., Springer, Berlin, Germany, 215pp.

Behrisch, R. (ed.) (1981), *Sputtering by particle bombardment I. Physical sputtering of single-element solids, Vol. 47*, Springer-Verlag, New York, United States, 281pp.

Belcher, J. W., and L. J. Davis (1971), Large-amplitude Alfvén waves in the interplanetary medium, 2, *J. Geophys. Res.*, *76*, 3534–3563.

Belton, M. J. S. (2003), *New Frontiers in the Solar System:An Integrated Exploration Strategy*, The National Academies Press, Washington, D.C., United States, 232 pp.

Bennett, S. (1993), *A History of Control Engineering 1930-1955*, Institution of Electrical Engineers, London, United Kingdom, 262 pp.

Berger, L., R. F. Wimmer-Schweingruber, and G. Gloeckler (2011), Systematic measurements of ion-proton differential streaming in the solar wind, *Phys. Rev. Lett.*, *106*, id. 151103.

Binder, K., and D. W. Heermann (2010), *Monte Carlo Simulation in Statistical Physics: An Introduction*, Springer, Berlin, Germany, 200pp.

Blum, P., J. Pfleiderer, and C. Wulf-Mathies (1975), Neutral gases of interstellar origin in interplanetary space, *Planet. Space Sci.*, *23*, 93–105.

Blum, P. W., and H. J. Fahr (1970), Interaction between interstellar hydrogen and the solar wind, *Astron. Astrophys.*, *4*, 280-290.

Bochsler, P., et al. (2010), Kinetic temperatures of iron ions in the solar wind observed with STEREO/PLASTIC, in *Twelfth International Solar Wind Conference* , edited by Maksimovic, K. Issautier, N. Meyer-Vernet, M. Moncuquet, and F. Pantellini, Conference Proceedings, vol. 1216, pp. 257-260, American Institute of Physics, Melville, N. Y., doi:10.1063/1.3395850.

Bodmer, R. (1996), The helium isotopic ratio as a test for minor ion fractionation in the solar wind acceleration process: SWICS/ULYSSES data compared with results from a multifluid model, University of Bern, Switzerland, Ph.D. Thesis.

Bürgi, A., et al. (1993), Charge exchange of low energy ions in thin carbon foils II: results for ions of B, C, F, Ne, Na, Si, S, Cl, Ar, K, and Fe, *J. Appl. Phys.*, *73*, 4130–4139.

Burlaga, L. F., F. B. McDonald, and N. F. Ness (1993), Cosmic ray modulation and the distant heliospheric magnetic field: Voyager 1 and 2 observations from 1986 to 1989, *J. Geophys. Res.*, *98*(A1), 1–11, doi:199310.1029/92JA01979.

Busch, F. (1980), A position-sensitive transmission time detector, *Nucl. Instrum. Meth.*, *171*, 71–74.

Bzowski, M., et al. (2012), Neutral interstellar helium parameters based on IBEX-Lo observations and test particle calculations, *Astrophys. J. Suppl. Ser.*, *198*, article 12, 27 pp.

Cameron, A. E., and D. F. Eggers (1948), An ion "velocitron", *Rev. Sci. Instrum.*, *19*, 605–607, doi:10.1063/1.1741336.

Campana, J. E. (1987), Time-of-flight mass spectrometry: a historical overview, *Instrum. Sci. Tech.*, *16*, 1–14, doi:10.1080/10739148708543625.

Carignan, G., and W. Pinkus (1968), Ogo-F04 experiment description, *Tech. Note 08041-3-T*, University of Michigan, Ann Arbor, Michigan, United States.

Carlson, C. W., and J. P. McFadden (1998), Design and application of imaging plasma instruments, in *Measurement Techniques in Space Plasmas: Particles, Geophys. Monog.*, *102*, 125-140.

Chambers, A. (2005), *Modern Vacuum Physics*, CRC Press, Florida, United States, 360pp.

Coleman, P. J. (1968), Turbulence, viscosity, and dissipation in the solar-wind plasma, *Astrophys. J.*, *153*, 371-388.

Comon, P., and C. Jutten (2010), *Handbook of blind source separation: independent component analysis and applications*, Academic Press, Oxford, United Kingdom, 856pp.

Cook, R. L., T. Porter, and L. Carpenter (1984), Distributed ray tracing, *SIGGRAPH Comput. Graph.*, *18*, 137–145, doi:10.1145/964965.808590.

Cranmer, S. R., A. A. van Ballegooijen, and R. J. Edgar (2007), Self-consistent coronal heating and solar wind acceleration from anisotropic magnetohydrodynamic turbulence, *Astrophys. J. Suppl. Ser.*, *171*, 520–551.

Dahl, D. A. (2000), SIMION for the personal computer in reflection, *Int. J. Mass Spectrom.*, *200*, 3–25, doi:10.1016/S1387-3806(00)00305-5.

Davis, G. M. (2002), *Noise reduction in speech applications*, CRC Press, Florida, United States, 407pp.

Dawson, P. H. (Ed.) (1997), *Quadrupole Mass Spectrometry and Its Applications*, 1st ed., Springer, Berlin, Germany, 349pp

DeGroot, M. H., and M. J. Schervish (2011), *Probability and Statistics*, 4th ed., Addison Wesley, Boston, Massachusetts, United States, 912pp.

Delcourt, D., F. Leblanc, K. Seki, N. Terada, T. Moore, and M. Fok (2007), Ion energization during substorms at Mercury, *Planet. Space Sci.*, *55*, 1502–1508, doi: 10.1016/j.pss.2006.11.026.

Devant, G., P. Fercocq, G. Lepetit, and O. Maulat (1989), Procede d'alimentation en tension des spectrographs de mass du type quadripolaire, French patent 2,620,568.

Douglas, D. J., and N. V. Konenkov (2002), Influence of the 6th and 10th spatial harmonics on the peak shape of a quadrupole mass filter with round rods, *Rapid Comm. Mass Spectrom.*, *16*, 1425–1431, doi:10.1002/rcm.735.

Fahr, H. J. (1978), Change of interstellar gas parameters in stellar-wind-dominated astrospheres solar case, *Astron. Astrophys.*, *66*, 103–117.

Fahr, H. J. (1971), The interplanetary hydrogen cone and its solar cycle variations, *Astron. Astrophys.*, *14*, 263-274.

Fahr, H. (1968), On the influence of neutral interstellar matter on the upper atmosphere, *Astrophys. Space Sci.*, *2*, 474–495.

Feldman, W. C., J. J. Lange, and F. Scherb (1972), Interstellar helium in interplanetary space, in *Solar Wind*, edited by C. P. Sonnett, P. J. Coleman, Jr., and J. M. Wilcox, Special Publication 308, NASA, Washington, D. C., pp. 684-697.

Fisk, L. A., G. Gloeckler, and N. A. Schwadron (2010), On theories for stochastic acceleration in the solar wind, *Astrophys. J.*, *720*, 533–540.

Fisk, L. A., and L. Zhao (2008), The heliospheric magnetic field and the solar wind during the solar cycle, in *Symposium S257: Universal Heliophysical Processes*, edited by N. Gopalswamy and D. F. Webb, Proceedings of the International Astronomical Union, vol. 4, Cambridge University Press, pp. 109-120.

Fisk, L. A., and G. Gloeckler (2006), The common spectrum for accelerated ions in the quiet-time solar wind, *Astrophys. J.*, *640*, L79–L82.

Fisk, L. A., N. A. Schwadron, and T. H. Zurbuchen (1999), Acceleration of the fast solar wind by the emergence of new magnetic flux, *J. Geophys. Res.*, *104*, 19,765–19,772.

Forbush, S. E. (1954), World-wide cosmic-ray variations, 1937-1952, *J. Geophys. Res.*, *59*, 525–542.

Foster, R. M. (1924), *A reactance theorem*, American Telephone and Telegraph Co., New York, United States, pp. 259-267.

Fowler, R. H., and L. Nordheim (1928), Electron emission in intense electric fields, *Royal Soc. Lond. Proc. Ser. A*, *119*, 173–181.

Frank, L. A. (1971), Plasma in the earth's polar magnetosphere, *J. Geophys. Res.*, *76*, 5202–5219, doi:197110.1029/JA076i022p05202.

Freeman, J. W. (1988), Estimates of solar wind heating inside 0.3 AU, *Geophys. Res. Lett.*, *15*, 88–91.

Frisch, U. (1996), *Turbulence*, Cambridge University Press, Cambridge, United Kingdom, 312 pp.

Funsten, H. O. (1995), Formation and survival of H- and C- ions transiting ultrathin carbon foils at keV energies, *Phys. Rev. B*, *52*, 8703.

Funsten, H. O., and D. J. McComas (1998), Limited resource plasma analyzers: Miniaturization concepts, in *Measurement Techniques in Space Plasmas: Particles, Geophys. Monog.*, *102*, 157–167.

Funsten, H. O., D. J. McComas, and B. L. Barraclough (1993), Ultrathin foils used for low-energy neutral atom imaging of the terrestrial magnetosphere, *Opt. Engin.*, *32*, 3090–3095.

Galvin, A. B., et al. (2008), The Plasma and Suprathermal Ion Composition (PLASTIC) investigation on the STEREO observatories, *Space. Sci. Rev.*, *136*, 437–486.

Gary, S. P., B. E. Goldstein, and J. T. Steinberg (2001), Helium ion acceleration and heating by Alfvén/cyclotron fluctuations in the solar wind, *J. Geophys. Res.*, *106*, 24,955–24,963.

Geiss, J., and G. Gloeckler (2003), Isotopic composition of H, He and Ne in the protosolar cloud, *Space Sci. Rev.*, *106*, 3–18, doi:10.1023/A:1024651232758.

Geiss, J., et al. (1995), The southern high-speed stream: Results from the SWICS instrument on Ulysses, *Science*, *268*, 1033–1036.

Geiss, J., P. Hirt, and H. Leutwyler (1970), On acceleration and motion of ions in corona and solar wind, *Solar Phys.*, *12*, 458–483.

Gershman, D. J., J. A. Gilbert, J. M. Raines, G. Gloeckler, S. C. Solomon, and T. H. Zurbuchen (2012a), Background noise in space-based time-of-flight sensors - part II: noise event forward modeling and removal through post-processing, *Space Sci. Rev.*, in prep.

Gershman, D. J., G. Gloeckler, J. A. Gilbert, J. M. Raines, L. A. Fisk, S. C. Solomon, E. Stone, and T. H. Zurbuchen (2012b), Observations of interstellar helium pickup ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

Gershman, D. J., M. Rubin, B. P. Block, M. Benna, P. R. Mahaffy, and T. Zurbuchen (2012c), Enabling the next generation of spaceborne quadrupole mass spectrometers, in *Earth and Space 2012: Engineering for Extreme Environments*, American Society of Civil Engineers, Pasadenca, CA.

Gershman, D. J., T. H. Zurbuchen, L. A. Fisk, J. A. Gilbert, J. M. Raines, B. J. Anderson, C. W. Smith, H. Korth, and S. C. Solomon (2012d), Solar wind alpha particles and heavy ions in the inner heliosphere, *J. Geophys. Res.*, in prep.

Gershman, D. J., B. P. Block, M. Rubin, M. Benna, P. R. Mahaffy, and T. H. Zurbuchen (2011), Higher order parametric excitation modes for spaceborne quadrupole mass spectrometers, *Rev. Sci. Instrum.*, *82*, 125, 109, doi: doi:10.1063/1.3669781.

Gershman, D. J., and T. H. Zurbuchen (2010), Modeling extreme ultraviolet suppression of electrostatic analyzers, *Rev. Sci. Instrum.*, *81*, 045,111, doi: 10.1063/1.3378685.

Gilbert, J. A., D. J. Gershman, G. Gloeckler, R. A. Lundgren, T. M. Orlando, J. McLain, and T. H. Zurbuchen (2012a), Background noise in space-based time-of-flight sensors - part i: Characterization and design considerations, *Space Sci. Rev.*, in prep.

Gilbert, J. A., S. T. Lepri, E. Landi, and T. H. Zurbuchen (2012), First measurements of the complete heavy ion charge state distributions of C, O, and Fe associated with interplanetary coronal mass ejections, *Astrophys. J.*, *751*, 1-8.

Gilbert, J. A., R. A. Lundgren, M. H. Panning, S. Rogacki, and T. H. Zurbuchen (2010), An optimized three-dimensional linear-electric-field time-of-flight analyzer, *Rev. Sci. Instrum.*, *81*, 3302.

Glasserman, P. (2004), *Monte Carlo methods in financial engineering*, Springer, Berlin, Germany, 596pp.

Gloeckler, G., et al. (2004), Observations of the helium focusing cone with pickup ions, *Astron. Astrophys.*, *426*, 845–854.

Gloeckler, G., and J. Geiss (2001), Heliospheric and interstellar phenomena deduced from pickup ion observations, *Space Sci. Rev.*, *97*, 169–181, doi: 10.1023/A:1011867320416.

Gloeckler, G., L. A. Fisk, J. Geiss, N. A. Schwadron, and T. H. Zurbuchen (2000a), Elemental composition of the inner source pickup ions, *J. Geophys. Res.*, *105*, 7459–7464.

Gloeckler, G., J. Geiss, N. A. Schwadron, L. A. Fisk, T. H. Zurbuchen, F. M. Ipavich, R. von Steiger, H. Balsiger, and B. Wilken (2000b), Interception of Comet Hyakutake's ion tail at a distance of 500 million kilometres, *Nature*, *404*, 576–578.

Gloeckler, G., et al. (1998), Investigation of the composition of solar and interstellar matter using solar wind and pickup ion measurements with SWICS and SWIMS on the ACE spacecraft, *Space Sci. Rev.*, *86*, 497–539, doi:10.1023/A:1005036131689.

Gloeckler, G., and J. Geiss (1998a), Interstellar and inner source pickup ions observed with SWICS on Ulysses, *Space Sci. Rev.*, *86*(1), 127–159, doi: 10.1023/A:1005019628054.

Gloeckler, G., and J. Geiss (1998b), Measurement of the abundance of Helium-3 in the sun and in the local interstellar cloud with SWICS on Ulysses, *Space Sci. Rev.*, *84*, 275–284.

Gloeckler, G., and J. Geiss (1996), Abundance of 3He in the local interstellar cloud, *Nature*, *381*, 210–212.

Gloeckler, G., N. A. Schwadron, L. A. Fisk, and J. Geiss (1995a), Weak pitch angle scattering of few MV rigidity ions from measurements of anisotropies in the distribution function of interstellar pickup H+, *Geophys. Res. Lett.*, *22*, 2665–2668.

Gloeckler, G., et al. (1995b), The solar wind and suprathermal ion composition investigation on the WIND spacecraft, *Space Sci. Rev.*, *71*(1), 79–124, doi: 10.1007/BF00751327.

Gloeckler, G., J. Geiss, H. Balsiger, L. A. Fisk, A. B. Galvin, F. M. Ipavich, K. W. Ogilvie, R. von Steiger, and B. Wilken (1993), Detection of interstellar pick-up hydrogen in the solar system, *Science*, *261*, 70–73.

Gloeckler, G., et al. (1992), The solar wind ion composition spectrometer, *Astron. Astrophys. Suppl. Ser.*, *92*, 267–289.

Gloeckler, G., et al. (1985), The charge-energy-mass spectrometer for 0.3-300 keV/e ions on the AMPTE CCE, *IEEE Trans. Geo. Rem. Sens.*, *GE-23*, 234–240, doi: 10.1109/TGRS.1985.289519.

Gloeckler, G., and K. Hsieh (1979), Time-of-flight technique for particle identification at energies from 2–400 keV/nucleon, *Nucl. Instrum. Meth.*, *165*, 537–544, doi:10.1016/0029-554X(79)90636-0.

Goldstein, B. E., M. Neugebauer, and E. J. Smith (1995), Alfvén waves, alpha particles, and pickup ions in the solar wind, *Geophys. Res. Lett.*, *22*, 3389–3392.

Goldstein, M. L., D. A. Roberts, and W. H. Matthaeus (1995), Magnetohydrodynamic turbulence in the solar wind, *Annu. Rev. Astron. Astrophys.*, *33*, 283–326.

González, R. C., and R. E. Woods (2008), *Digital Image Processing*, 3rd ed., Prentice Hall, New Jersey, United States, 976pp.

Goral, C. M., K. E. Torrance, D. P. Greenberg, and B. Battaile (1984), Modeling the interaction of light between diffuse surfaces, in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '84, p. 213–222, ACM, New York, NY, USA, doi:10.1145/800031.808601.

Gosling, J. T., D. J. McComas, D. A. Roberts, and R. M. Skoug (2009), A one-sided aspect of Alfvénic fluctuations in the solar wind, *Astrophys. J. Lett.*, *695*, L213–L216.

Graham, R. L., D. E. Knuth, and O. Patashnik (1994), *Concrete mathematics: a foundation for computer science*, Addison-Wesley, Boston, Massachusetts, United States, 672pp.

Gruntman, M. (1997), Energetic neutral atom imaging of space plasmas, *Rev. Sci. Instrum.*, *68*, 3617–3656, doi:doi:10.1063/1.1148389.

Gruntman, M. (1995), Extreme-ultraviolet radiation filtering by freestanding transmission gratings, *Appl. Opt.*, *34*, 5732–5737.

Haerendel, G., G. Paschmann, N. Sckopke, H. Rosenbauer, and P. C. Hedgecock (1978), The frontside boundary layer of the magnetosphere and the problem of reconnection, *J. Geophys. Res.*, *83*, 3195–3216, doi:197810.1029/JA083iA07p03195.

Hamilton, K., C. W. Smith, B. J. Vasquez, and R. J. Leamon (2008), Anisotropies and helicities in the solar wind inertial and dissipation ranges at 1 AU, *J. Geophys. Res.*, *113*, A01106.

Hamilton, D. C., G. Gloeckler, F. M. Ipavich, R. A. Lundgren, and R. B. Sheldon (1990), New high-resolution electrostatic ion mass analyzer using time of flight, *Rev. Sci. Instrum.*, *61*, 3104–3106.

Hartle, R. E., E. C. S. Jr, K. W. Ogilvie, J. D. Scudder, A. J. Lazarus, and S. K. Atreya (1982), Titan's ion exosphere observed from Voyager 1, *J. Geophys. Res.*, *87*, 1383–1394, doi:198210.1029/JA087iA03p01383.

Hartle, R. E., S. A. Curtis, and G. E. Thomas (1975), Mercury's helium exosphere, *J. Geophys. Res.*, *80*, 3689–3692, doi:197510.1029/JA080i025p03689.

Harvey, B., H. H. F. Smid, T. Pirard, B. Harvey, H. H. F. Smid, and T. Pirard (2010), *Emerging space powers*, Springer, Berlin, Germany, 732pp.

Hasselkamp, D. (1981), Ion induced electron emission from metal surfaces, *Nucl. Instrum. Meth.*, *180*, 349–356.

He, X. D., K. E. Torrance, F. X. Sillion, and D. P. Greenberg (1991), A comprehensive physical model for light reflection, *SIGGRAPH Comput. Graph.*, *25*, 175–186, doi: 10.1145/127719.122738.

Hefti, S., et al. (1998), Kinetic properties of solar wind minor ions and protons measured with SOHO/CELIAS, *J. Geophys. Res.*, *103*, 29,697–29,704.

Heroux, L., and H. E. Hinteregger (1978), Aeronomical reference spectrum for solar UV below 2000 a, *J. Geophys. Res.*, *83*, 5305–5308.

Herrero, F. A. (1992), Light-trap design using multiple reflections and solid-angle attenuation: application to a spaceborne electron spectrometer, *Appl. Opt.*, *31*, 5331–5340, doi:10.1364/AO.31.005331.

Hoffman, J. H., R. R. Hodges, F. S. Johnson, and D. E. Evans (1973), Lunar atmospheric composition results from Apollo 17, in *4th Lunar Science Conference, vol. 3*, March 5-8, Houston, Texas, United States, 11pp.

Hovestadt, D., et al. (1995), CELIAS - charge, element and isotope analysis system for SOHO, *Solar Phys.*, *162*, 441–481.

Hsieh, K. C., E. Keppler, and G. Schmidtke (1980), Extreme ultraviolet induced forward photoemission from thin carbon foils, *J. Appl. Phys.*, *51*, 2242–2246, doi: doi:10.1063/1.327849.

Ipavich, F. M., A. B. Galvin, G. Gloeckler, D. Hovestadt, B. Klecker, and M. Scholer (1986), Comet Giacobini-Zinner: in situ observations of energetic heavy ions, *Science*, *232*, 366–369.

Ipavich, F. M., L. S. Ma Sung, and G. Gloeckler (1982), *Tech. Report TR-82-172*, University of Maryland, Maryland, United States.

Ipavich, F. M., R. A. Lundgren, B. A. Lambird, and G. Gloeckler (1978), Measurements of pulse-height defect in Au-Si detectors for H, He, C, N, O, Ne, Ar, Kr from about 2 to 400 keV/nucleon, *Nucl. Instrum. Meth.*, *154*, 291–294.

Isenberg, P. A. (1997), A hemispherical model of anisotropic interstellar pickup ions, *J. Geophys. Res.*, *102*, 4719–4724.

Isenberg, P. A., and J. V. Hollweg (1982), Finite amplitude Alfvén waves in a multi-ion plasma - propagation, acceleration, and heating, *J. Geophys. Res.*, *87*, 5023–5029.

Issautier, K., N. Meyer-Vernet, M. Moncuquet, and S. Hoang (1998), Solar wind radial and latitudinal structure: Electron density and core temperature from Ulysses thermal noise spectroscopy, *J. Geophys. Res.*, *103*, 1969–1979.

Jackson, J. D. (1998), *Classical Electrodynamics Third Edition*, 3rd ed., Wiley, Hoboken, New Jersey, United States, 808pp.

Jensen, H. W. (1996), Global illumination using photon maps, in *Rendering Techniques '96, Proceedings of the Seventh Eurographics Workshop on Rendering* edited by X. Pueyo, and P. Schröder, Eurographics Association, Switzerland, 294 pp.

Jiang, Y., X. Fang, Y. Rang, and D. Tian (2008), Development and application of digital control technology of resonant excitation for quadrupole ion trap, *C. J. Anal. Chem.*, *36*, 715–718, doi:10.1016/S1872-2040(08)60037-7.

Kabin, K. (2000), Interaction of Mercury with the solar wind, *Icarus*, *143*, 397–406, doi:10.1006/icar.1999.6252.

Kajiya, J. T. (1986), The rendering equation, in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '86, p. 143–150, ACM, New York, New York, United States, doi:10.1145/15922.15902.

Kallenbach, R., J. Geiss, G. Gloeckler, and R. von Steiger (2000), Pick-up ion measurements in the heliosphere - a review, *Astrophys. Space Sci.*, *274*, 97–114.

Kalos, M. H., and P. A. Whitlock (2009), *Monte Carlo Methods*, 2nd ed., Wiley, Hoboken, New Jersey, United States, 215pp.

Kasper, J. C., A. J. Lazarus, and S. P. Gary (2008), Hot solar-wind helium: Direct evidence for local heating by Alfvén-Cyclotron dissipation, *Phys. Rev. Lett.*, *101*, doi:10.1103/PhysRevLett.101.261103.

Kasprzak, W. T., H. B. Niemann, A. E. Hedin, S. W. Bougher, and D. M. Hunten (1993), Neutral composition measurements by the Pioneer Venus neutral mass spectrometer during orbiter re-entry, *Geophys. Res. Lett.*, *20*, 2747–2750, doi: 199310.1029/93GL02241.

Kerr, R. A. (2003), Failed CONTOUR comet mission melted itself down, *Science*, *302*(5645), 546–546, doi:10.1126/science.302.5645.546a.

Killen, R., et al. (2007), Processes that promote and deplete the exosphere of Mercury, *Space Sci. Rev.*, *132*, 433–509, doi:10.1007/s11214-007-9232-0.

Kissel, J., and F. R. Krueger (1987), The organic component in dust from Comet Halley as measured by the PUMA mass spectrometer on board Vega 1, *Nature*, *326*, 755–760.

Kohl, J. L., et al. (1998), UVCS/SOHO empirical determinations of anisotropic velocity distributions in the solar corona, *Astrophys. J. Lett.*, *501*, doi:10.1086/311434.

Kohl, J. L., et al. (1997), First results from the SOHO ultraviolet coronagraph spectrometer, *Solar Phys.*, *175*, 613–644.

Konenkov, N., A. Korolkov, and M. Machmudov (2005), Upper stability island of the quadrupole mass filter with amplitude modulation of the applied voltages, *J. Am. Soc. Mass Spectrom*, *16*, 379–387, doi:10.1016/j.jasms.2004.11.023.

Konenkov, N., L. Cousins, V. Baranov, and M. Sudakov (2001), Quadrupole mass filter operation with auxiliary quadrupolar excitation: theory and experiment, *Int. J. Mass Spectrom.*, *208*, 17–27, doi:10.1016/S1387-3806(01)00375-X.

Korde, R., and J. Geist (1987), Quantum efficiency stability of silicon photodiodes, *Appl. Opt.*, *26*, 5284–5290, doi:10.1364/AO.26.005284.

Koutroumpa, D., M. R. Collier, K. D. Kuntz, R. Lallement, and S. L. Snowden (2009), Solar wind charge exchange emission from the helium focusing cone: Model to data comparison, *Astrophys. J.*, *697*, 1214–1225.

Kulov, S. K., S. A. Kesaev, I. R. Bugulova, J. L. Pergamentsev, V. J. Boyadjidy, N. V. Berishvili, K. J. Ahpolov, and S. G. Manukov (2005), Quality of microchannel plate working surfaces, *Proc. SPIE 5834*, 203, doi:10.1117/12.628886

Kuo, H. (1996), *White noise distribution theory*, CRC Press, Florida, United States, 378pp.

Lallement, R., J. C. Raymond, J. Bertaux, E. Quémerais, Y. Ko, M. Uzzo, D. McMullin, and D. Rucinski (2004), Solar cycle dependence of the helium focusing cone from SOHO/UVCS observations: electron impact rates and associated pickup ions, *Astron. Astrophys.*, *426*, 867–874.

Leamon, R. J., C. W. Smith, N. F. Ness, W. H. Matthaeus, and H. K. Wong (1998), Observational constraints on the dynamics of the interplanetary magnetic field dissipation range, *J. Geophys. Res.*, *103*, 4775-4787.

Leamon, R. J., C. W. Smith, N. F. Ness, and H. K. Wong (1999), Dissipation range dynamics: kinetic Alfvén waves and the importance of β_e, *J. Geophys. Res.*, *104*, 22,331–22,344.

Li, H., and K. E. Torrance (2005), A practical, comprehensive light reflection model, *Tech. Rep. PCG-05-03*, Cornell University, Ithaca, New York, United States.

Lide, D. R. (2004), *CRC handbook of chemistry and physics: a ready-reference book of chemical and physical data*, CRC Press, Florida, United States, 2712pp.

Lin, R. P., et al. (1995), A Three-Dimensional plasma and energetic particle investigation for the wind spacecraft, *Space Sci. Rev.*, *71*, 125–153.

Lukyanov, A., S. Barabash, and M. Holmström (2004), Energetic neutral atom imaging at Mercury, *Adv. Space Res.*, *33*, 1890–1898, doi:10.1016/j.asr.2003.05.035.

Lunine, J. I. (2001), The occurrence of Jovian planets and the habitability of planetary systems, *Proc. Nat. Acad. Sci.*, *98*(3), 809–814, doi:10.1073/pnas.98.3.809.

Mabry, D. J., S. J. Hansel, and J. B. Blake (1993), The SAMPEX data processing unit, *IEEE Trans. Geo. Rem. Sens.*, *31*, 572–574, doi:10.1109/36.225524.

Madey, T. E. (1971), Electron-Stimulated desorption and work function studies of clean and cesiated (110) GaAs, *J. Vac. Sci. Tech.*, *8*, 39.

Mahaffy, P. R., et al. (2012), The sample analysis at Mars investigation and instrument suite, *Space Sci. Rev.*, in revision.

Mahaffy, P. (2002), Neutral gas and ion measurements by the CONTOUR mission, in *34th COSPAR Scientific Assembly, vol. 34*, Houston, Texas, United States.

Mahaffy, P. R., H. B. Niemann, A. Alpert, S. K. Atreya, J. Demick, T. M. Donahue, D. N. Harpold, and T. C. Owen (2000), Noble gas abundance and isotope ratios in the atmosphere of Jupiter from the Galileo probe mass spectrometer, *J. Geophys. Res.*, *105*(E6), PP. 15,061–15,071, doi:200010.1029/1999JE001224.

Mahaffy, P., T. Donahue, S. Atreya, T. Owen, and H. Niemann (1998), Galileo probe measurements of D/H and 3He/4He in Jupiter's atmosphere, *Space Sci. Rev.*, *84*, 251–263, doi:10.1023/A:1005091806594.

Maksimovic, M., S. P. Gary, and R. M. Skoug (2000), Solar wind electron suprathermal strength and temperature gradients: Ulysses observations, *J. Geophys. Res.*, *105* 18,337–18,350.

Manly, B. F. J. (2007), *Randomization, bootstrap and Monte Carlo methods in biology*, CRC Press, Florida, United States, 424pp.

March, R. E., and J. F. J. Todd (2005), *Quadrupole ion trap mass spectrometry*, 2nd ed., Wiley, Hoboken, New Jersey, United States, 392pp.

Marchand, P., and P. Marmet (1964), Opération d'un filtre de masse quadrupolaire en condition optimum, *Can. J. Phys.*, *42*, 1914–1919.

Marsch, E. (1991), *Turbulence in the Solar Wind*, in *Reviews in Modern Astronomy*, edited by G. Klare, Springer, Berlin, Germany, 43pp.

Marsch, E., K. H. Mühlhäuser, H. Rosenbauer, R. Schwenn, and F. M. Neubauer (1982a), Solar wind helium ions: Observations of the Helios solar probes between 0.3 and 1 AU, *J. Geophys. Res.*, *87*, 35–51.

Marsch, E., K. H. Mühlhäuser, R. Schwenn, H. Rosenbauer, W. Pilipp, and F. M. Neubauer (1982b), Solar wind protons: three-dimensional velocity distributions and derived plasma parameters measured between 0.3 and 1 AU, *J. Geophys. Res.*, *87*, 52–72.

McComas, D. J., R. W. Ebert, H. A. Elliott, B. E. Goldstein, J. T. Gosling, N. A. Schwadron, and R. M. Skoug (2008), Weaker solar wind from the polar coronal holes and the whole Sun, *Geophys. Res. Lett.*, *35*, L18103, doi:10.1029/2008GL034896.

McComas, D., et al. (2004a), The Interstellar Boundary Explorer (IBEX), in *Physics of the Outer Heliosphere Conference*, edited by V. Florinski, N. V. Pogorelov, and G. P. Zank, Conference Proceedings vol. 719, American Institute of Physics, Melville, N. Y., pp. 162-181.

McComas, D. J., et al. (2004b), The interstellar hydrogen shadow: Observations of interstellar pickup ions beyond Jupiter, *J. Geophys. Res.*, *109*, A02104, doi:10.1029/2003JA010217.

McComas, D., S. Bame, P. Barker, W. Feldman, J. Phillips, P. Riley, and J. Griffee (1998a), Solar Wind Electron Proton Alpha Monitor (SWEPAM) for the Advanced Composition Explorer, *Space Sci. Rev.*, *86*, 563–612.

McComas, D. J., et al. (1998b), Ulysses' return to the slow solar wind, *Geophys. Res. Lett.*, *25*, 1–4.

McComas, D. J., and J. E. Nordholt (1990), New approach to 3-D, high sensitivity, high mass resolution space plasma composition measurements, *Rev. Sci. Instrum.*, *61*, 3095–3097.

McDonald, G. E. (1975), Spectral reflectance properties of black chrome for use as a solar selective coating, *Solar Energy*, *17*(2), 119–122, doi:10.1016/0038-092X(75)90067-5.

McKenzie, J. F., M. Banaszkiewicz, and W. I. Axford (1995), Acceleration of the high speed solar wind., *Astron. and Astrophys.*, *303*, L45-L48.

McLain, J. L., A. L. Sprague, G. A. Grieves, D. Schriver, P. Travinicek, and T. M. Orlando (2011), Electron-stimulated desorption of silicates: A potential source for ions in Mercury's space environment, *J. Geophys. Res.*, *116*, 03,007.

McMullin, D. R., et al. (2004), Heliospheric conditions that affect the interstellar gas inside the heliosphere, *Astron. Astrophys.*, *426*, 885–895.

McMullin, D. R., D. L. Judge, E. Phillips, M. Hilchenbach, P. Bochsler, P. Wurz, E. Moebius, and F. Ipavich (2002), Measuring the ionization rate of in-flowing interstellar helium with the SOHO/CELIAS/SEM, in *From Solar Min to Max: Half a Solar Cycle with SOHO*, edited by A. Wilson, Special Publication 508, European Space Agency, Noordwijk, The Netherlands, pp. 489–491.

Meier, R. R. (1977), Some optical and kinetic properties of the nearby interstellar gas, *Astron. and Astrophys.*, *55*, 211–219.

Menzel, D., and R. Gomer (1964), Desorption from metal surfaces by low-energy electrons, *J. Chem. Phys.*, *41*, 3311–3328.

Miseki, K. (1993), Quadrupole mass spectrometer, United States patent 5,227,629.

Mitchell, D. G., et al. (2000), High energy neutral atom (HENA) imager for the IMAGE mission, *Space Sci. Rev.*, *91*, 67–112.

Möbius, E., et al. (2012), Interstellar gas flow parameters derived from Interstellar Boundary Explorer-Lo observations in 2009 and 2010: Analytical analysis, *Astrophys. J. Suppl. Ser.*, *198*, article 11, 18 pp.

Möbius, E., et al. (2010), He pickup ions in the inner Heliosphere—Diagnostics of the local interstellar gas and of interplanetary conditions, in *Pickup Ions throughout the Heliosphere and Beyond*, edited by J. Roux, G. Zank, A. J. Coates, and V. Florinski, Conference Proceedings vol. 1302, American Institute of Physics, Melville, N. Y., pp. 37–43.

Möbius, E., et al. (2004), Synopsis of the interstellar He parameters from combined neutral gas, pickup ion and UV scattering observations and related consequences, *Astron. Astrophys.*, *426*, 897–907.

Möbius, E., et al. (1998), The solar energetic particle ionic charge analyzer (SEPICA) and the data processing unit (S3DPU) for SWICS, SWIMS and SEPICA, *Space Sci. Rev.*, *86*, 449–495.

Möbius, E. (1996), The local interstellar medium viewed through pickup ions, recent results and future perspectives, *Space Sci. Rev.*, *78*, 375–386.

Möbius, E., D. Hovestadt, B. Klecker, M. Scholer, G. Gloeckler, and F. M. Ipavich (1985), Direct observation of He+ pick-up ions of interstellar origin in the solar wind, *Nature*, *318*, 426–429.

Möbius, E., D. Rucinski, D. Hovestadt, and B. Klecker (1995), The helium parameters of the very local interstellar medium as derived from the distribution of He+ pickup ions in the solar wind., *Astron. Astrophys.*, *304*, 505.

Möbius, E., et al. (1985b), The Time-of-Flight spectrometer SULEICA for ions of the energy range 5-270 keV/Charge on AMPTE IRM, *IEEE Trans. Geo. Rem. Sens.*, *3*, 274–279, doi:10.1109/TGRS.1985.289527.

Moore, T., et al. (2000), The low-energy neutral atom imager for IMAGE, *Space Sci. Rev.*, *91*, 155–195, doi:10.1023/A:1005211509003.

Mukherjee, P., M. Kang, T. H. Zurbuchen, L. J. Guo, and F. A. Herrero (2007), Fabrication of high aspect ratio Si nanogratings with smooth sidewalls for a deep UV-blocking particle filter, *J. Vac. Sci. Tech. B*, *25*, 2645.

Neugebauer, M. (1976), The role of Coulomb collisions in limiting differential flow and temperature differences in the solar wind, *J. Geophys. Res.*, *81*, 78–82.

Neugebauer, M., A. J. Lazarus, H. Balsiger, S. A. Fuselier, F. M. Neubauer, and H. Rosenbauer (1989), The velocity distributions of cometary protons picked up by the solar wind, *J. Geophys. Res.*, *94*, 5227–5239.

Niemann, H. B., et al. (2005), The abundances of constituents of Titan's atmosphere from the GCMS instrument on the huygens probe, *Nature*, *438*, 779–784.

Niemann, H. B., et al. (1998), The Planet-B neutral gas mass spectrometer, *Earth Planet Space*, *50*, 785–792.

Niemann, H., et al. (1997), The gas chromatograph mass spectrometer aboard Huygens, in *Huygens: Science, Payload and Mission*, Proceedings of an European Space Agency conference. Edited by A. Wilson (1997)., p.85

Niemann, H. B., D. N. Harpold, S. K. Atreya, G. R. Carignan, D. M. Hunten, and T. C. Owen (1992), Galileo probe mass spectrometer experiment, *Space Sci. Rev.*, *60*, 111–142, doi:10.1007/BF00216852.

Niemann, H. B., J. R. Booth, J. E. Cooley, R. E. Hartle, W. T. Kasprzak, N. W. Spencer, S. H. Way, D. M. Hunten, and G. R. Carignan (1980), Pioneer Venus orbiter neutral gas mass spectrometer experiment, *IEEE Trans. Geo. Rem. Sens.*, *1*, 60–65, doi:10.1109/TGRS.1980.350282.

Niemann, H. B., R. E. Hartle, W. T. Kasprzak, N. W. Spencer, D. M. Hunten, and G. R. Carignan (1979), Venus upper atmosphere neutral composition - preliminary results from the Pioneer Venus orbiter, *Science*, *203*, 770–772.

Noda, H., T. Terasawa, Y. Saito, H. Hayakawa, A. Matsuoka, and T. Mukai (2001), Observation of the interstellar helium cone by the NOZOMI spacecraft, *Space Sci. Rev.*, *97*, 423–426.

Ogilvie, K. W., et al. (1995), SWE, a comprehensive plasma instrument for the WIND spacecraft, *Space Sci. Rev.*, *71*, 55–77.

Ogilvy, J. A. (1991), *Theory of Wave Scattering From Random Rough Surfaces,*, 1 ed., Taylor & Francis, London, United Kingdom, 292pp.

Pan, J., J. Lv, Z. Zhang, J. Sun, and D. Su (2010), Ion feedback suppression for microchannel plate applied to third generation image intensifiers, *C. J. Electron.*, *19*(4).

Paolini, F. R., and G. C. Theodoridis (1967), Charged particle transmission through spherical plate electrostatic analyzers, *Rev. Sci. Instrum.*, *38*, 579–588, doi: doi:10.1063/1.1720771.

Paresce, F. (1975), Quantum efficiency of a channel electron multiplier in the far ultraviolet, *Appl. Opt.*, *14*, 2823–2824, PMID: 20155117.

Parker, E. N. (1965), Dynamical theory of the solar wind, *Space Sci. Rev.*, *4*, 666–708.

Parker, E. N. (1958), Dynamics of the interplanetary gas and magnetic fields., *Astrophys. J.*, *128*, 664-676.

Paul, W., H. P. Reinhard, and U. von Zahn (1958), Das elektrische massenfilter als massenspektrometer und isotopentrenner, *Z. Phys.*, *152*, 143–182.

Paul, W., and H. Steinwedel (1953), Ein neues massenspektrometer ohne magnetfeld, *Z. Natur. Teil A*, *8*, 448.

Pelz, D. T., C. A. Reber, A. E. Heden, and G. R. Carignan (1973), A neutral-atmosphere composition experiment for the atmosphere Explorer-C -D, and -E, *Rad. Sci.*, *8*, 277.

Pfeffer, W., B. Kohlmeyer, and W. F. W. Schneider (1973), A fast zero-time detector for heavy ions using the channel electron multiplier, *Nucl. Instrum. Meth.*, *107*, 121.

Pollock, C. J., et al. (2000), Medium energy neutral atom (MENA) imager for the IMAGE mission, *Space Sci. Rev.*, *91*, 113–154.

Raines, J. M., et al. (2012), Distribution and compositional variations of plasma ions in Mercury's space environment: The first three Mercury years of MESSENGER observations, *J.Geophys. Res.*, in prep.

Raines, J. M., J. A. Slavin, T. H. Zurbuchen, G. Gloeckler, B. J. Anderson, D. N. Baker, H. Korth, S. M. Krimigis, and R. L. McNutt Jr (2011), MESSENGER observations of the plasma environment near Mercury, *Planet. Space Sci.*, *59*(15), 2004–2015, doi:10.1016/j.pss.2011.02.004.

Read, F. H., N. J. Bowring, P. D. Bullivant, and R. R. A. Ward (1998), Penetration of electrostatic fields and potentials through meshes, grids, or gauzes, *Rev. Sci. Instrum.*, *69*, 2000–2006.

Ritzau, S. M., and R. A. Baragiola (1998), Electron emission from carbon foils induced by keV ions, *Phys. Rev. B*, *58*, 2529–2538.

Roberts, D. A., M. L. Goldstein, L. W. Klein, and W. H. Matthaeus (1987), Origin and evolution of fluctuations in the solar wind: Helios observations and Helios-Voyager comparisons, *J. Geophys. Res.*, *92*,12,023–12,035.

Rohner, U., L. Saul, P. Wurz, F. Allegrini, J. Scheer, and D. McComas (2012), A simple 3D plasma instrument with an electrically adjustable geometric factor for space research, *Meas. Sci. Tech.*, *23*, 901, doi:10.1088/0957-0233/23/2/025901.

Rosenbauer, H., et al. (1977), A survey on initial results of the Helios plasma experiment, *J. Geophys. Res.*, *42*, 561–580.

Rucinski, D., M. Bzowski, and H. J. Fahr (2003), Imprints from the solar cycle on the helium atom and helium pickup ion distributions, *Annal. Geophys.*, *21*, 1315–1330.

Rucinski, D., and H. J. Fahr (1989), The influence of electron impact ionization on the distribution of interstellar helium in the inner heliosphere - possible consequences for determination of interstellar helium parameters, *Astron. Astrophys.*, *224*, 290–298.

Samson, J. A. R., Z. X. He, L. Yin, and G. N. Haddad (1994), Precision measurements of the absolute photoionization cross sections of He, *J. Phys. B*, *27*, 887–898.

Sarantos, M., P. H. Reiff, T. W. Hill, R. M. Killen, and A. L. Urquhart (2001), A Bx-interconnected magnetosphere model for Mercury, *Planet. Space Sci.*, *49*, 1629–1635, doi:10.1016/S0032-0633(01)00100-3.

Saul, L., P. Wurz, and R. Kallenbach (2009), A measurement of the adiabatic cooling index for interstellar helium pickup ions in the inner heliosphere, *Astrophys. J.*, *703*, 325–329.

Saul, L., E. Mobius, P. Isenberg, and P. Bochsler (2007), On pitch-angle scattering rates of interstellar pickup ions as determined by in situ measurement of velocity distributions, *Astrophys. J.*, *655*, 672–677.

Saul, L., et al. (2003), SOHO CTOF observations of interstellar He+ pickup ion enhancements in solar wind compression regions, in *Solar Wind Ten*, edited by M. Velli, F. Malara, and R. Bruno, Conference Proceedings vol. 679, American Institute of Physics, Melville, N. Y., pp. 778-781.

Schaefer, R. T., J. A. MacAskill, M. Mojarradi, A. Chutjian, M. R. Darrach, S. M. Madzunkov, and B. J. Shortt (2008), Digitally synthesized high purity, high-voltage radio frequency drive electronics for mass spectrometry, *Rev. Sci. Instrum.*, *79*, 107, doi:doi:10.1063/1.2981691.

Schneider, W. (1970), Mass-identification of alpha-particles and heavy ions by time-of-flight methods, *Nucl. Instrum. Meth.*, *87*, 253–259.

Schulte, J., P. V. Shevchenko, and A. V. Radchik (1999), Nonlinear field effects in quadrupole mass filters, *Rev. Sci. Instrum.*, *70*, 3566–3571, doi:10.1063/1.1149960.

Schwadron, N. A. (1998), A model for pickup ion transport in the heliosphere in the limit of uniform hemispheric distributions, *J. Geophys. Res.*, *103*, 20,643–20,650.

Schwenn, R., and E. Marsch (Eds.) (1990), *Physics of the Inner Heliosphere I: Large-Scale Phenomena (Physics and Chemistry in Space)*, 1st ed., Springer, Berlin, Germany, 282pp.

Schwenn, R., H. Rosenbauer, and H. Miggenrieder (1975), The plasma experiment on board Helios/E 1, *Raumfahrtforschung*, *19*, 226–232.

Scime, E. E., D. J. McComas, E. H. Anderson, and M. L. Schattenburg (1995), Extreme-ultraviolet polarization and filtering with gold transmission gratings, *Appl. Opt.*, *34*, 648–654.

Scott, D. W. (1992), *Multivariate density estimation: theory, practice, and visualization*, Wiley, Hoboken, New Jersey, United States, 317pp.

Shenoi, B. A. (2005), *Introduction to Digital Signal Processing and Filter Design*, 1 ed., Wiley, Hoboken, New Jersey, United States, 423pp.

Silverman, B. W. (1986), *Density estimation for statistics and data analysis*, CRC Press, Florida, United States, 176pp.

Simpson, J. A. (1983), Elemental and isotopic composition of the galactic cosmic rays, *Ann. Rev. Nucl. Part. Sci.*, *33*, 323–382.

Slavin, J. A., et al. (2010), MESSENGER observations of extreme loading and unloading of Mercury's magnetic tail, *Science*, *329*(5992), 665–668, doi: 10.1126/science.1188067.

Slavin, J. A., et al. (2008), Mercury's magnetosphere after MESSENGER's first flyby, *Science*, *321*(5885), 85–89, doi:10.1126/science.1159040.

Smirnov, A. (1999), *Processing of multidimensional signals*, Springer, Berlin, Germany, 271 pp.

Smith, C. W., B. J. Vasquez, and J. V. Hollweg (2012), Observational constraints on the role of cyclotron damping and kinetic Alfvén waves in the solar wind, *Astrophys. J.*, *745*, 1-8.

Smith, C. W., J. L'Heureux, N. F. Ness, M. H. Acuña, L. F. Burlaga, and J. Scheifele (1998), The ACE magnetic fields experiment, *Space Sci. Rev.*, *86*, 613–632.

Smith, E. J., A. Balogh, M. Neugebauer, and D. McComas (1995), Ulysses observations of Alfvén waves in the southern and northern solar hemispheres, *Geophys. Res. Lett.*, *22*, 3381–3384.

Solomon, S. C., R. L. McNutt, R. E. Gold, and D. L. Domingue (2007), MESSENGER mission overview, *Space Sci. Rev.*, *131*, 3–39.

Solomon, S. C., et al. (2001), The MESSENGER mission to Mercury: scientific objectives and implementation, *Planet. Space Sci.*, *49*, 1445–1465, doi:10.1016/S0032-0633(01)00085-X.

Stone, E., A. Frandsen, R. Mewaldt, E. Christian, D. Margolies, J. Ormes, and F. Snow (1998), The Advanced Composition Explorer, *Space Sci. Rev.*, *86*, 1–22.

Sudakov, M., N. Konenkov, D. Douglas, and T. Glebova (2000), Excitation frequencies of ions confined in a quadrupole field with quadrupole excitation, *J. Am. Soc. Mass Spectrom.*, *11*, 10–18, doi:10.1016/S1044-0305(99)00111-7.

Then, A. (1990), Formation and behavior of surface layers on electron emission glasses, *J. Non Crys. Sol.*, *120*, 178–187.

Theodoridis, G. C., and F. R. Paolini (1969), The angular response of spherical plate electrostatic analyzers, *Rev. Sci. Instrum.*, *40*, 621–631, doi:doi:10.1063/1.1684022.

Thomas, G. E. (1978), The interstellar wind and its influence on the interplanetary environment, *Ann. Rev. Earth Planet. Sci.*, *6*, 173–204.

Totten, T. L., J. W. Freeman, and S. Arya (1995), An empirical determination of the polytropic index for the free-streaming solar wind using Helios 1 data, *J. Geophys. Res.*, *93*, 13–17.

Tuzlukov, V. P. (2002), *Signal processing noise*, CRC Press, Florida, United States, 663pp.

Vallerga, J., R. Lallement, M. Lemoine, F. Dalaudier, and D. McMullin (2004), EUVE observations of the helium glow: Interstellar and solar parameters, *Astron. Astrophys.*, *426*, 855–865.

Vampola, A. L. (1998), Measuring energetic electrons: what works and what doesn't, in *Measurement Techniques in Space Plasmas: Particles, Geophys. Monog.*, *102*, 339-355.

Vankka, J., and K. A. Halonen (2001), *Direct Digital Synthesizers: Theory, Design and Applications*, 1 ed., Springer, Berlin, Germany, 216pp.

Vaseghi, S. V. (2008), *Advanced Digital Signal Processing and Noise Reduction*, Wiley, Hoboken, New Jersey, United States, 480pp.

Vasyliunas, V. M. (1968), A survey of low-energy electrons in the evening sector of the magnetosphere with OGO 1 and OGO 3, *J. Geophys. Res.*, *73*, 2839–2884.

Vasyliunas, V. M., and G. L. Siscoe (1976), On the flux and the energy spectrum of interstellar ions in the solar system, *J. Geophys. Res.*, *81*, 1247–1252.

Verdini, A., and M. Velli (2007), Alfvén waves and turbulence in the solar atmosphere and solar wind, *Astrophys. J.*, *662*, 669–676.

von Steiger, R., and T. H. Zurbuchen (2006), Kinetic properties of heavy solar wind ions from Ulysses-SWICS, *Geophys. Res. Lett.*, *33*, L09103, doi: 10.1029/2005GL024998.

von Steiger, R., N. A. Schwadron, L. A. Fisk, J. Geiss, G. Gloeckler, S. Hefti, B. Wilken, R. R. Wimmer-Schweingruber, and T. H. Zurbuchen (2000), Composition of quasi-stationary solar wind flows from Ulysses/Solar Wind Ion Composition Spectrometer, *J. Geophys. Res.*, *105*, 27,217–27,238.

Waite, J. H., et al. (2004), The Cassini ion and neutral mass spectrometer (INMS) investigation, *Space Sci. Rev.*, *114*, 113–231, doi:10.1007/s11214-004-1408-2.

Weller, C. S., and R. R. Meier (1981), Characteristics of the helium component of the local interstellar medium, *Astrophys. J.*, *246*, 386–393.

Whitted, T. (1980), An improved illumination model for shaded display, *Commun. ACM*, *23*(6), 343–349, doi:10.1145/358876.358882.

Wiens, R. C., P. Bochsler, D. S. Burnett, and R. F. Wimmer-Schweingruber (2004), Solar and solar wind isotopic compositions, *Earth. Planet. Sci. Lett.*, *226*, 549–565, doi:10.1016/j.epsl.2004.07.011.

Wilhelm, K., et al. (1995), SUMER - solar ultraviolet measurements of emitted radiation, *Solar Phys.*, *162*, 189–231.

Wilken, B. (1984), A compact time-of-flight mass-spectrometer with electrostatic mirrors, *Nucl. Instrum. Meth. Phys. Res.*, *222*, 587–600.

Williams, D. J., R. W. McEntire, S. Jaskulek, and B. Wilken (1992), The Galileo energetic particles detector, *Space Sci. Rev.*, *60*, 385–412, doi:10.1007/BF00216863.

Winslow, R. M., B. J. Anderson, C. L. Johnson, J. A. Slavin, H. Korth, M. E. Purucker, D. N. Baker, and S. C. Solomon (2012), Mercury's magnetopause and bow shock from MESSENGER observations, *J. Geophys. Res.*, in prep.

Witte, M., M. Banaszkiewicz, H. Rosenbauer, and D. McMullin (2004), Kinetic parameters of interstellar neutral helium: updated results from the Ulysses/GAS instrument, *Adv. Space Res.*, *34*, 61–65.

Wiza, J. L. (1979), Microchannel plate detectors, *Nucl. Instrum. Meth.*, *162*, 587–601.

Woods, T. N. (2010), Irradiance variations during this solar cycle minimum, in *SOHO-23: Understanding a Peculiar Solar Minimum Conference*, edited by S. R. Cranmer, J. T. Hoeksema, and J. L. Kohl, Conference Series vol. 428, Astronomical Society of the Pacific, San Francisco, CA, pp. 63-71.

Woods, T. N., P. C. Chamberlin, J. W. Harder, R. A. Hock, M. Snow, F. G. Eparvier, J. Fontenla, W. E. McClintock, and E. C. Richard (2009), Solar Irradiance Reference Spectra (SIRS) for the 2008 Whole Heliosphere Interval (WHI), *Geophys. Res. Lett.*, *36*, id. L01101, 1-5.

Wu, F. M., and D. L. Judge (1979), Temperature and flow velocity of the interplanetary gases along solar radii, *Astrophys. J.*, *231*, 594–605.

Wüest, M. (1998), Time-of-flight ion composition measurement technique for space plasmas, in *Measurement Techniques in Space Plasmas: Particles, Geophys. Monog.*, *102*, 141-155.

Wurz, P., J. Whitby, U. Rohner, J. Martín-Fernández, H. Lammer, and C. Kolb (2010), Self-consistent modeling of Mercury's exosphere by sputtering, micrometeorite impact and photon-stimulated desorption, *Planet. Space Sci.*, *58*, 1599–1616, doi:10.1016/j.pss.2010.08.003.

Young, D. T., et al. (2004), Cassini plasma spectrometer investigation, *Space Sci. Rev.*, *114*, 1–112.

Zhao, X., Z. Xiao, and D. Douglas (2010), Mass analysis with islands of stability with linear quadrupoles incorporating higher order multipole fields, *J. Am. Soc. Mass Spectrom.*, *21*, 393–402, doi:10.1016/j.jasms.2009.11.002.

Zhao, X., Z. Xiao, and D. J. Douglas (2009), Overcoming field imperfections of quadrupole mass filters with mass analysis in islands of stability, *Anal. Chem.*, *81*(14), 5806–5811, doi:10.1021/ac900711b.

Ziegler, J. F., M. D. Ziegler, and J. P. Biersack (2010), SRIM - the stopping and range of ions in matter (2010), *Nucl. Instrum. Meth. Phys. Res. B*, *268*, 1818–1823.

Zurbuchen, T. H., et al. (2011), MESSENGER observations of the spatial distribution of planetary ions near Mercury, *Science*, *333*(6051), 1862–1865, doi: 10.1126/science.1211302.

Zurbuchen, T. H., et al. (2008), MESSENGER observations of the composition of Mercury's ionized exosphere and plasma environment, *Science*, *321*, 90–92.

Zurbuchen, T. H., L. A. Fisk, N. A. Schwadron, and G. Gloeckler (2000), Observations of non-thermal properties of heavy ions in the solar wind, in *Acceleration and Transport of Energetic Particles Observed in the Heliosphere*, edited by R. A. Mewaldt, M. Miller, J. R. Jokipii, M. A. Lee, T. H. Zurbuchen, and E. Móbius, Conference Proceedings, vol. 528, pp. 215-220, American Institute of Physics, Melville, N. Y.

Zurbuchen, T. H., G. Gloeckler, J. C. Cain, S. E. Lasley, and W. Shanks (1998), Low-weight plasma instrument to be used in the inner heliosphere, in *Conference on Missions to the Sun II*, ed. by C.M. Korendyke. Proceedings of the Society of Photo-Optical Instrumentation Engineers, Bellingham, Wash., vol. 3442, 1998, pp. 217224.

Zurbuchen, T. H., P. A. Bochsler, and F. Scholze (1995), Reflection of ultraviolet light at 121.6 nm from rough surfaces, *Opt. Engin.*, *34*, 1303–1315, doi: doi:10.1117/12.199865.

Zurbuchen, T. H. (1992), Development of the analyzer system of WIND-MASS, University of Bern, Switzerland, M. S. Thesis.