



LAWRENCE  
LIVERMORE  
NATIONAL  
LABORATORY

# **BioAerosol Mass Spectrometry: Reagentless Detection of Individual Airborne Spores and other Bioagent Particles Based on Laser Desorption/Ionization Mass Spectrometry**

P.T. Steele

September 2004

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

BioAerosol Mass Spectrometry: Reagentless Detection of Individual Airborne Spores  
and other Bioagent Particles Based on Laser Desorption/Ionization Mass Spectrometry

By

PAUL THOMAS STEELE  
B.S. (Harvey Mudd College) 1998  
M.S. (University of California, Davis) 2000

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Applied Science - Engineering

in the




OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

  
\_\_\_\_\_  
  
\_\_\_\_\_  
  
\_\_\_\_\_

Committee in Charge

2004

BioAerosol Mass Spectrometry: Reagentless Detection of Individual Airborne Spores and other Bioagent Particles Based on Laser Desorption/Ionization Mass Spectrometry

**Abstract**

Better devices are needed for the detection of aerosolized biological warfare agents. Advances in the ongoing development of one such device, the BioAerosol Mass Spectrometry (BAMS) system, are described here in detail. The system samples individual, micrometer-sized particles directly from the air and analyzes them in real-time without sample preparation or use of reagents. At the core of the BAMS system is a dual-polarity, single-particle mass spectrometer with a laser based desorption and ionization (DI) system. The mass spectra produced by early proof-of-concept instruments were highly variable and contained limited information to differentiate certain types of similar biological particles. The investigation of this variability and subsequent changes to the DI laser system are described. The modifications have reduced the observed variability and thereby increased the usable information content in the spectra.

These improvements would have little value without software to analyze and identify the mass spectra. Important improvements have been made to the algorithms that initially processed and analyzed the data. Single particles can be identified with an impressive level of accuracy, but to obtain significant reductions in the overall false alarm rate of the BAMS instrument, alarm decisions must be made dynamically on the basis of multiple analyzed particles. A statistical model has been developed to make these decisions and

the resulting performance of a hypothetical BAMS system is quantitatively predicted. The predictions indicate that a BAMS system, with reasonably attainable characteristics, can operate with a very low false alarm rate (orders of magnitude lower than some currently fielded biodetectors) while still being sensitive to small concentrations of biological particles in a large range of environments. Proof-of-concept instruments, incorporating some of the modifications described here, have already performed well in independent testing.

## Acknowledgements

“I looked, and there before me was a pale horse! Its rider was named Death, and Hades was following close behind him. They were given power over a fourth of the earth to kill by sword, famine and plague, and by the wild beasts of the earth.”  
The Apocalypse of John

There ain't no avoiding plagues and diseases (or those other things), but it probably can't hurt to try. At the very least, my haphazard and occasionally non-harmful attempts to help build a better bug detector (to avert “plagues”) have resulted in this heap of what I'll loosely refer to as “information”. This thing wouldn't exist, if it weren't for gobs of help from gobs of people. It might be easier to just thank every sentient entity in existence and then list the one or two who haven't been of assistance, but I'll try to do the reverse, more or less. For all you people whom I'm going to forget to thank, let me say thanks now.

I guess this all got started when I was born. I think my parents deserve a good hunk of the credit (or blame) for that. Thanks for all your help and love and money over all these many, many, many years of school. I suppose I should even thank the sister too. Thanks sis.

I'll spare those of you who have made it this far the details of the next few decades of my existence. Somehow I ended up at LLNL and the good group of guys at the FALCON laser were kind enough to put up with my complete ignorance of anything relating to lasers. Somehow they even kept me from maiming myself too horribly. Thanks guys.

After a big glob of FALCONers fled the lab (or perhaps my detestable presence), I set out to explore the magical world of quantum dots. I've pretty much forgotten everything I ever learned about dots, but the experience I gained in the lab (turning little knobs on

various optical mounts for hours and hours in the dark) has been invaluable. Thanks Howard Lee, Boyd Taylor and Alan Chin.

After decimating that group (my second), I had rather limited contact with non-imaginary humans for a couple of months (I also had very limited contact with sun light). Luckily, Steve Vernon retrieved me from the corner of the equipment depository (otherwise known as my office) and gave me something new to fiddle with. I didn't manage to drive Steve (i.e. group number 3) away, but my mere aura did manage to scare away all the money. Fortunately, Steve managed to find a final resting place for me: the BAMS group (lucky group number 4). Thanks Steve.

For reasons I may never understand, Matthias Frank was actually willing to take me under his wing and haul my carcass towards doctorhood (perhaps Steve bribed him... if so thanks again Steve). Just as mystifying, Eric Gard has never attempted to excise me from the BAMS group like some benign but unsightly, 250-pound tumor. The rest of the group may have been stuck with me, but they've been more than kind too. David Fergenson has been amazingly tolerant of the perversions of code that I've spawned and the various programming atrocities I've committed. Maurice Pitesky has prepared countless spores for me and as far as I can tell hasn't tried to slip me anything too lethal even once. Herb Tobias, Abneesh Srivastava and Keith Coffee have all put up with the continual stream of random questions that seem to spill out of my head and have even provided seemingly reasonable answers on numerous occasions. There ain't enough space to list all the rest of the helpful and nice people in the BAMS group (like Jim, for example) so I'll just hafta lump all of you together and say THANKS!!! I've got to single out Matthias though because I think I've caused him more suffering and trauma than any

one else (although, as of late, Lisa Constable has had to wade through a life-sucking bog of paperwork for me). Thanks Matthias for continually looking out and sticking up for me these past couple years and for all the help with the thesis too. I suspect that this thing may not earn me a Noble prize, but without your help, the people in the admissions department at Davis would probably be shredding all evidence of the fact that they actually admitted me.

Now that I've mentioned school, I've got to thank Richard Freeman for serving as my advisor. Thanks also to the Student Employee program for paying me. Being able to afford luxuries like food has been very, very nice. I like food.

The fact that the seemingly random string of projects I've been involved with actually fit together and prepared me for the experiments described here must constitute absolute proof of a divine power. I've absolutely got to thank El Shaddai El Elyon.

This work was performed under the auspices of the U.S. Department of Energy by University of California Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.



## **Executive Summary**

Bioterrorism is not a new phenomenon, but the probability that a group will try to inflict massive casualties and the probability that they will succeed seem to be growing. Better devices are needed for the detection of airborne biological particles and this thesis describes the ongoing development of one such device at LLNL, the BioAerosol Mass Spectrometry (BAMS) system. This system samples individual, micrometer-sized particles directly from the air and analyzes them in real-time without sample preparation or use of reagents. At the core of the BAMS system is a dual-polarity, single-particle mass spectrometer with a laser based desorption and ionization (DI) system. Mass spectra produced by early proof-of-concept instruments were highly variable, which complicated the differentiation of certain types of similar particles. This thesis describes the investigation of this variability and changes made to the DI laser system that have reduced the variability and thereby increased the usable information content in the spectra. These changes will enable a more sensitive and selective detector.

These improvements would have little value, however, without software to analyze and identify the mass spectra. The initial algorithms developed by the BAMS group often performed surprisingly well, but an understanding of the mass spectra has continually improved and shortcomings of the software have become apparent. This thesis describes significant changes to the algorithms that initially processed and analyzed the data. To ultimately obtain false alarm rates orders of magnitudes less than those of existing detectors, however, an altogether different type of modification is needed. A statistical model has been developed to appropriately make alarm decisions on the basis of multiple analyzed particles. This ensures that the BAMS system will be able to dynamically adapt

itself to its environment to obtain the maximum possible sensitivity while satisfying a particular false alarm rate requirement. A closely related performance model indicates that a BAMS system, with reasonably attainable characteristics, will operate with a very low rate of false alarm (orders of magnitude lower than some currently fielded biodetectors) while still being sensitive to small concentrations of biological particles in a large range of environments. BAMS has not yet reached its full potential, but this thesis will show that significant progress has been made. The first improvements developed in this thesis have already been implemented in several instruments, one of which has taken part and performed well in independent testing.

## **Chapter Overview**

Chapter 1 describes the imposing problem that is faced. Biological “agents” have already demonstrated the ability to kill millions of people. Current detectors are simply inadequate to prevent bioterrorists, or even nature itself, from inflicting massive casualties in the future. The nature of the aerosols that must be dealt with are described, as are existing detectors. This ultimately motivates the choice of a particular type of architecture for the BAMS system.

Chapter 2 describes the basic elements of most of the experiments used to develop and refine BAMS. Biological and non-biological samples are routinely procured and prepared. Aerosols are then generated and analyzed by proof-of-concept instruments. The “proof-of-concept instruments” used for all of the experiments in this thesis are modified commercial Aerosol Time-of-Flight Mass Spectrometers (ATOFMS) and are described in detail.

Chapter 3 discusses the software that analyzes the raw data. The data is loaded, smoothed, calibrated and compressed to be useful for further analysis. Further analysis frequently involves clustering routines that are discussed in detail. A very basic component of any routine that clusters or identifies mass spectra is a metric that quantifies the similarity of mass spectra. A unique new metric is proposed.

Chapter 4 utilizes the software described in chapter 3 to argue that the laser system used for desorption and ionization in the commercial spectrometers causes unnecessary variability in the mass spectra. This is because a non-uniform laser profile allows individual particles to interact with significantly different fluences from shot to shot. Interactions with different fluences cause different amounts of energy to be absorbed and it is shown that differences in absorbed energy are correlated with significant changes in spectral features.

In chapter 5, the basic conclusion reached in chapter 4 is acted upon. The DI system is modified to produce a flattop profile. It produces less spectral variability than the original profile and the improvement is quantified. The flatter profile is used to determine the fluence thresholds required to desorb and ionize significant numbers of ions from several different types of particles. Not surprisingly, particles of any given type are found to have a range of thresholds. This would appear to be a fundamental source of variability that cannot be avoided by modification of the instrument. The determination of this threshold and its range of values for spores, in particular, is believed to be a new and useful measurement.

Chapter 6 describes initial results from new DI lasers with different wavelengths and pulse lengths. The goal is to produce mass spectra with sufficient information content to

easily differentiate closely related particle types. Spectral changes are observed, but the current mass spectrometer prevents, or makes difficult, the observation of new ions with significantly higher masses than those already produced and observed earlier. A model of the ATOFMS instrument is developed that shows this plainly.

The final chapter (chapter 7) examines the potential effectiveness of the BAMS system in real applications. A mathematical performance model is described and implemented in software. Given a desired probability of detection and a maximum false alarm rate, for example, the model determines the range of conditions in which a hypothetical BAMS system can successfully operate. Different scenarios will place different requirements upon the system so the model plays a critical and essential role in determining the tradeoffs between false alarm rate, sensitivity and speed.

# Table of Contents

Chapter 1. Introduction, Motivation and Background .....	1
1.1 History.....	1
1.2 Aerosols .....	6
1.3 Bacterial Spores .....	10
1.4 The Present State of Biological Detection.....	13
1.5 The Appeal of Single Particle Mass Spectrometry.....	20
Chapter 2. Experimental Basics: Samples, Aerosolization and a Mass Spectrometer. ....	22
2.1 Spore Preparation.....	23
2.2 Particle Types Used for Experiments .....	24
2.3 Aerosol Generation.....	25
2.4 The Present Mass Spectrometers .....	28
2.5 The Standard DI Laser and Triggering Scheme .....	34
2.6 Challenges to the Production of a Useful BAMS System.....	37
Chapter 3. Data Processing, Clustering, Identification and Limiting Statistics.....	39
3.1 Initial Processing: Smoothing and Baseline Subtraction.....	40
3.2 Calibration.....	42
3.2.1 Derivation of the BAMS Time-of-Flight Equation .....	43
3.2.2 Standard Calibration Methods and Limitations in an ATOFMS .....	46
3.2.3 Calibration Refinement (a.k.a. Autocalibration).....	49
3.3 Peak Finding and Vector Formation.....	54
3.4 Clustering and Identification.....	57
3.4.1 Structure of the Clustering Algorithm .....	58
3.4.2 Matching Spectra, Updating Neurons and Merging Clusters .....	59
3.4.3 Identification.....	62
3.5 Statistics and Quantification of Mass Spectral Variability.....	63
3.5.1 Statistical Fluctuations in a Single Peak .....	64
3.5.2 Quantification of Variability in a Data Set .....	67
Chapter 4. Spectral Consistency and Laser Power Dependence.....	73
4.1 Desorption/Ionization Laser.....	74
4.2 Data Collection and Initial Processing.....	80
4.3 <i>B.at.</i> Spore Mass Spectra and Their Energy Dependence .....	80
4.4 The Magnitude of the Profile Induced Variations .....	84
4.4 Conclusions .....	90
Chapter 5. Laser Profile Modifications, Fluence Thresholds and Improved Spectral Consistency.....	91
5.1 Flattopping and Characterizing the DI Laser Profile .....	92
5.2 DI Fluence Thresholds for <i>Bacillus</i> Spores .....	97
5.3 DI Fluence Thresholds for BSA, MS2 and DPA.....	109
5.4 Direct Observation of Variability Reduction with Flattop .....	114
5.5 Conclusions .....	122
Chapter 6. Pulse Length and Wavelength Effects on Mass Spectra .....	124
6.1 266 nm Wavelength, ~130 fs, ~6 ns and 40ns Pulse Length Experiments .....	125
6.2 355 nm Wavelength, ~6 ns Pulse Length Experiments .....	133

6.3 3.05 $\mu\text{m}$ Wavelength, $\sim 4$ ns Pulse Length Experiments .....	136
6.4 ATOFMS Performance for High Mass Ions .....	141
6.5 Conclusions .....	145
Chapter 7. Performance Predictions .....	147
7.1 Introduction.....	147
7.2 Probability of False Alarm.....	150
7.3 Probability of Detection.....	153
7.4 Aerosol Properties and Instrument Parameters .....	155
7.5 Performance Predictions .....	159
7.6 Preconcentrators and Prescreeners.....	163
7.7 Tradeoffs between $P_d$ , $P_{fa}$ and $T$ .....	169
Outlook.....	173
Appendix:.....	176
A.1 Smoothing Function.....	177
A.2 Baseline Subtraction Function .....	178
A.3 Autocalibration Function .....	179
A.4 Peak Finding and Vector Formation Function.....	181
A.5 New Baseline and Vector Formation Function.....	183
A.6 The BART Clustering Function.....	186
A.7 Performance Model.....	190
A.8 Poisson Simplification of Eq. 23 for Two Particles Types .....	198
References .....	200

## Chapter 1. Introduction, Motivation and Background

### 1.1 History

Bioterrorism is not a new phenomenon, but the probability that a group will try to inflict massive casualties and the probability that they will succeed seem to be growing. Just as troubling is the fact that nature is quite capable of spreading disease without a terrorist's help. Better detectors are needed to rapidly identify infectious agents irrespective of their source. History provides ample evidence of the misery that unchecked bacteria, viruses and toxins can cause.

In 541 A.D., the Justinian plague began in Pelusium, Egypt. Within 160 years, North Africa, Europe, central and southern Asia, and Arabia had experienced population losses of 50 to 60%<sup>1</sup>. In 1346, the Black Death (Figure 1 and Figure 2) began to spread in Europe. It eventually killed a third of the population. In 1855, the third plague pandemic erupted in China. More than 12 million people in China and India died. The outbreak ultimately spread to every inhabited continent<sup>2</sup>. Some argue that plague may have killed



**Figure 1.** Plague patient with ulcerated, swollen cervical lymph node (bubo). Image obtained from CDC.



**Figure 2.** Plague was known as the "Black Death" because of gangrene that often developed with advanced disease. Image obtained from CDC.

as many as 200 million people throughout recorded history<sup>1</sup>. Although victims of pneumonic plague could spread the disease directly, most cases were caused by the lowly flea<sup>2</sup>. The bite of an infected flea injected thousands of plague bacteria into the skin. The bacteria traveled to the lymph nodes and reproduced, usually causing bubonic plague. Modern medicine has reduced the scale of natural outbreaks, but they continue to occur<sup>2</sup>.

Nature, however, has never been entirely to blame for the spread of plague. In 1346, the trade city of Kaffa was besieged by the forces of Kipchak khan Janibeg<sup>3</sup>. The siege succeeded in eroding the living conditions within the city, but in 1347, the surrounding army, rather than the inhabitants of the city, started to die at a rapid rate. The plague had come. So many soldiers were killed that the siege could not be maintained. Before Janibeg left, he used his catapults to hurl the corpses of plague victims into the city. Whether the infected corpses or other natural vectors were truly responsible for the outbreak in the city is unclear, but the squalid living conditions resulting from the siege ensured that the spread of the disease was rapid and devastating. Four ships that were thought to be “untainted” fled from the city of Kaffa home to Italy. Unfortunately, the ships were not as clean as thought. They carried the plague to Europe.

Plague has killed vast numbers of people in a horrific manner, but even worse pestilences have tormented mankind. The foremost among these was smallpox.

“[Smallpox was] the most terrible of all the ministers of death. The havoc of the plague had been far more rapid; but the plague had visited our shores only once or twice within living memory; and the small pox was always present, filling the churchyards with corpses, tormenting with constant fears all whom it had not yet stricken, leaving on those whose



lives it spared the hideous traces of its power, turning the babe into a changeling at which the mother shuddered, and making the eyes and cheeks of the betrothed maiden objects of horror to the lover.”<sup>4</sup>

Some have claimed that “almost everyone eventually contracted the disease”<sup>5</sup>. It was, without doubt, highly contagious<sup>6</sup>. Even if direct contact was avoided, virions could be inhaled. Lesions would appear and leave extensive and permanent scars if the victims were fortunate enough to survive (Figure 3 and Figure 4). In the unvaccinated, the fatality rate was 30% or more<sup>2</sup> (often much more). Even in recent times the disease inflicted tremendous suffering. The following observations were made in Bombay less than fifty years ago.

“The head was usually covered by what appeared to be a single pustule; the nose and the lips were glued together. When the tightly filled vesicles burst, the pus soaked through the bedsheet, became smeared on the blanket and formed thick, yellowish scabs and crusts on the skin. When



**Figure 3.** Smallpox lesions on the hands of a patient in Accra, Ghana. Image obtained from CDC.



**Figure 4.** A smallpox patient in Teheran, Iran, 1962. Image obtained from CDC.

the pulse was taken tags of skin remained stuck to the fingers... Swallowing was so painful that the patients refused all nourishment and, in spite of agonizing thirst, often refused all fluids... Wails and groans filled the rooms.”<sup>7</sup>

After a systematic program of eradication, the last endemic case of smallpox was reported in Somalia in October 1977. The last known case of smallpox, of any type, originated from a laboratory accident in England in September 1978<sup>7</sup>. The only samples known to remain are maintained by the CDC in the United States and by the Institute for Viral Preparations in Moscow. It is alleged, however, that the Soviets weaponized and produced huge quantities of the virus<sup>8</sup>. It is uncertain how much was destroyed or if unknown sources remain elsewhere.

In comparison to plague and smallpox, anthrax was historically little more than a nuisance. It was primarily a disease of herbivores that grazed in areas where the soil was contaminated with spores. Human infection resulted chiefly from contact with infected animals. Spores deposited in the skin would germinate, reproduce and release toxins<sup>9-11</sup> leading to local edema and an ulcerated sore (Figure 5). Less than one in five cases of this cutaneous infection proved fatal<sup>12</sup>. The more severe inhalational form of infection almost never occurred outside of specialized labor groups such as goat hair mill workers.

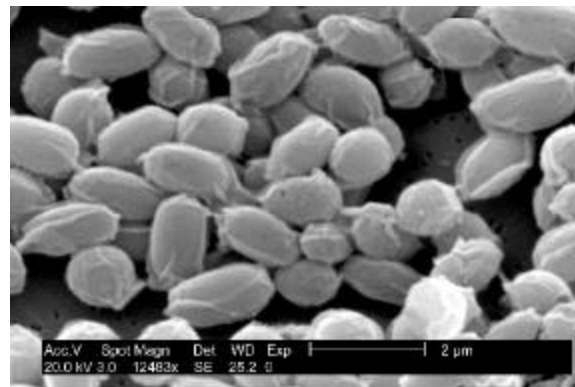
In spite of this, anthrax’s potential as a biological weapon has long been clear. It was easy to obtain and culture, and its spores were incredibly rugged. An airborne release of the spores effectively produced inhalational infection and thus a high fatality rate if medical care was not promptly administered. Germany is alleged to have used anthrax against animals in WWI while Japan is thought to have used anthrax in China in WWII.

The United States, the Soviet Union and Iraq all weaponized anthrax<sup>13</sup>, and other countries are known or are suspected to have worked with it as well. In recent times, however, none of those countries actually used anthrax directly against their enemies (or at least no one detected them doing so).

In September of 2001, something new happened. Five letters filled with *Bacillus anthracis* spores (Figure 6) were sent through the U.S. mail to journalists and politicians<sup>14</sup>. Twenty-two confirmed or suspected cases of anthrax resulted, half of which were inhalational. Five people died. Those responsible have not yet been caught. A troubling observation is that at least some of the letters appear to have contained “weaponized” spores characterized by high concentration, uniform size, and a coating to reduce clumping<sup>13, 15</sup>. A large volume of a similar preparation of spores could easily be released over a major city. A 1993 publication of the Office of Technology Assessment estimated that 100 kg of anthrax released as a line source under ideal conditions could kill 1-3 million people in a densely populated area like Washington, D.C.<sup>16</sup>. In 1997, the CDC estimated that the direct economic cost of such an attack could reach \$26.2 billion



**Figure 5.** A cutaneous anthrax lesion. Anthrax gets its name from the Greek word for coal. Image obtained from CDC.



**Figure 6.** Spores from the Sterne strain of *Bacillus anthracis*. The spores used in the 2001 attacks were of the Ames strain. Image obtained from CDC.

per 100,000 persons exposed<sup>17</sup>.

Interestingly, anthrax itself does not directly kill the infected; it is the toxins that it produces. There are, however, far more potent toxins than those created by anthrax. *Clostridium botulinum* is a bacterium that produces botulinum toxin. If the bacteria are consumed or infect a wound, toxin is released into the bloodstream leading to disease (i.e. botulism). Food-borne botulism has almost certainly occurred since ancient times<sup>18</sup>. Fortunately, the pure toxin is not encountered naturally; it is widely recognized as the most poisonous substance known. The LD<sub>50</sub> for a 70 kg man is thought to be 70 µg orally, 0.70-0.90 µg inhalationally, and only 0.09-0.15 µg intravenously<sup>19</sup>. This is 15,000 times more toxic than VX nerve agent<sup>20</sup>. It is not surprising that various countries have diligently worked to manufacture it. After the Persian Gulf War, Iraq admitted to producing 19,000 L of concentrated botulinum toxin solution<sup>19</sup>. The majority of this had been loaded into weapons with the intent of airborne dispersal.

Given the scale of past outbreaks, known efforts to weaponize biological agents and the readiness of certain groups to inflict massive casualties, it is clear that effective biodetectors are needed. Since an aerosolized release of agent is the most direct means to affect a large number of people, the need for a bioaerosol detector is particularly acute.

## 1.2 Aerosols

This thesis describes the development of a detector for biological aerosols so it is important to understand what an aerosol is and to understand what types of aerosols might be encountered after a bioterrorist attack. Aerosol research is by no means a new field and a number of books have been written on the subject<sup>21-25</sup>. By definition, an

aerosol is a collection of solid or liquid particles suspended in a mixture of gases that has some stability against gravitational settling. Particle sizes may range from nanometers to a hundred micrometers or more, but particles on either the end of the distribution do not generally persist for long.

In practice, it is very useful to group aerosol particles on the basis of their size. Particles less than one or two microns in diameters are called “fine” particles. These are divided into two modes. Particles with a diameter less than or equal to  $0.1\ \mu\text{m}$  form the “nuclei mode”. Particles with diameters between  $0.1\ \mu\text{m}$  and  $2.0\ \mu\text{m}$  form the “accumulation mode”. Larger particles constitute the “coarse mode”. The relative populations of these modes varies significantly from location to location, as does the overall number of aerosol particles. Nonetheless, aerosol particle distributions presented in terms of number, surface area or volume do in fact tend to be modal (hence the above names). One set of experiments reviewed by Higgins<sup>21</sup> studied smog size distributions in Pasadena in 1969. In a purely numerical sense, the vast majority of particles belonged to the nuclei mode. Most of the smog’s surface area, however, resulted from particles in the accumulation mode. The volume (or equivalently mass) distribution was bimodal. The greatest fraction of the volume was contained in the accumulation mode, but a distinct and significant contribution was due to the coarse mode.

Particles in the nuclei mode may be formed directly by combustion processes. They may also be formed by the condensation of gases. These small particles generally do not persist for long because condensation and coagulation cause them to grow rapidly. The growth of particles in the nuclei mode is actually an important source of particles in the accumulation mode, although there are certainly other sources as well. Particles in the

accumulation mode persist for much longer periods of time because they grow relatively slowly and are not rapidly removed by other means. Rain, for example, does not efficiently remove particles less than  $\sim 5 \mu\text{m}$  in diameter (it can be quite efficient, however, for larger particles)<sup>25</sup>. Johnston<sup>26</sup> states that typically encountered fine particles (from both modes) are primarily composed of elemental carbon and condensable organic and inorganic species such as sulfate, nitrate and ammonium.

Coarse particles are usually generated by mechanical means; these include, for example, wind, sea spray, volcanic eruptions and mining. Coarse particles are largely composed of soil, minerals, sea salt and biogenic debris<sup>26</sup>. The largest particles settle rapidly because of gravity. Although coarse particles constitute a significant fraction of the mass of typical aerosols, fine particles have far more surface area and are far more numerous. As a result, fine particles generally have a greater influence on human health and atmospheric processes and have been the subject of more intensive research.

Bioaerosols, as the name implies, have a biological origin. The surfaces of living and dead plants are important natural sources of airborne bacteria and fungal spores<sup>24</sup>. Wind, waves and even rain may aerosolize microorganisms from natural accumulations of water. Bioaerosols occur naturally everywhere from the middle of the ocean to the middle of the arctic. They consist of particles such as bacterial spores and cells, viruses, fungal spores, protozoa, pollen, fragments of insects and skin scales. These include both fine and coarse particles. Viruses are among the smallest bioaerosol particles; some species are only a few tens of nanometers in size. On the other end of the size spectrum, pollen grains can be over  $100 \mu\text{m}$  in diameter. As with other types of aerosols, bioaerosol concentrations vary significantly from location to location. Typical outdoor

concentrations of airborne bacteria range from 100 to 1000 cfu/m<sup>3</sup> (cfu=colony forming units)<sup>25</sup>. Bioaerosol concentrations as high as 10<sup>10</sup> cfu/m<sup>3</sup> may occur in environments such as textile mills<sup>24</sup>.

Biological weapons constitute a subset of the particles that can form biological aerosols. They include pathogenic bacterial spores, vegetative cells, fungal spores, viruses and toxins and frequently contain other materials to stabilize and protect the basic biological agent as well as to prevent clumping and aid dispersal<sup>16, 20, 27, 28</sup>. Even the basic organisms and biological materials by themselves have a broad range of properties. Anthrax (*Bacillus anthracis*), for example, is likely to be dispersed in spore form. (Spores have a diameter of ~1 μm and a number of unique properties discussed in the next section.) Plague (*Yersinia pestis*) is larger (~2 μm) and has only a vegetative state. Smallpox is a 0.3 μm diameter virus, but it is unlikely to be released or detected in the form of single virions<sup>24</sup>. Spores and vegetative cells are amazingly complex structures, but viruses are little more than protein and nucleic acid (with a lipid envelope in some cases). Toxins such as ricin are basically nothing more than protein.

The rapid, reliable and sensitive identification of any type of biological aerosol particle is a daunting task. Many of these particles are “living objects and undergo physical, chemical, and biological changes which are time and space dependent, and are further connected with environmental properties.”<sup>23</sup> To make matters worse, it was already noted that it is unlikely that a biological warfare agent would be released in a pure form. There is abundant information available on how to enhance the airborne dispersal of biological materials with the addition of various chemicals. Data regarding the spraying of microbial pesticides is just one example<sup>29</sup>.

In the case of anthrax, it is known that Van der Waals forces cause unprocessed spores to clump together. Large particles are not deposited efficiently in human lungs and also settle rapidly from the air. Both are undesirable properties if maximal lethality is desired. Silica powders and nanoparticles have long been used to prevent agent particles from coming close enough together for Van der Waals forces to become significant. The U.S. Army has experimented with silica nanoparticles such as WR-50, WR-51, Cab-O-Sil and Sipernat D 13; the Soviet Union used Aerosil while Iraq used both Cab-O-Sil and Aerosil<sup>15</sup>. Bentonite (a clay containing silica particles) can also be used<sup>15</sup>. Military scientists have stated that the “weaponized” anthrax letters sent to Senator Daschle’s office contained silica<sup>15</sup>. In the Senate anthrax letter, there is also evidence that the bond between the silica nanoparticles and spores was further enhanced by the use of a sol-gel or polymerized glass<sup>15</sup>. Some believe that the spores may have even been electrostatically charged<sup>30</sup> to further aid their dispersal. At any rate, the end result of the processing was a powder far more potent than a simple combination of anthrax spores, cells and residual growth medium.

### **1.3 Bacterial Spores**

Many bacteria are highly pathogenic; anthrax is singled out and considered particularly useful as a biological weapon because of its ability to produce spores. Spores are incredibly rugged, dormant forms of bacteria. They are resistant to boiling, freezing, drying, radiation, high pressures, acids and disinfectants. They can be aerosolized, exposed to sunlight and harsh environmental conditions and still remain viable. Anthrax belongs to the genus *Bacillus* and to the family *Bacillaceae*. All six of the genera in the



*Bacillaceae* family can form spores. The “normal” state of these bacteria is vegetative. Vegetative cells have active metabolisms, reproduce and actively carry out a whole array of chemical processes just like “regular” bacterial cells. What makes *Bacillaceae* unique is their response to harsh environmental conditions. Depletion of a critical nutrient, for example, can trigger a remarkable transformation. Through the process of sporulation, the vegetative cell packages DNA and other essential molecules into a compact structure that is initially contained inside the cell wall. This spore (more precisely referred to as an endospore) is ultimately released into the environment when the mother cell lyses and dies. Other types of organisms can produce structures that serve roughly similar purposes, but these structures form on the cell exterior and are known as exospores.

Sporulation has been most thoroughly studied in *Bacillus subtilis*. Historically, the process has been described as lasting eight hours during which the cell progresses through seven morphological stages (generally labeled with Roman numerals). The particular enumeration of the steps is not as important as the fact that the physical and chemical properties of the cell are changing, which may affect the responses of some biodetectors. The BioAerosol Mass Spectrometry (BAMS) group, for example, has observed that the chemical properties (i.e. the mass spectra) of spores continue to evolve well after eight hours have passed<sup>31</sup>. A number of important changes in *B. subtilis* spores during sporulation are well documented, and these merit a brief discussion here because of their relevance to various detection techniques.

In stage IV (i.e. the fourth morphological stage of development), the final ellipsoidal shape of the spore develops as two layers of peptidoglycan are formed<sup>32</sup>. This well-defined shape can be used to differentiate spores from certain types of natural

background particles<sup>33</sup>. (The spore shape and size change slightly, however, with relative humidity<sup>34</sup>.) At the same time, or shortly after the peptidoglycan layers begin to form, proteins and other molecules formed in the cytoplasm of the mother cell<sup>35</sup> are deposited on the surface of the outer membrane of the spore forming a “coat”. This coat becomes apparent at approximately the fifth hour of sporulation<sup>36</sup> (marking stage V).

The coat is important because it varies significantly from species to species. In *B. subtilis*, two layers are apparent. In *B. anthracis* two layers may be present, but they are not easily discernible by electron microscopy<sup>35</sup>. Other *Bacillus* species can have more than two layers. The coats of different *Bacillus* species also have different numbers of polypeptide species. *B. cereus* has one major protein, while *B. subtilis* has more than 25<sup>36</sup>. These differences may serve as an additional means of differentiating closely related *Bacillus* species.

At the same time the coat is being formed, small acid soluble proteins (SASPs) and dipicolinic acid (DPA) are produced and packed into the core of the spore<sup>37</sup>. DPA is an important biomarker that differentiates *Bacillus* spores from most other bacteria. It is utilized in several detectors and is apparent in the data produced by the current BAMS instruments. MALDI experiments have shown that the SASPs can be used to differentiate a number of *Bacillus* species<sup>38</sup>. In stage VI (also known as maturation), the resistant properties of the spore appear, the forespore becomes dormant and the potential for germination is fully realized. In stage VII, the mature spore is released by lysis of the mother cell.

Some *Bacillus* species (e.g. *B. anthracis*, *B. cereus* and *B. thuringiensis*) have an outer covering called an exosporium. Its function is not well understood. In the case of anthrax,

it does not appear to be important to sporulation or germination<sup>39</sup>. Regardless of its particular biological function, however, it may also help differentiate certain *Bacillus* species.

Of all the potential biological warfare agents, *B. anthracis* is one of the most important to analyze, understand and ultimately identify. Because of its lethality, however, it is not particularly convenient to work with. *Bacillus atrophaeus* (*B.at.*) is a close relative of *B. anthracis* that is expected to share many of its features. It is not, however, pathogenic. For these reasons, it has been used frequently by the defense community and is utilized in many of the experiments described in this thesis. *B.at.* was formerly known as *Bacillus subtilis* var. *niger* and earlier as *Bacillus globigii* (*BG*).

#### **1.4 The Present State of Biological Detection**

Natural epidemics have caused millions of deaths. Intentional dispersal of anthrax, plague, smallpox or other biological agents could be devastating. Rapid, sensitive and accurate sensors are needed for the detection of many types of attacks and, in particular, for airborne releases. Clearly one of the greatest challenges is that miniscule amounts of material must be detected in very short periods of time. Consider, for example, that the LD<sub>5</sub> for anthrax (the dose per person that would kill 5% of the exposed population) may be as low as 14 spores<sup>40</sup>. Since an average person inhales approximately 10 liters of air per minute, a one-minute exposure to an aerosol containing just 1 or 2 anthrax spores (i.e. 1 or 2 picograms of material) per liter of air could be fatal for 1 in every 20 people. In a city of moderate size, that could mean thousands of fatalities. The situation is even worse for agents such as tularemia (Table 1). Clearly this makes it very difficult to create an

instrument that can “detect to warn” (i.e. provide sufficient warning for individuals near the instrument to avoid inhalation of an infectious dose of agent). The required response time is obviously scenario dependent, but it is usually considered to be on the order of one minute. This is very difficult, if not impossible, to obtain with many techniques.

Even if the ability to detect to warn proves unobtainable in the near future, the ability to “detect to treat” (i.e. to detect the presence of agent and initiate treatment before symptoms develop) still has significant value. The smallpox vaccine, for example, significantly reduces the probability of mortality even when administered several days after smallpox infection<sup>7</sup>. Anthrax can also be effectively managed so long as treatment is begun before symptoms develop. The same holds true for the treatment of plague<sup>2</sup>. Fortunately, symptoms do not appear immediately for any of these diseases. Even a detector that takes a day or more to respond still has value.

An additional challenge for biodetectors is that natural background aerosols are always present. A detector cannot simply determine whether an aerosol is present or not, it must

CDC Category A Bioterrorism Agents/Diseases	
Agent	Approx. Infectious Dose
Anthrax ( <i>Bacillus anthracis</i> )	8000 CFU (Inhalational LD <sub>50</sub> ) <sup>40</sup>
Botulism ( <i>Clostridium botulinum</i> toxin)	0.7-0.9 µg (Inhalational LD <sub>50</sub> ) <sup>19</sup>
Plague ( <i>Yersinia pestis</i> )	100-500 organisms by aerosol <sup>41</sup>
Smallpox (variola major)	10-100 virions by aerosol (estimated) <sup>41</sup>
Tularemia ( <i>Francisella tularensis</i> )	10 organisms <sup>42</sup>
Viral Hemorrhagic Fevers (e.g. Ebola, Marburg, Lassa...)	a “few” virions (for Marburg in monkeys) <sup>43</sup>

**Table 1.** CDC Category A bioterrorism agents and diseases are believed to pose the greatest potential threat because of their ease of dissemination or transmission from person to person, high mortality rate and potential to cause panic and social disruption.

determine whether a miniscule amount of agent is present in a background that is usually orders of magnitude more concentrated and that may contain a large variety of materials. Components of the biological background may be closely related to certain agents but still remain completely harmless. Other background materials may adhere to or alter any agent that is present. Still other particles may interfere with the detector in unexpected ways. In some locations, true agent organisms are present naturally and this should not be interpreted as evidence of an attack. It is not difficult to understand, therefore, why many techniques have been developed to identify microorganisms in the lab, but few have made it to the field.

Historically, most fielded point detectors have been classified as triggers or identifiers. In general, the purpose of a trigger is only to rapidly indicate that an agent may be present. The trigger makes little if any effort to specifically identify the agent and may frequently “alarm” when no agent is present. A crude trigger might be nothing more than a simple particle counter. The appealing features of triggers are generally that they are fast, relatively inexpensive and potentially reagentless. When a trigger believes that an agent may be present, it triggers an identifier. The purpose of the identifier is to identify the specific agent (if any) that is present. Unlike the trigger, the identifier must have a low false alarm rate. This typically means that more time consuming analyses must be performed and that reagents, or at least some type of sample preparation, must be utilized. Consequently, continuous operation is often impractical (hence the need for a trigger).

A relatively small number of basic techniques (in a variety of slightly different forms) are being investigated as potential triggers and identifiers. A few of the more established

techniques will be mentioned here, starting with potential triggers. Stand-off detection systems will not be described at all.

Various optical techniques can be used as triggers. Laser induced fluorescence, for example, has been used to analyze individual particles very rapidly<sup>44-49</sup>. Spatially resolved light-scattering can reveal information about the size and shape of particles<sup>33, 50</sup>. Such measurements are reagentless and fast. Unfortunately, the information that they provide is insufficient to consistently differentiate organisms at the species level. In fact, it may be difficult to differentiate many particle types at any level.

A specific implementation of a fluorescence-based trigger is the Biological Aerosol Warning System (BAWS). An early version of BAWS uses a frequency quadrupled, Q-switched Nd:YAG laser to excite fluorescence from individual aerosol particles<sup>46</sup>. Photomultiplier tubes (PMTs) detect the fluorescence in two bands: 300-400 nm and 400-600 nm. Based on the PMT signals, rough classification of particles is possible. BAWS's most appealing feature is that it is sensitive to concentrations of 5-10 particles per liter of air and has an alarm time of less than one minute<sup>46</sup>. Newer versions have incorporated a number of improvements, but the false alarm rate is still believed to be high. Another system based on fluorescence detection is the FLuorescent Aerodynamic Particle Sizer (FLAPS). It uses one laser to track and size aerosol particles and a separate laser to excite fluorescence<sup>51</sup>. The excitation is provided by a CW helium-cadmium laser (325 nm), and the fluorescence is detected in a band extending from 420 to 580nm. The original system could detect at least some fluorescence from individual *Bacillus* spores. A newer variation of FLAPS (the UV APS or FLAPS II) has replaced the older excitation source with a pulsed, frequency-tripled, neodymium-based laser<sup>52, 53</sup>. FLAPS II can easily

analyze single spores, but a cross-sensitivity to non-bacterial organic materials has been observed<sup>52</sup>. Other fluorescence instruments exist<sup>54</sup>, but BAWS and FLAPS are perhaps the two most widely used.

Numerous molecular and microbiology techniques have been used as identifiers to detect and fully identify biological particles. In general, their chief limitation is the time required for analysis, which may range from hours to days when sample collection, preparation, and actual analysis are considered<sup>55,56</sup>. Traditional microbiology techniques, such as culturing, are also labor-intensive and limited in that they can detect only living cells. 1% or less of the microbes in a natural sample of dirt or water can typically be cultured<sup>24</sup>. Molecular methods, such as the Polymerase Chain Reaction (PCR), in-situ hybridization, and immunoassays, are extremely sensitive and specific at the species level, or below, but require sample collection and processing as well as specialized reagents.

A number of these techniques have successfully made it to the field. The Biological Aerosol Sentry and Information System (BASIS)<sup>57</sup> was deployed at the 2002 Winter Olympics Games in Salt Lake City, UT. A network of simple aerosol sampling instruments collect and store aerosol samples on dry filters. The filters are collected daily, or more frequently if desired, and taken to a central laboratory. The filters are there processed to release the nucleic acids of any organisms or viruses collected. PCR is then used to determine if pathogens are present. The BioWatch program is patterned after BASIS and is deployed in several cities<sup>58</sup>. The time for detection in both systems is obviously limited by the frequency of filter collection and may be 24 hours or more. The Autonomous Pathogen Detection System (APDS) is much faster<sup>59</sup>. It is a point detector

that integrates sample collection, multiplexed immunoassays and confirmatory PCR as well as sample archiving and data reporting in a fully autonomous unit. Immunoassays can be performed on 30-60 minute intervals and have successfully performed in field studies using live anthrax and plague<sup>59</sup>. This is certainly an impressive feat, but the system still requires various reagents to operate and is too slow to detect to warn in many scenarios.

Traditional mass spectrometry is well suited to the detection of biological agents, in certain respects, due to its high information content and extreme sensitivity to small samples. Unfortunately, most mass spectrometry techniques suffer from relatively long analysis times when sample collection, culturing, and preparation are factored in. Another practical concern is that a vacuum system is required, which makes it somewhat difficult to produce small, rugged and energy efficient instruments. One general approach has been to analyze the fatty acid methyl esters in microorganisms using Pyrolysis Mass Spectrometry (PyMS)<sup>60-63</sup>. Another approach has been to use classical Matrix Assisted Laser Desorption/Ionization (MALDI) to analyze a broad range of biomolecules<sup>64-66</sup>. Analysis times are typically on the order of minutes to hours even when samples are analyzed without culturing. Both approaches require  $\sim 10^3$ - $10^4$  cells for typical analyses, but they do have the potential to provide species level identification of bacteria<sup>67</sup>. Both approaches are sensitive to environmental contamination.

The Block II Chemical Biological Mass Spectrometer (CBMS) is an instrument capable of detecting and identifying both chemical and biological agents<sup>62</sup>. Biological agents are identified using PyMS. Respirable particles are collected and concentrated with an opposed jet virtual impactor. The particles are then directed to a quartz pyrolysis



tube where they are heated and treated with tetramethylammonium hydroxide (TMAH, in methanol). The resulting products are transported to an ion trap mass spectrometer for final analysis. The whole cycle takes about 5 minutes. The speed is very appealing, but since all aerosol particles in the proper size range are collected at once there would appear to be significant potential for background interference. The actual false alarm rate is unknown. Another system worthy of mention is a MALDI-based miniature time-of-flight mass spectrometer developed at the John Hopkins Applied Physics Laboratory<sup>68-71</sup>. The system collects aerosol particles on a VCR tape, deposits a small amount of matrix on top of each sample (containing many collected particles) and then analyzes each sample in a small mass spectrometer. The device is intended to be fast and portable, but little information is available concerning its actual performance. As with the CBMS, many particles are analyzed at once which may cause difficulties.

A BioAerosol Mass Spectrometry (BAMS) system is intended to combine the speed and rough classification abilities of a trigger system like FLAPS with the specificity of a highly specialized mass spectrometer for full identification of individual particles. It does not require any reagents and it has the potential to be very fast. At the very least, the BAMS system will be a trigger with a greatly reduced false alarm rate. The ultimate goal, however, is to detect to warn without the need for a separate identifier (BAMS itself will be the identifier). The specific details of the current BAMS instruments are described in the next chapter, but a general review of single particle mass spectrometry is given in the next section since this is a critical core technology.

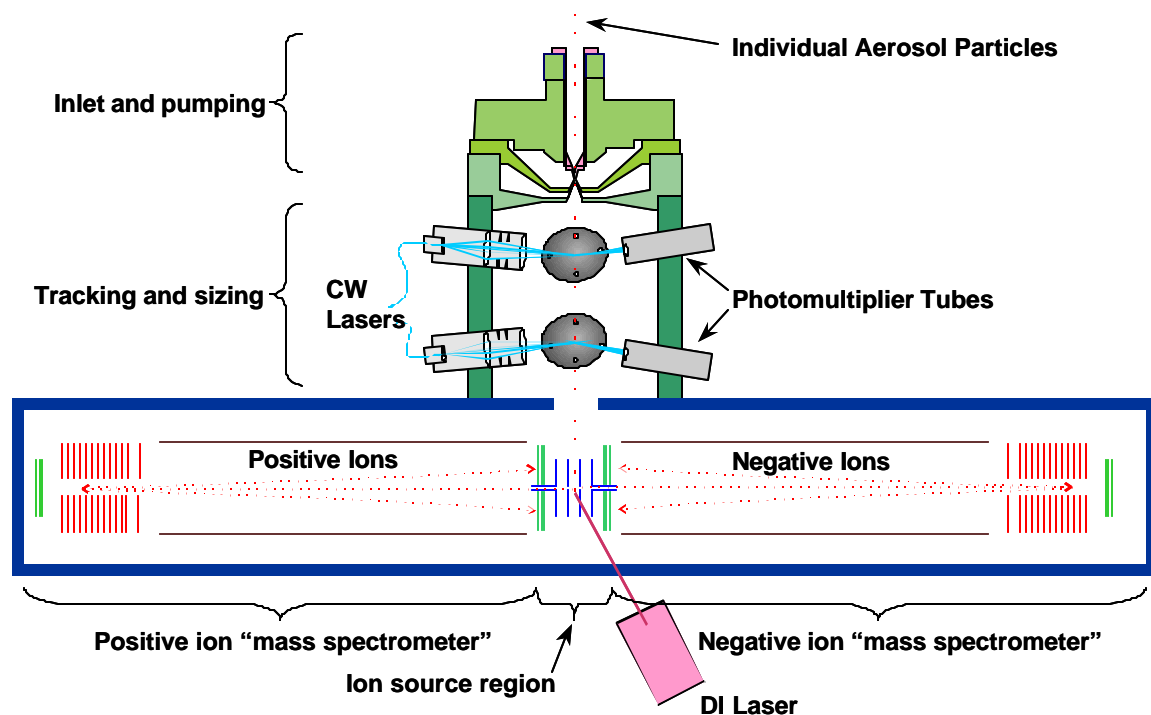
## 1.5 The Appeal of Single Particle Mass Spectrometry

Typical aerosol samples are not likely to contain biological warfare agent particles (in most locations). Nevertheless, there will almost always be significant numbers of harmless background particles. Since the background may interfere with recognition of agent, it is advantageous to analyze or at least screen aerosol particles individually. This means, however, that the detection system must be able to analyze individual particles very rapidly so that any agent particles that are present can be found and identified in a reasonable amount of time. Single-particle mass spectrometry is an established technique for the rapid chemical analysis of individual aerosol particles<sup>72</sup>. With little or no modification these instruments are capable of producing mass spectra from single bioaerosol particles<sup>73-76</sup>. The real challenge is to produce spectra that are sufficiently rich in information content and consistent enough to allow reliable agent identification (i.e. species level differentiation and identification of single particles).

The general technique of single particle aerosol mass spectrometry has been developed and implemented in forms such as RSMS<sup>77</sup>, PALMS<sup>78</sup>, LAMPAS<sup>79, 80</sup>, RTAMS<sup>81</sup>, and ATOFMS<sup>82</sup>, which members of the BAMS group at LLNL helped develop. Figure 7 shows a simple diagram of an ATOFMS instrument with several features common in single particle mass spectrometers. In most modern systems, aerosol particles are sucked directly from the atmosphere into vacuum through some type of inlet<sup>83</sup>. One or more stages of differential pumping enable a high vacuum to be maintained in spite of the constant flow of air into the system. As particles approach the ion source region of the mass spectrometer they cross and scatter light from one or two CW laser beams<sup>84</sup>. The scattered light can be used to determine the particle's size, speed and location. This information is then used to trigger one or two high-intensity pulsed lasers that desorb and

ionize molecules from the particles. The full spectrum of ions can then be measured, at once, in systems using a time-of-flight mass spectrometer. These instruments characterize individual particles in a fraction of a second without reagents or sample preparation so they clearly have some potential as rapid bioaerosol detectors.

New techniques have been developed to coat single aerosol particles “on-the-fly” with a matrix before they enter vacuum<sup>85-87</sup>. This is much faster than traditional MALDI and it still has the potential to produce high mass ions, which is very appealing. Unfortunately, it also requires the use of consumables, which is an undesirable characteristic for a fully autonomous detector. As a result, the current BAMS system does not use MALDI, but this could be changed easily if the use of reagent is acceptable in some particular future application.

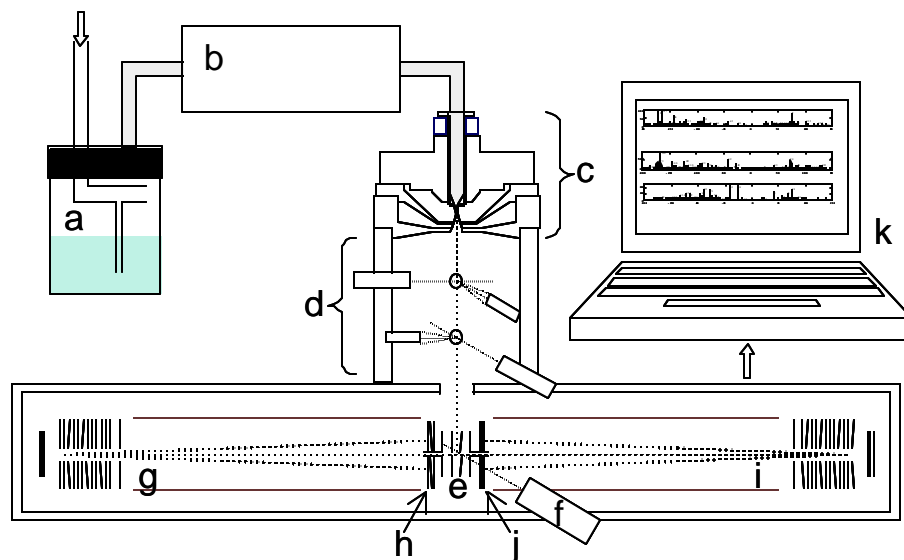


**Figure 7.** A simple diagram of an ATOFMS showing several features common in single particle mass spectrometers. Aerosol particles drawn into an inlet pass through several stages of pumping and are focused into a beam that passes downward through a tracking and sizing region and into the ion source region of a mass spectrometer. The ATOFMS is a dual polarity mass spectrometer that effectively contains two mass spectrometers: one for positive ions and one for negative ions.

## **Chapter 2. Experimental Basics: Samples, Aerosolization and a Mass Spectrometer.**

This chapter covers the basic physical components of virtually every laboratory experiment that the BAMS group performs. Clearly a variety of biological materials must be worked with. Many of these can simply be purchased in a usable form, but bacterial spores and other organisms are grown and prepared in the BAMS group's own facilities. The preparation of bacterial spores is described in section 2.1. The properties of several other particle types (utilized in chapter 5) are described in section 2.2. Artificially generated aerosols are needed to benchmark proof-of-concept systems as well as to better understand the fundamental interactions between biological particles and lasers that enable the systems to work. The methods used to aerosolize samples are discussed in section 2.3. The "proof-of-concept" mass spectrometers used for the experiments in this thesis are modified commercial single particle mass spectrometers. They are described in section 2.4. Details of the desorption/ionization (DI) laser and triggering systems are elaborated upon in section 2.5. Section 2.6 discusses several changes in the commercial instruments that are necessary to produce useful BAMS systems (which will ultimately be quite distinct from the current instruments). The specific changes listed are addressed in subsequent chapters.

The basic layout for a typical experiment is shown in Figure 8. Spores, for example, are aerosolized from solution, dried and sampled into the single particle mass spectrometer ultimately producing mass spectra that are saved to disk. The saved data can be retrieved for instant analysis and on-line identification, or stored and accumulated for later analysis off-line. While the basic physical elements of experiments are described in



**Figure 8.** The laboratory setup used to obtain mass spectra from individual aerosol particles. A nebulizer (a) produces an aerosol of wet particles. A drier (b) removes the water and the dry particles are sucked into the mass spectrometer inlet (c). They pass through a sizing region (d) and arrive at the ion source region (e). A pulsed laser (f) creates ionized molecules. Positive ions (g) are accelerated and then reflected towards a set of microchannel plates (MCP) (h). Negative ions (i) are accelerated in the opposite direction and detected with a separate set of MCPs (j). The digitized MCP signals are finally stored and analyzed on a computer (k).

this chapter, the methods used to analyze the data produced by actual experiments are described in chapter 3.

## 2.1 Spore Preparation

Growing microorganisms in such a way that reproducible results can be obtained is not a trivial matter. The BAMS group includes several microbiologists who prepare the vast majority of the samples used. The basic steps involved in the preparation of spores are given below using *Bacillus atrophaeus* (*B.at.*) as a particular example. As mentioned in chapter 1, *B.at.* was formerly known as *B. globigii* and is frequently used as a surrogate for anthrax (*Bacillus anthracis* or *B.a.*) by the Department of Defense.

*B.at.* cells (ATCC #9372, Dugway Proving Ground, Dugway, UT) are typically grown to mid-log phase in tryptone yeast extract broth ( $\frac{1}{4}\times$ TY) and then aliquoted into 75 ml of

fresh broth in a 1:25 dilution. The cells sporulate in a shaker incubator at 32°C until approximately 90% of the cells are refractile (3-4 days). Phase contrast microscopy and spore staining are used to confirm that spores are present. Gram stains are used to identify vegetative cells and/or non-refractile cells that are still developing into spores. Spores are harvested by centrifugation at 8000 g for 12 minutes, and washed in cold double-distilled water. After three washes the spores are reconstituted in double distilled-water at concentrations of approximately  $10^6$  spores/ml (as determined using a Petroff-Hauser counting chamber). To confirm the purity of the cultures, a 500 base pair region of the 16s rDNA can be sequenced. The same basic process can be adapted for other *Bacillus* species and different growth media.

## 2.2 Particle Types Used for Experiments

*B.at.* spores, prepared as just described, are used in chapter 4. A number of additional types of particles are analyzed in chapter 5. *B.at.* spores prepared in  $\frac{1}{4}\times$ TY are studied as well as *B.at.* spores prepared in resuspension (rs) media, *Bacillus thuringiensis* (*B.t.*) spores grown in resuspension media, clumps of MS2 virions, clumps of bovine serum albumin (BSA) and clumps of dipicolinic acid (DPA). Two different growth media were used to quickly see what types of effects different preparation routines might have on the *B.at.* spore mass spectra. Tryptone yeast extract broth is an undefined medium. Resuspension media is a defined medium. *B.t.* (ATCC No. 16494) is not used as a surrogate for a particular agent but rather as serves as a surrogate for all of the naturally occurring particle types that might easily be misidentified as *B.a.* (or *B.at.* in the current experiments). *B.t.* is in fact frequently used as an insecticide and is likely to be found in “natural” environments.

For the studies here, MS2 serves as a crude simulant for viral agents such as smallpox and Ebola. MS2 is a single stranded RNA bacteriophage that infects *Escherichia coli*. It is approximately 26nm in diameter, contains 3569 base pairs in its genome, and has a capsid containing 180 copies of its single coat protein. (Only three other proteins are encoded by its genome<sup>37</sup>.) It is perhaps unfortunate that the coat protein is structurally unique compared to other RNA viruses<sup>88</sup>. A single virus is far too small to detect so clumps of MS2 virions are aerosolized with a Collison nebulizer (described in section 2.3). The sample used here was obtained from the ATCC (No. 15597-B1) and may contain significant impurities from buffers and salts.

BSA serves as a simulant for toxins such as ricin and botulinum toxin. Albumin is a protein produced in the liver that constitutes a significant fraction of the protein found in blood plasma. One of its several functions is to maintain the pH level of the blood stream. Most forms of albumin contain relatively low amounts of tryptophan and methionine and high amounts of cystine and the charged amino acids<sup>89</sup>. The lack of tryptophan in particular may be relevant since tryptophan is known to absorb well at 266 nm, the wavelength of the “standard” DI laser described in section 2.5. Albumin can be derived from many different animals; bovine serum albumin (BSA) from Sigma (Fraction V,  $\geq 96\%$ ) was used here. It was also aerosolized using a Collison nebulizer.

## 2.3 Aerosol Generation

Although the spore samples can be dried to form a powder, it is usually most convenient to work with them, or other types of samples, in the form of a liquid. Liquid solutions are easily aerosolized using a Collison nebulizer<sup>90, 91</sup>. In the nebulizer, a rapid

flow of air is directed across the top of a narrow tube whose bottom end is submerged in the solution to be aerosolized (Figure 8). The airflow generates a low-pressure region above the solution drawing it towards the top of the tube where direct exposure to the airflow causes small droplets to be formed. Large droplets may also be formed, but these drop out of the aerosol quickly and run back into the solution reservoir. The wet bioaerosols produced by the nebulizer are dried with a diffusion drier containing activated silica gel desiccant and then piped through copper tubing to the aerosol mass spectrometer inlet.

The Collison nebulizer is a convenient means of aerosolizing a solution, but it is known to have several properties that may be undesirable in certain applications. The first of these is that it does not produce a monodisperse aerosol; the range of particle sizes directly produced is actually quite large. For the particular experiments described here, however, this is usually not a problem. When spores are aerosolized, for example, the initial aerosol particles do range significantly in size, but the particles consist mainly of water and most contain either no spores or only a single spore. (Bigger drops containing many spores may be formed, but most of them quickly fall out of the aerosol.) When the water is removed with a drier, single spores are the predominant product and these particles are, by their nature, relatively uniform in size.

In a typical experiment, using a solution of water and *B.at.* spores, the mean aerodynamic diameter of the particles analyzed by the mass spectrometer was determined to be 0.92  $\mu\text{m}$ , which is consistent with the value expected for single spores. The aerodynamic diameter of a particle is equivalent to the diameter of a unit density sphere with the same settling velocity as the particle. Eighty percent of the particles had a



diameter within  $\pm 0.05 \mu\text{m}$  of the mean, but occasionally much larger diameters (2-3  $\mu\text{m}$ ) were observed. These most likely represent clusters of two or more spores. Approximately one in ten particles had a smaller diameter, but few of these are believed to represent intact spores. They may be fragments of spores or vegetative cells or other types of residual impurities.

If a soluble analyte is aerosolized, rather than spores, the resulting dried particles may vary significantly in size. Nonetheless, the aerodynamic diameter of each particle analyzed by the mass spectrometer is measured (as described below) so it is possible to choose a size-selected subset for more careful post-analysis. This is done in chapter 5 for nebulized clumps of DPA. In some situations, even this may not be necessary. The inlets of the current instruments significantly limit the range of particles sizes that can be analyzed irrespective of the size range produced.

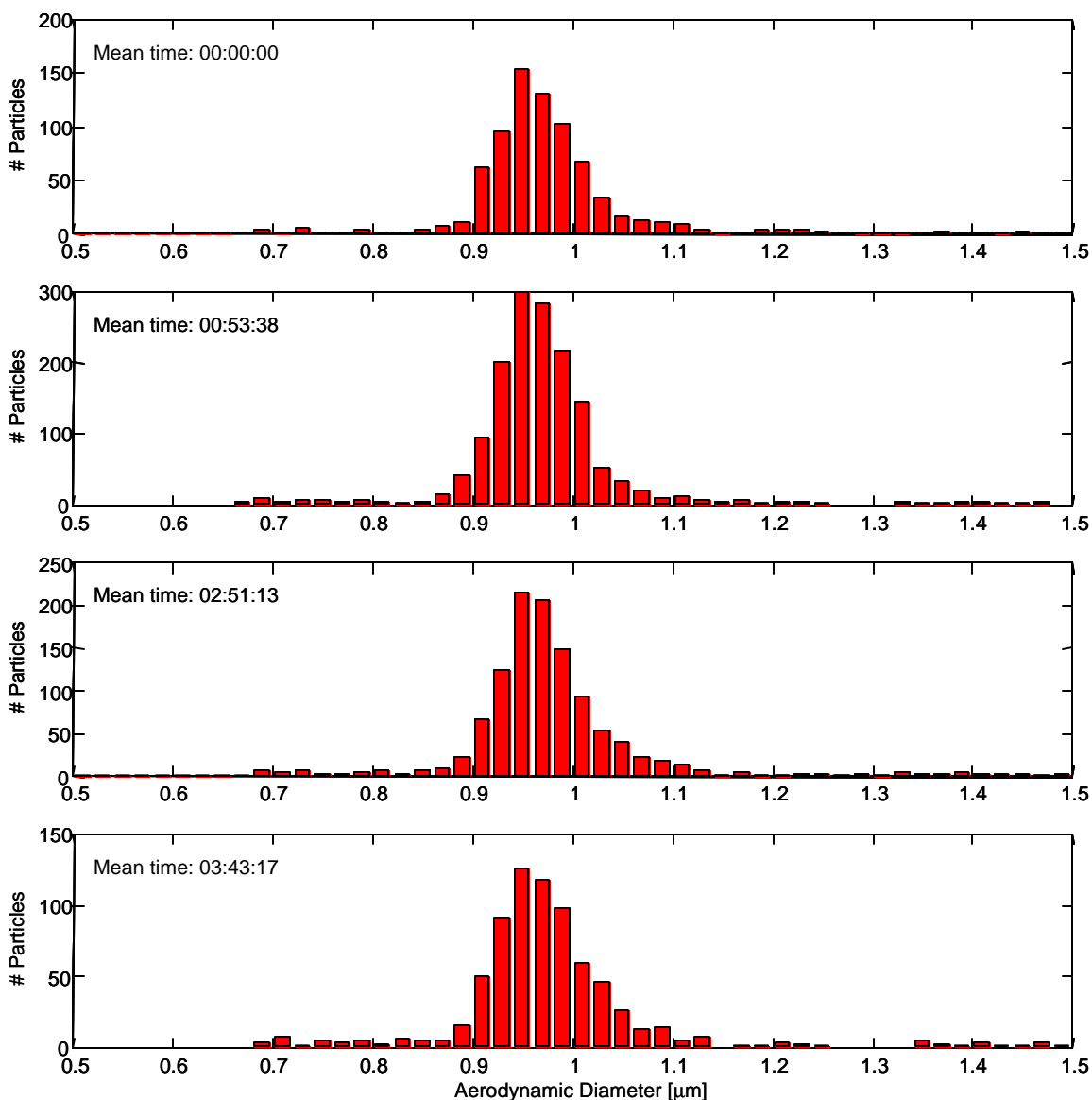
A second limitation of the Collison nebulizer is that the aerosolized biological particles it produces may carry thousands of elementary charges<sup>92</sup>. This is relevant in the current instruments because the particles travel for several centimeters through a high electric field as they approach the center of the mass spectrometer (e in Figure 8). Although the field is theoretically sufficient to push some highly charged particles away from the DI laser, it has been possible to acquire mass spectra at an acceptable rate from many types of particles so the nebulizer is simply used directly in practice. If the charges prove to have a significant effect in some future experiment, it is possible to use a charge neutralizer to correct the problem.

A third potential shortcoming of the Collison nebulizer is that it may damage delicate biological particles<sup>92, 93</sup>. The rapid flow of air necessary to produce small droplets may

produce large shear forces that rupture biological particles or strip off loosely connected outer layers or appendages. If vegetative cells are to be aerosolized, this is cause for great concern. Fortunately, spores are much more rugged. Although spores may not escape entirely undamaged, the size distribution of *B.at.* spores is observed to remain fairly constant over the course of an extended experiment (Figure 9). In such experiments the bulk of the solution is expected to pass through the nebulizer tube many times so if the spores were broken apart the size distribution would become broader with time. If more delicate samples must be aerosolized, a bubbling aerosol generator could be used<sup>94, 95</sup>. It is believed to be more gentle than the Collison nebulizer and is also observed to produce less highly charged particles<sup>92</sup>.

## 2.4 The Present Mass Spectrometers

The two single particle mass spectrometers used in the experiments here were both originally Model 3800 Aerosol Time-of-Flight Mass Spectrometers (ATOFMS) from TSI. These instruments are licensed versions of the ATOFMS developed by Dr. Kimberly Prather's group at UC Riverside. Several members of the BAMS group were once members of the Prather group and had a large role in the development of the original instrument. A detailed description of the Prather group's single particle mass spectrometer already exists<sup>82</sup>, but a description of the closely related TSI spectrometer will be given here for completeness and to point out its unique characteristics. Relevant modifications of the instruments will be described as appropriate. Figure 8 shows many of the features that are described below.



**Figure 9.** Size distributions of *B.at.* spores measured by the single particle mass spectrometer over the course of an experiment. Almost four hours elapse with minimal changes in the size distribution indicating that the spores are not being fragmented significantly.

The spectrometer draws air and entrained aerosol particles from the environment (or from an aerosolization device) through a converging nozzle<sup>26</sup> into vacuum at a rate of ~1 L/min. The diameter of the nozzle is approximately 340 μm. A supersonic expansion from the nozzle into vacuum focuses the aerosol particles into a vertically orientated beam that passes downward through two stages of differential pumping, through a sizing region (which acts as an additional pumping stage) and finally into the ion source region

at the center of the dual polarity mass spectrometer. The supersonic expansion imparts each particle with a velocity dependent on its aerodynamic diameter. Small particles travel faster than larger particles. An aerosol particle that has been properly focused and accelerated crosses two 0.36 mm diameter ( $1/e^2$ ), 532 nm, CW laser beams in the sizing region (at  $\sim 10^{-4}$  Torr) causing two bursts of scattered light that are detected by separate photomultiplier tubes. Ellipsoidal mirrors (which are not obvious in Figure 8) ensure that a large fraction of the scattered light is collected. The times at which the scattered light bursts occur are used to determine the particle's position, velocity and, with proper calibration, its aerodynamic diameter<sup>84, 96</sup>. Once the particle's position and velocity are known, the system predicts when the particle will reach the ion source region (at  $\sim 10^{-7}$  Torr) and triggers the single 266 nm DI laser accordingly (as discussed in section 2.5). The distance between the sizing lasers is  $\sim 6$  cm and the distance from the lower tracking laser to the center of the ion source region is  $\sim 12$  cm.

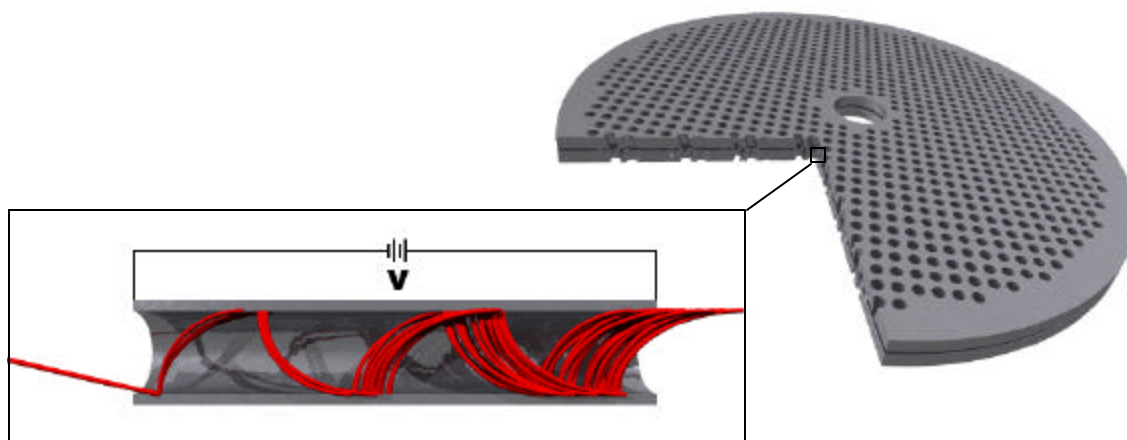
A single pulse from the DI laser both desorbs and ionizes molecules from individual aerosol particles. The ionized molecules are then extracted from the single ion source region into effectively two opposing reflectron time-of-flight mass spectrometers<sup>97, 98</sup>. One spectrometer analyzes positive ions and the other analyzes negative ions. For both polarities, a two-stage, static extraction potential<sup>99</sup> is used such that singly charged ions are imparted with  $\sim 6.4$  keV of energy over a 1 cm path. Since all particles are imparted with the same amount of energy, small particles have higher velocities and cover a fixed distance in less time than larger particles. This is, of course, the well-known basis of time-of-flight mass spectrometry. In such systems, the mass-to-charge ratio of an ion can generally be derived from its flight time using an equation of the form

$$\frac{m}{q} = c(t - t_0)^2.$$

Eq. 1

$m$  is the mass,  $q$  is the charge and  $t$  is the flight time.  $c$  and  $t_0$  are constants determined by the voltages, distances and timing electronics used in the mass spectrometer (chapter 3 provides more details). The distance the ions travel is determined primarily by the length of the drift tube and reflectron, which have a combined length of  $\sim 70$  cm in both polarities (the ions traverse the drift tube twice, however, and penetrate only a finite distance into the reflectron). The ions are ultimately detected with microchannel plates (MCPs)<sup>100</sup> located on either side of the source region.

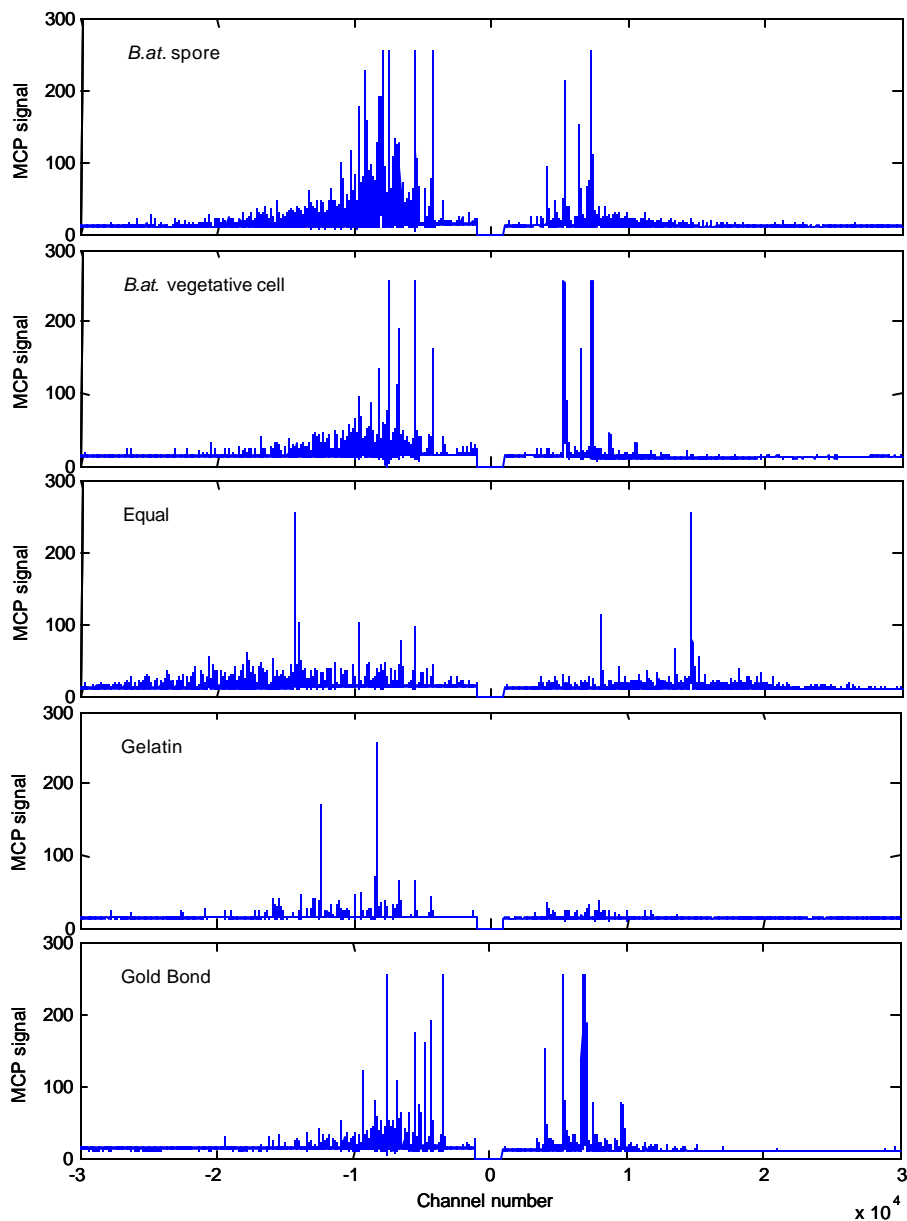
The MCPs are annular as shown in Figure 10. The central aperture allows ions that are created in the source region to pass into the drift region. Upon traversing the drift region, being reflected, and crossing the drift region again some of the ions hit the active region of the MCP. The remaining ions that fail to hit the active area cannot be detected. This is



**Figure 10.** A simple illustration of an annular microchannel plate detector and an enlarged view of one of its channels. Ions pass through the hole in the middle of the detector before they are reflected and return to hit individual channels in the active area. (The relative size of the channels has been greatly enlarged for clarity). The inset shows that an initial ion (on the left) hits the channel wall producing a cascade of ever increasing numbers of electrons. This electron avalanche ultimately produces a detectable burst of charge. Two MCP plates are stacked in a chevron configuration in the current detectors.

a significant problem for certain masses as discussed in chapter 6. Each ion that hits a channel of the MCP has a certain probability of producing one or more secondary electrons that are then multiplied within the channel producing a detectable burst of charge. RC coupling causes the brief current to generate a voltage and this voltage is the quantity that is ultimately digitized and stored. One initial ion produces an average of  $\sim 10^6$  final electrons in the two-plate, chevron detectors employed in the current systems. The response is not perfectly consistent, however, a roughly Gaussian distribution of values is produced with a bias-voltage dependent mean. The width of the distribution (FWHM) is roughly equal to the mean. This distribution does not include saturation effects. If a channel in the MCP has been hit by an ion, that channel (and potentially neighboring channels) will become charge depleted and unable to respond properly to ions that hit later within the acquisition of a single spectrum. This can cause inefficient detection of large ions if numerous smaller ions are produced by the same aerosol particle.

Once triggered (as described in section 2.5), the data acquisition (DAQ) system records the current (or more precisely the voltage) generated by the MCPs every 2 ns for 60  $\mu$ s producing 30,000 data points per MCP (60,000 data points for the full dual polarity mass spectrum, Figure 11). The instrument used in chapter 4, has one 8-bit Acqiris digitizer for each polarity. The instrument used in subsequent chapters has dual 8-bit digitizers (Cougar 1000 from Acqiris) used in parallel with different gain settings to produce  $\sim 12$ -bit data. A small but non-zero DC offset is intentionally added to the baseline when digitizing the data so that the baseline and baseline noise are properly recorded.



**Figure 11.** Raw mass spectral data from individual aerosol particles. The horizontal axis shows the digitizer channel number, which is proportional to the flight time (and not related to the MCP channels). Each digitizer channel is 2 ns wide. The vertical scale shows the MCP signal in arbitrary units (which are proportional to the magnitude of the negative voltages produced and hence the number of ions detected). Data from positive and negative ions were recorded separately and combined for this figure. Negative ions were defined to have negative flight times (i.e. negative channel numbers) for the purposes of plotting.

For each particle, the raw mass spectrum, laser pulse energy, particle size and several other parameters are saved to disk where they can be retrieved for automated real-time identification or stored for later analysis, as described in the next chapter. Although the identification and analysis are performed using software developed by the BAMS group,

the software that actively runs the current mass spectrometers and saves the files is a commercial product (TSI's MS Control) that came installed with the spectrometers.

## 2.5 The Standard DI Laser and Triggering Scheme

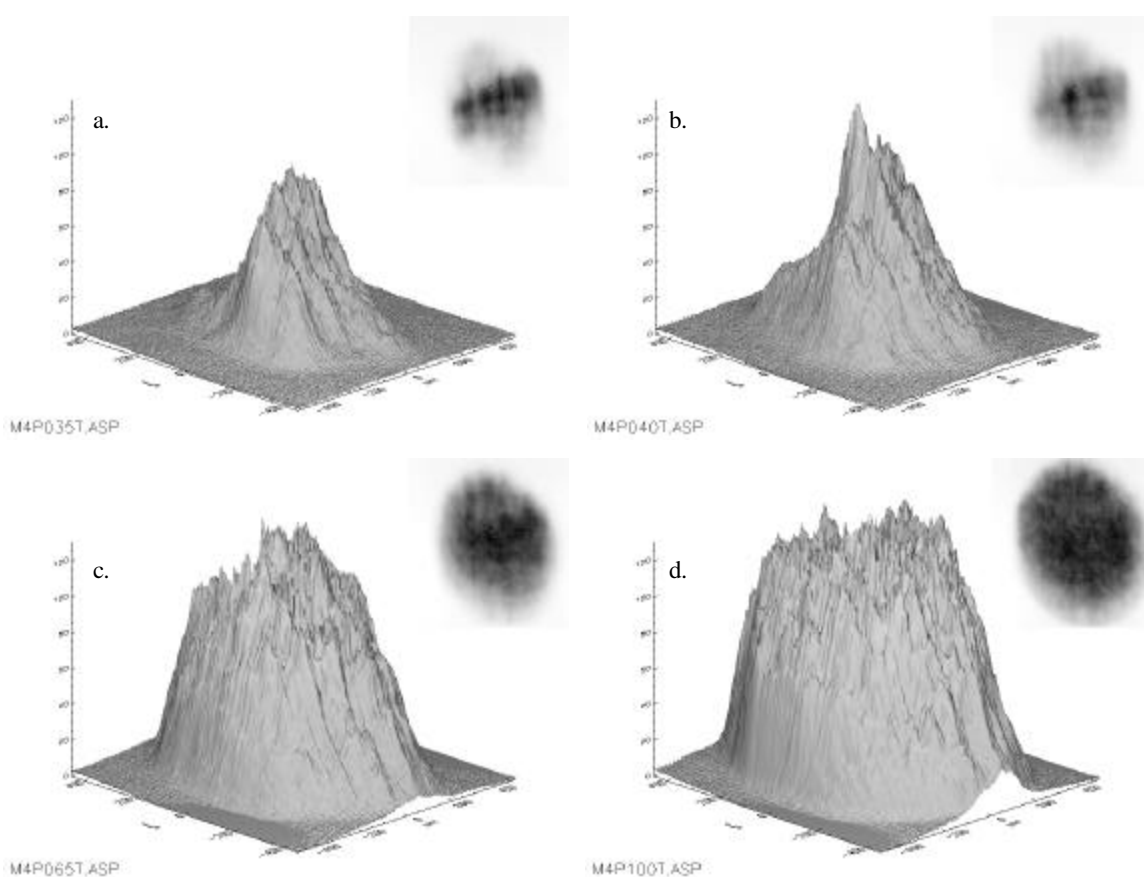
The “standard” DI laser used for many of the measurements reported here was a Big Sky Laser Technologies Ultra with an integrated fourth harmonic package. This is a Q-switched, frequency-quadrupled Nd:YAG laser system that emits pulses with a wavelength of 266 nm, a pulse length of ~6 ns and a roughly Gaussian spatial profile. The 266 nm wavelength is particularly useful since it is nearly coincident with absorption peaks in dipicolinic acid (DPA)<sup>101</sup> and several of the amino acids present in bacterial spores. Typical pulse energies range from 0.2 to 1.0 mJ (at the point where the pulses interact with aerosol particles). Unfortunately, in the default mode of operation set by Big Sky, the Ultra's pulse energy is controlled by adjusting the power delivered to its flash lamp. This is less than optimal because the laser beam profile is observed to change with pulse energy as shown in Figure 12. The effects of the beam profile on mass spectra and efforts to improve the profile will be discussed in chapters 4 and 5. (Chapter 5 actually utilizes a second, nearly identical Ultra with an improved harmonics package.) Other DI lasers will be discussed in chapter 6.

In the standard TSI configuration, the laser pulses are reflected off of two mirrors (to allow full freedom in positioning the beam) and then sent through a lens and window to focus the pulses at a point inside the vacuum system of the mass spectrometer. The pulses then expand to a diameter on the order of 400  $\mu\text{m}$  at which point they interact with the tracked aerosol particles (the laser diameter can easily be changed). The expanding pulses



continue through another window and are incident on a pulse energy meter (a J25LP2 from Molectron) secured to the back of the mass spectrometer.

It is important to discuss the exact scheme used to trigger the standard DI laser and the DAQ system since this is critical for successful operation of the instrument. If the laser is fired too early or too late, it will not hit tracked particles. If the DAQ system is not properly synchronized with the DI laser, the calibration of the mass spectrometer will change; in particular, the  $t_0$  parameter in Eq. 1 will vary. (Calibration is discussed more fully in chapter 3). A complication in this process is that proper triggering of the laser requires not one but two signals (the flash lamp trigger and the Q-switch trigger) and the



**Figure 12.** Images of the 266nm Ultra beam profile at different pulse energies. The laser energy increases monotonically from a. ( $\sim 0.4$  mJ) to d. ( $\sim 2$  mJ). The vertical scale shows the fluence (in arbitrary units) at each point in the base plane. The base of each figure is approximately  $900 \mu\text{m} \times 900 \mu\text{m}$ . The most important features to note are simply that the shape of the profile changes dramatically and that there are a broad range of fluences produced at every setting.

first of these (the flash lamp trigger) is required almost 200  $\mu\text{s}$  before the laser emits light.

Since the distance from the second tracking laser to the middle of the mass spectrometer is twice the distance from the first tracking laser to the second, the time delay between the two scattering events is simply doubled and added to the time of the second scattering event to determine the time when the particle will reach the center of the ion source. The instrument supplies one trigger signal, 185  $\mu\text{s}$  before the particle arrives, to trigger the standard DI laser's flash lamp. The instrument supplies a second signal, when the particle arrives at the middle of the ion source, to trigger the laser's Q-switch. A finite delay exists between the Q-switch trigger and the emission of the laser pulse, but this is small ( $<100$  ns) and is compensated for by a slight vertical translation of the DI laser beam.

The laser itself provides a Q-switch sync pulse used to trigger the data acquisition system. This is the default triggering scheme configured by TSI and was used in chapter 4. Unfortunately, the amount of time between the Q-switch sync and the actual emission of laser light varies with laser energy. This means that the ion creation time varies with respect to the start of data acquisition and thus the mass calibration of the spectrometer changes every time the pulse energy is adjusted.

The second TSI mass spectrometer was used with different DI lasers for subsequent chapters and an improved triggering scheme was implemented. The timing circuit was modified so that the signal previously used to trigger the standard laser's flash lamp occurred even earlier in time (300  $\mu\text{s}$  before particle arrival rather than 185  $\mu\text{s}$ ). This allowed external delay generators to be used to create all three required triggers with

complete freedom. One trigger starts the flashlamp (if the laser employs a flash lamp). The next trigger fires the Q-switch and the final trigger starts the data acquisition system. An oscilloscope and a photodiode are used to synchronize the triggering of the data acquisition system with the emission of the laser pulse (rather than the Q-switch sync). This allows the laser power or the whole DI laser to be changed without altering the mass calibration.

## **2.6 Challenges to the Production of a Useful BAMS System**

An ATOFMS cannot simply be rolled into the field and used as a biodetector. This thesis addresses several of the problems that need to be solved in order for a future BAMS system to successfully operate as a powerful detector of biological aerosols. The desorption and ionization processes used to generate ions must be optimized to produce mass spectra that have greater information content and that are also less variable than the spectra produced by the unaltered commercial systems. In the present context, greater information content basically means that ions with higher masses must be generated (since these are more likely to be characteristic of a particular species than small ions such as sodium). Software algorithms must also be developed and refined to effectively identify particles on the basis of their mass spectra and a few other properties. Ultimately, software is needed that will take the results of many individual particle identifications and determine whether there is sufficient cause to sound an alarm or not. All of these issues are addressed in this thesis. Other challenges exist, but other members of the BAMS group are actively addressing them. The group is already building a new, far more

advanced instrument, based on our own designs, whose performance should far exceed that of any existing single particle mass spectrometer.

## Chapter 3. Data Processing, Clustering, Identification and Limiting Statistics

The BAMS system will rapidly produce huge amounts of data that must be processed. Automated software algorithms are essential. Their ultimate goal is to identify individual aerosol particles on the basis of dual polarity mass spectra and a few related statistics such as particle size. This task might be relatively easy, if particles of a given type (e.g. spores of a particular *Bacillus* species) always produced identical spectra. Unfortunately, this simply does not happen in practice. Identical particles produce only roughly similar mass spectra and significant differences between the spectra may be present. Dealing with this variability is perhaps the biggest single challenge to successful application of the BAMS technique.

The BAMS software more or less supports two basic modes of operation: known particle analysis and unknown particle identification. In the first mode, known samples are aerosolized and sampled by the instrument. (This type of operation includes all of the experiments performed in this thesis.) Some basic initial processing creates mass spectra that can be examined visually or used in more quantitative applications. Frequently, it is advantageous to group similar spectra into “clusters”. Clustering has several applications, but certainly one of the most important is that average spectra from the clusters can be collected and stored to form a library of spectral types. In the second mode of operation unknown particles are sampled by the instrument and identified by comparison to the library. The basic processing of the raw spectra and many of the calculations performed in the two modes of operation are actually very similar to one another.

This chapter describes the various mathematical operations performed on the raw data produced in either of the modes just described. The algorithms and even the instruments have been continually evolving so an assortment of processing and analysis methods have been used throughout this thesis; methods used to analyze the data in chapter 4, for example, were replaced with newer methods for use in chapter 5. At any rate, certain operations are always performed to load the raw spectral data from file, smooth it (if desired), remove the baseline offset (section 3.1) and then optionally refine the time-of-flight calibration so that peaks can be assigned accurate mass-to-charge ratios (as described in section 3.2). The dual polarity mass spectra are then represented in compressed form by pairs of high-dimensional vectors (section 3.3). Once the spectra have been represented in vector form, additional sets of operations can be used to either cluster similar spectra or to identify unknown spectra (section 3.4).

Since the identification and clustering algorithms operate on the vector representations of the spectra, any variability in the vectors is cause for concern. Potential sources of variability are minimized whenever possible, but there is no reason to believe that the mass spectra (or their vector representations) will ever be perfectly consistent. Simple statistical arguments are developed to help quantify and deal with this variability more effectively in section 3.5.

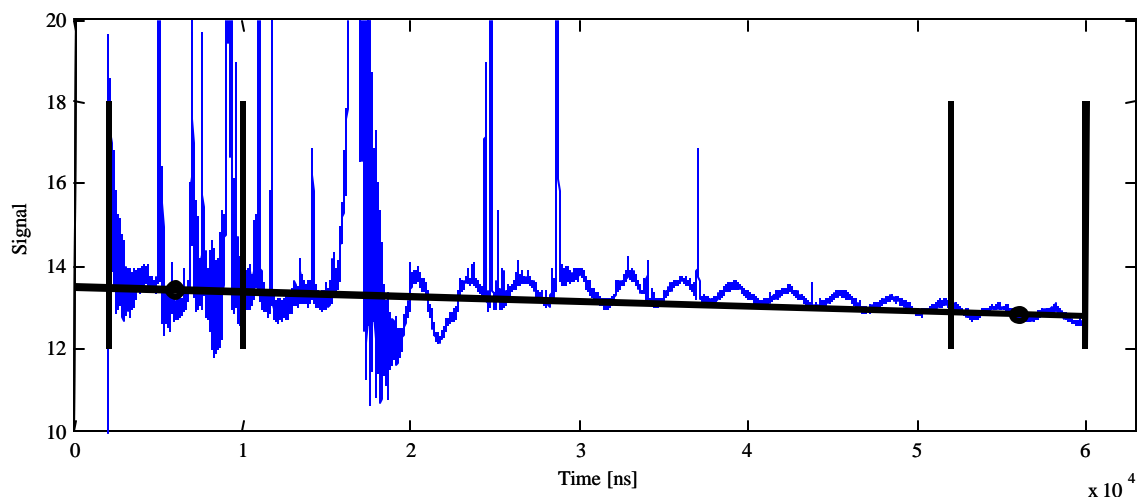
### **3.1 Initial Processing: Smoothing and Baseline Subtraction**

The commercial software that runs the ATOFMS instrument creates a separate data file for each dual polarity mass spectrum collected (as described in the previous chapter). Each file contains a header holding the particle size, measured laser energy and time

stamp among other statistics. The bulk of each file consists of the raw mass spectral data: 30,000 data points representing the positive ion signal at different flight times and 30,000 data points representing the negative ion signal at matching flight times. To be precise, the two sets of data representing the raw “half” spectra are actually just long lists of the MCP voltage readings. The position of a reading in the list indicates the flight time (the first point corresponds to  $t \approx 0$  ns, the second point to  $t \approx 2$  ns and so on). The sum of the voltage readings corresponding to a single peak is proportional, on average, to the number of detected ions that generated the peak. Several examples of raw mass spectra were shown in Figure 11.

The data for the two half spectra obtained from each aerosol particle are treated independently in the initial processing since this is simple and has historically produced useful results. It is clear, however, that some correlations must exist between the two half spectra, which may prove useful in the future. In the first step of the processing, a smoothing algorithm convolves each half spectrum with a narrow Gaussian to reduce some of the point-to-point fluctuations in the data. This is equivalent to a Gaussian frequency filter<sup>102</sup>. To minimize processing time, the Gaussian (in the time domain) contains only five discrete points and has a  $1/e$  full-width of 4 ns (appendix A.1 contains the actual code).

A simple baseline correction is next employed to remove the roughly constant baseline offset. The baseline is not perfectly constant because large ion fluxes may cause ringing and other non-ideal behavior in the detector circuitry (Figure 13). 4000 continuous data points near the start of each half spectrum are selected and sorted by amplitude. The second lowest quarter of the sorted data is averaged and the result is used as the baseline



**Figure 13.** An average of  $\sim 200$  positive raw half spectra is shown to make imperfections in the baseline clear (the baseline correction is always applied to individual spectra in practice however). The two pairs of vertical bars indicate the regions used to calculate two starting baseline values. The thick, nearly horizontal line indicates the baseline estimate for the spectrum. Oscillations in the actual baseline are not desirable, but they are relatively small compared to the mean DC offset ( $\sim 13$  units) and are insignificant compared to the height of the tallest peak ( $>3500$  units).

value at the middle of the 4000-point region. The remaining data in the region is excluded from the average in order to prevent real peaks, chemical noise or detector ringing from unduly influencing the calculated value of the baseline. A second set of 4000 points is selected at the end of the spectrum and another baseline value is obtained. A straight line is then constructed between the two points forming the full baseline estimate, which is then subtracted from the data. In most cases, this does a fairly good job of providing a flat baseline with a mean near zero. A further refinement will be discussed later.

### 3.2 Calibration

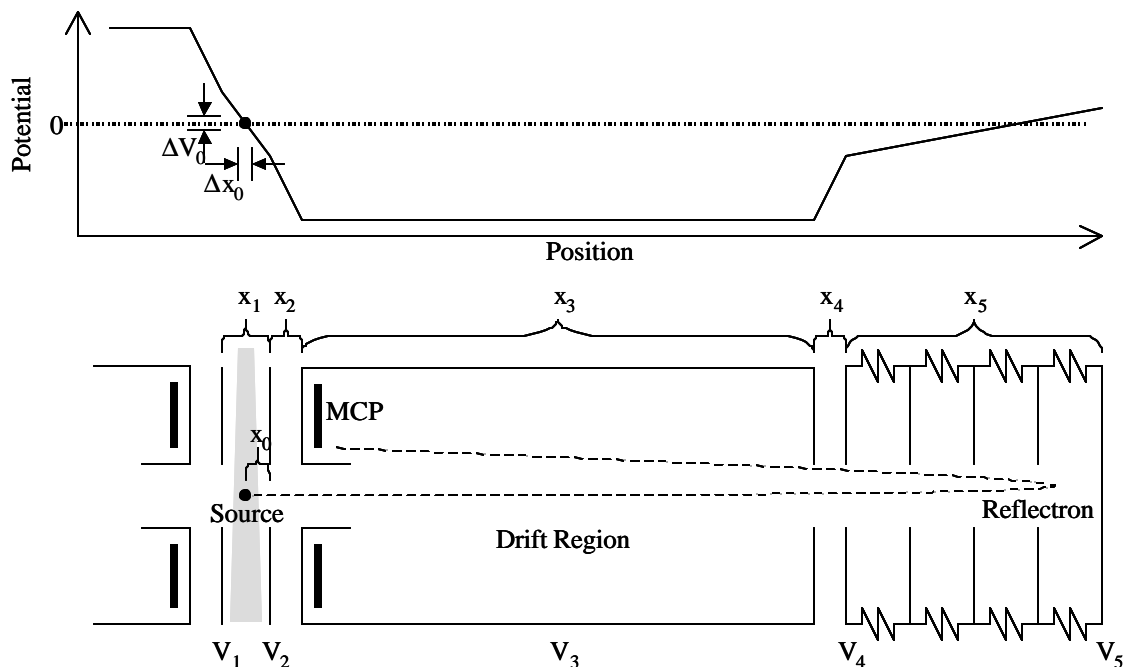
At this point, the spectral data consists of slightly processed voltages measured at a series of flight times. A calibration equation must be obtained so that each of these flight times can be associated with a specific mass-to-charge ratio. An equation for the expected flight time of an ion in the ATOFMS is derived in section 3.2.1. It is simplified to reveal a general form of equation that is actually more useful for calibration purposes. Once this



is done, standard calibration methods are described in section 3.2.2 along with several unique problems encountered in the ATOFMS that require more advanced methods to be applied. In short, each spectrum has a slightly different calibration. This necessitates an “autocalibration” routine such as the one described in section 3.2.3.

### 3.2.1 Derivation of the BAMS Time-of-Flight Equation

The total flight time of an ion in a mass spectrometer, such as that shown in Figure 14, can be predicted very easily if the electric fields are assumed to result from simple linear gradients between applied potentials (i.e. no fringing effects). In Figure 14,  $x_1$  to  $x_5$  represent the lengths of the various regions of the mass spectrometer.  $V_1$  to  $V_5$  represent the applied voltages. Note that the entire drift region is held at a constant potential. The measured flight time of an ion, with mass  $m$  and charge  $q$ , initially located at  $x_0$  with



**Figure 14.** The basic configuration of the ion source, drift tube, reflectron and MCP detector for one polarity of the BAMS system is shown (not to scale). The detectors and part of the drift region for the other polarity are also shown. The gray trapezoid in the source crudely represents the focused stream of aerosol particles. The black dot at its center indicates the region where particles interact with the DI laser. The simple plot at top shows the basic pattern of accelerating potentials.

initial axial velocity  $s_0$  and trigger error  $t_0$  (the time interval between the ion's creation and the moment the digitizer starts acquiring the MCP signals) can be found by sequentially calculating the acceleration in each region, the velocity at the end of each region and finally the time spent in each region.

The accelerations are found using the simple equation  $F = ma = \Delta Vq/x$ .

$$\begin{aligned} a_1 &= \frac{V_1 - V_2}{x_1} \frac{q}{m} \\ a_2 &= \frac{V_2 - V_3}{x_2} \frac{q}{m} \\ a_3 &= 0 \\ a_4 &= \frac{V_3 - V_4}{x_4} \frac{q}{m} \\ a_5 &= \frac{V_4 - V_5}{x_5} \frac{q}{m} \end{aligned}$$

**Eq. 2**

The axial velocity or speed,  $s_f$ , at the end of each region is found using  $s_f = (s_i^2 + 2ax)^{1/2}$  ( $s_i$  is the initial velocity).

$$\begin{aligned} s_1 &= (s_0^2 + 2a_1x_0)^{1/2} \\ s_2 &= (s_1^2 + 2a_2x_2)^{1/2} \\ s_3 &= (s_2^2 + 2a_3x_3)^{1/2} = s_2 \\ s_4 &= (s_3^2 + 2a_4x_4)^{1/2} \\ s_5 &= 0 \end{aligned}$$

**Eq. 3**

Note that the location at which the ion turns in the reflectron does not have to be calculated explicitly. Finally, the flight time in each region can be calculated using

$$t = (s_f - s_i)/a.$$

$$\begin{aligned}
 t_1 &= \frac{s_1 - s_0}{a_1} \\
 t_2 &= \frac{s_2 - s_1}{a_2} \\
 t_3 &= \frac{x_3}{s_3} \\
 t_4 &= \frac{s_4 - s_3}{a_4} \\
 t_5 &= \frac{s_5 - s_4}{a_5}
 \end{aligned}$$

**Eq. 4**

These times correspond to the time spent in each region by the ion on a single crossing ( $t_5$  is the time to required to stop in the reflectron or equivalently the time to reach maximum penetration). The ion crosses several of the regions twice, therefore, the total recorded flight time is

$$T_{tot} = t_0 + t_1 + t_2 + 2t_3 + 2t_4 + 2t_5 .$$

**Eq. 5**

Eq. 5 involves an additional slight approximation because the ion does not quite retrace the entire length of the drift tube before hitting the MCP. Although all of these equations are quite simple, the full expansion of Eq. 5 is quite lengthy. If the initial axial ion velocity is zero, however, it is not difficult to show that the mass-to-charge ratio corresponding to a recorded flight time of  $t$  can be written as

$$\frac{m}{q} = c(t - t_0)^2 .$$

**Eq. 6**

The parameters  $c$  and  $t_0$  are constants determined by the properties of the mass spectrometer. Eq. 6 is perfectly equivalent to Eq. 5 so long as the initial axial velocity of

the ion is zero. If the velocity is non-zero, Eq. 5 cannot be represented so compactly. An approximation can be written as

$$\frac{m}{q} \approx c_1(t-t_0)^2 + c_2(t-t_0)^3.$$

Eq. 7

The parameters  $c_2$ ,  $c_1$ , and  $t_0$  could be written in terms of the voltages and distances used in the derivation of Eq. 5, but Eq. 5 is itself an approximation. The true value of Eq. 6 and Eq. 7 is that they provide a general form of equation that can be fit to a limited set of data (i.e. a few pairs of mass-to-charge ratios and associated flight times) to derive a calibration curve for an entire mass spectrum.

### 3.2.2 Standard Calibration Methods and Limitations in an ATOFMS

To obtain a calibration for most TOF spectrometers, it is usually sufficient to simply generate a mass spectrum from a material with a known (and ideally simple) composition. Since the calibrant material's composition is known, peaks in the resulting raw mass spectrum can be assigned tentative masses by eye with relative ease. Once several flight times have been associated with specific masses, a calibration equation (usually with the form of Eq. 6) is fit to the data and the masses for any of the remaining flight times can be found. In many cases it is even possible to incorporate a calibrant directly in the sample that is being analyzed to provide an accurate internal standard.

Calibration is generally not a topic that merits much discussion, but the dual polarity instruments used here accentuate a few problems that can often be ignored in standard mass spectrometers. The most significant of these problems is that the exact position at which ions are created varies from shot to shot and cannot be easily measured. The cause

of this uncertainty is that the DI laser has a finite width (this is crudely represented by the dot in the middle of the source in Figure 14). One aerosol particle can interact with the left side of the laser beam and the next particle can interact with the right. Since there is a static potential gradient in the ion source region, the time-of-flight for an ion depends on the axial position where it is created. To be perfectly clear, each axial position has a different potential relative to the drift tube so ions that are created from aerosol particles at different axial positions will have different velocities in the drift tube, different total flight times and thus slightly different calibrations. Even slight variations in the calibration are potentially significant. If two identical raw spectra are calibrated such that one or two of the prominent peaks in one of the spectra are offset by one mass unit (or potentially even less) from the matching peaks in the other spectrum, the software will not recognize that the raw spectra were identical or even similar. This can cause particles to be left unidentified, or worse yet, misidentified.

For the moment, Eq. 5 will be applied directly to confirm that the finite DI laser size is in fact a significant source of jitter in the calibration of the current instrument. Assume that the ions have no initial velocity and that the trigger error is zero. The relevant parameters of the mass spectrometer are approximated as follows

$$x_1=1 \text{ cm}, \quad x_2=0.5 \text{ cm}, \quad x_3=60 \text{ cm}, \quad x_4=0.5 \text{ cm}, \quad x_5=14 \text{ cm},$$

$$V_1=2000 \text{ V}, \quad V_2=-3000 \text{ V}, \quad V_3=-7000 \text{ V}, \quad V_4=-2000 \text{ V}, \quad V_5=2000 \text{ V}.$$

The flight times for ions with mass-to-charge ratios of 199 and initial positions spread axially over  $400 \mu\text{m}$  ( $\pm 200 \mu\text{m}$  relative to the middle of the ion source) range from 18.296 to 18.344  $\mu\text{s}$ . Flight times for ions with mass-to-charge ratios of 200 and the same initial positions range from 18.342 to 18.390  $\mu\text{s}$ . The ranges overlap, which is clearly not

desirable. On one laser shot, a particle may generate ions with a mass-to-charge ratio of 199 and a flight time of 18.343  $\mu\text{s}$ . On the very next laser shot, an entirely different type of particle could produce ions with a mass-to-charge ratio of 200 and exactly the same flight time. This will cause significant difficulties in subsequent analyses if the proper masses cannot be determined reliably. The overlap becomes even worse for larger masses.

There are several means to address this problem. Ideally, the incoming aerosol particles should be focused more tightly into the center of the ion source. This is a work in progress (and is not discussed in this thesis). A second option is to reduce the diameter of the laser beam, in order to reduce the width of the region in which aerosol particles are ionized, but this would also reduce the number of particles hit, which is unacceptable. A final option is to calibrate each spectrum individually. This would be relatively easy if certain marker peaks were always present in the mass spectra, but this does not usually occur, even with pure samples. Nonetheless, an “autocalibration” routine has been successfully developed that produces a refined calibration for each individual spectrum. It is described in the next section (3.2.2), but first it is important to mention a final peculiarity of the ATOFMS.

In an ATOFMS, the initial velocities imparted to the ions by the DI event are geometrically unconstrained (there is no sample substrate). This should not be a significant source of calibration jitter, but it will degrade the ultimate resolution of the mass spectrometer. If peaks overlap significantly, they will not be identified accurately in later analysis (as is explained in section 3.3). The unconstrained velocity distribution may also contribute to odd peaks shapes that are occasionally observed and complicate

accurate calibration. The reason for the peak broadening (and potentially related peak shaping) is that individual ions are just as likely to have an initial velocity in the direction of the electric field, as in the opposite direction. For ions with an initial axial velocity antiparallel to their acceleration, the electric field must bring them to a stop (in the axial direction) and then accelerate them back to the plane in which they were created. By the time the ions reach this plane their axial velocity has been fully reversed and they behave just as if they had started with the opposite velocity. A finite amount of time will have passed during this turnaround, however, that cannot be compensated for by the static fields in the rest of the spectrometer. The turnaround time is equal to

$$t_{\text{turn}} = \frac{2s_{0,\text{axial}}}{\left( \frac{v_1 - v_2}{x_1} \frac{q}{m} \right)}$$

Eq. 8

where  $s_{0,\text{axial}}$  is the initial component of the velocity along the axis of the spectrometer. The turn around time can be reduced by increasing the potential across the ion source, but this causes other problems in practice. Other members of the BAMS group have begun implementing a delayed extraction system that should help overcome this limitation. This new system may also enable the measurement of the initial ion velocity distribution so that its effects can be better understood. Unfortunately the system was not available for the experiments reported in this thesis.

### 3.2.3 Calibration Refinement (a.k.a. Autocalibration)

Each individual spectrum requires a potentially unique calibration, but it is too time consuming to calibrate each spectrum manually (i.e. to identify individual peaks in the

raw mass spectra, assign tentative masses based on knowledge or experience and then fit a calibration equation). Fortunately, a nearly equivalent result can be obtained by calibrating the average spectrum from a set of data manually and then using an automated routine to refine that “average” calibration for each individual spectrum. The only necessary assumption in the current autocalibration algorithm is that the majority of the peaks observed in each spectrum should have integer mass-to-charge ratios. Because of the small mass range of the ATOFMS (quantified in chapter 6) and the fact that multiply charge ions are seldom observed, this assumption should be relatively safe. Utilizing the “average” calibration and the aforementioned assumption, the goal of the current routine is to obtain two parameters, a and b, such that

$$\frac{m}{q} \approx (at + b)^2 .$$

**Eq. 9**

This is just a slight variation of Eq. 6.

At the core of the autocalibration routine is a “calibration quality function” that must be maximized to obtain the refined calibration parameters. Its primary arguments consist of a modified subset of the raw mass spectral data and test values for each of the two calibration parameters (a and b). The quality function is designed to have a large value when the correct calibration parameters are input for the spectral data provided and a small value when the incorrect parameters are input. Clearly the real challenge is to properly define the quality function. Once it is defined, the refined calibration parameters are easily found using the “fminsearch” function built into MATLAB v6.5 (which was used to develop all of the code described in this thesis). The code for the autocalibration algorithm and the quality function (named “MassFit” in the code) can be found in the



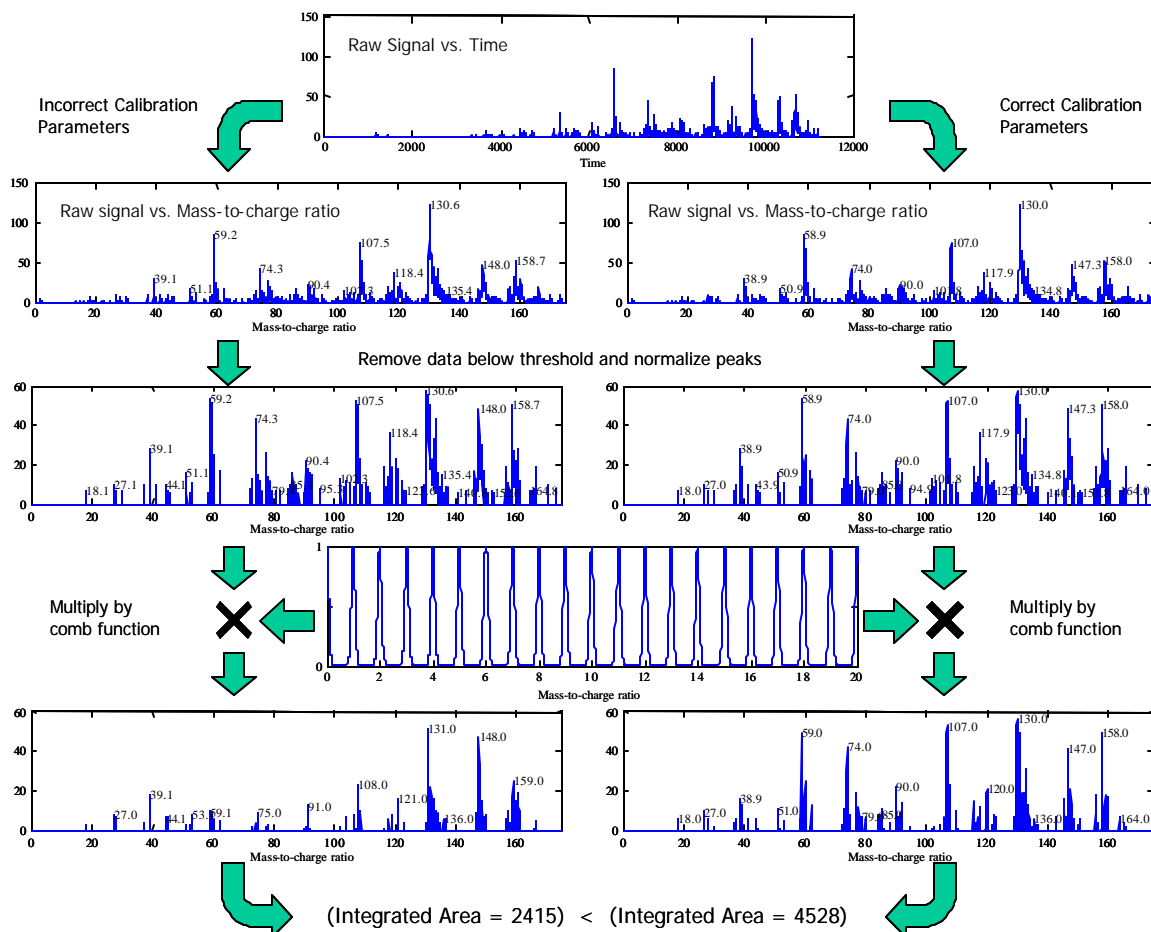
appendix (A.3). The quality function is described shortly, but it is important to first fully understand the inputs passed to the function.

To facilitate rapid execution of the autocalibration routine (and the quality function in particular), regions of the raw spectrum where the ion signal falls below a threshold value are discarded. Regions representing mass-to-charge ratios greater than  $\sim 250$  are also discarded since ions with such masses cannot be expected to have integer mass-to-charge ratios. The square root of the remaining data (i.e. the actual voltage values) is then taken to prevent large peaks from overpowering smaller peaks. The end result of all this is the modified spectral data that is actually passed to the quality function. The “average” calibration parameters are generally passed to the maximization routine as a starting value. The maximization routine itself, however, determines the actual calibration parameter values passed to the quality function.

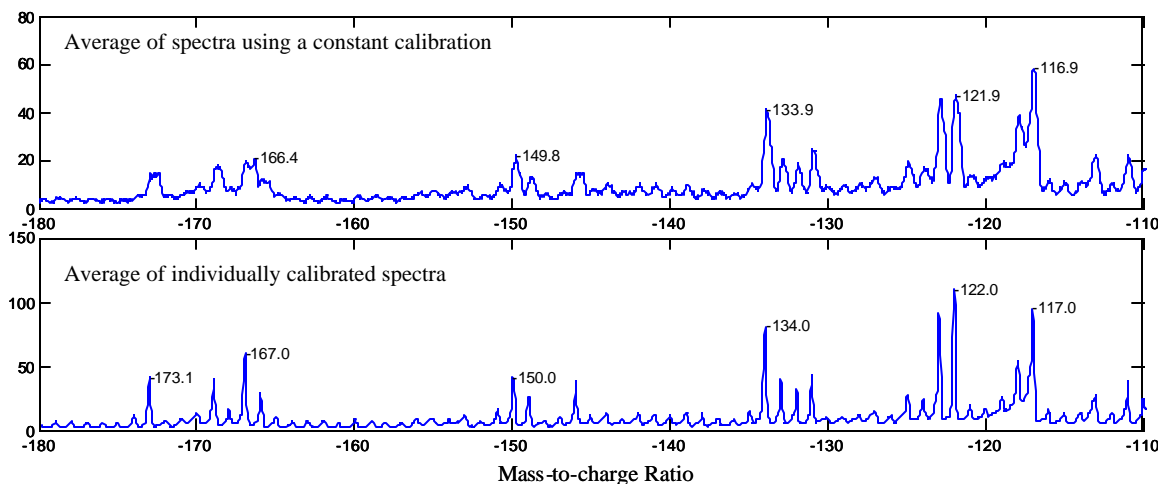
The quality function operates as follows. It takes the modified spectral data and calibrates it using whatever calibration parameters are provided by the maximization routine. This modified bit of mass spectrum is then multiplied by a comb-like artificial spectrum that has a “tooth” at each integer mass-to-charge ratio. (If large peaks are expected at non-integer mass-to-charge ratios, teeth can be added or removed from the comb.) The resulting product is integrated to find its total area. The area is weighted by a few additional factors, which prevent the new calibration parameters from deviating significantly from their average values, and the result is the value returned by the quality function. As already stated, the calibration parameters that maximize this value are assumed to be the correct calibration parameters for the spectrum under consideration. The important operations performed by the function are illustrated in Figure 15.

The reason the autocalibration routine works is fairly simple. Most of the prominent peaks in the spectra will fall on integer mass-to-charge ratios and overlap significantly with the teeth of the comb when the calibration parameters are correct. When the product is integrated, a large value results. When incorrect calibration parameters are used, the peaks are not located at integer mass-to-charge ratios and do not overlap well with the teeth of the comb. The area of the resulting product is reduced (Figure 15).

The resolution of averaged mass spectra is improved significantly by using the autocalibration function. Figure 16 shows average spectra produced with and without the



**Figure 15.** A graphical example showing that the correct calibration parameters produce a larger “integrated area” when applied to a raw spectrum than incorrect calibration parameters. Notice that most of the peaks have near integer mass-to-charge ratios when the correct calibration parameters are first applied. This is not the case when incorrect parameters are applied. Similar operations are performed by the autocalibration routine.



**Figure 16.** Averages of 1000 spectra obtained from single *Bacillus* spores. Note in particular that the peaks near  $m/q=-167$  are much more distinct when the spectra are individually calibrated (using the routine described) before averaging. The peaks are sharper because the jitter in the calibration has been reduced.

autocalibration routine. The peak at  $m/z=-173$  has a width of 0.8 mass units (FWHM) and a height of 10 units when the autocalibration routine is not used. When the routine is used, the peak has a width of 0.2 mass units and a height of 40. This represents an improvement in resolution of 4 and an improvement in the peak height-to-width ratio of 16. This clearly demonstrates that the calibration jitter has been reduced by the autocalibration routine. It should be pointed out, however, that the algorithm is not perfect. The routine can make the calibration worse when it is forced to operate on poor quality spectra that have only a few peaks or large amounts of “chemical noise”. Poor results are also more likely when a poor starting calibration is applied. In future versions of the algorithm, it may be possible to take advantage of certain correlations, which must exist between the positive and negative halves of each dual polarity spectrum, to increase the algorithm’s effectiveness in such cases. It should be clear, however, that the autocalibration routine cannot produce an average spectrum with higher mass resolution than the individual spectra.

### 3.3 Peak Finding and Vector Formation

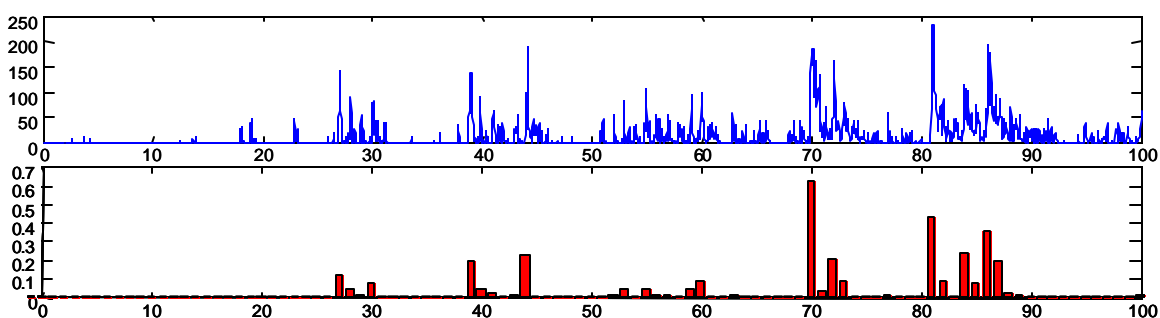
Once the raw spectra have been smoothed, corrected for baseline offsets, and calibrated, a peak finding algorithm is used to determine the location, height and area of each peak in each spectrum and the spectra are finally represented in compressed form as 350-dimensional vectors. Two methods to produce the vectors have been developed and used in this thesis: one requires the peak finding algorithm and the other does not. In general they produce similar results, but each method has certain advantages and shortcomings. The method dependent on the peak finder is used in chapter 4 and is described here first.

The location, height and area of the ion peaks in each spectrum are determined using a relatively simple algorithm (included in appendix A.4). In the case of 8 bit data (where signal values range from 0 to 255 in arbitrary units), any contiguous region of data where the signal is above a threshold value of 5 is considered a single peak. For ~12 bit data, the threshold is set at 50. The height assigned to each peak (i.e. above threshold region) is the maximum signal value in the region. The location assigned to the peak is the mass-to-charge ratio of the maximum signal point (as opposed to, for example, the middle of the peak region or its center of mass). The area of the peak is simply the sum of the data points in the region.

For each aerosol particle that produces a bipolar spectrum, two 350-element vectors are formed (one vector for positive ions and one for negative ions). Generally, it is preferable to characterize spectra by the area of their peaks (rather than their heights) because the area of a peak should be proportional to the number of ions that generated it

(on average). The  $n^{\text{th}}$  element of each vector, therefore, represents the integrated area of the ion peaks near the integer mass-to-charge ratio  $|m/q| = n$  where  $n$  ranges from 1 to 350. Ion peaks with mass-to-charge ratios greater than 350 (which are rare) are ignored, but this can be changed easily.

Some peaks fall between integer mass-to-charge ratios because they represent random chemical noise, multiply charged ions or are simply calibrated imperfectly; the exact cause is often not clear. Peaks with masses greater than a few hundred Daltons may not have integer masses to begin with. Based on the information present in a single mass spectrum, it is also frequently difficult to determine whether a small peak represents meaningful signal or not. Consequently, it is not always clear to which vector element a particular peak area should be assigned (or if it should be assigned to any element at all). The following rule was implemented to consistently associate peaks with one or more vector elements. If an identified peak falls within 0.3 units of an integer mass-to-charge ratio, its area is assigned entirely to the corresponding vector element. If a peak is more than 0.3 units away, its area is split between the two neighboring vector elements. All peaks identified in the proper mass range are included. The exact details of this procedure are apparent in the code in appendix A.4. By splitting noninteger peaks, the effects of any



**Figure 17.** A portion of the mass spectrum from a single spore is shown in the top plot. The bar plot below it shows the first 100 elements of the vector representing that portion of the spectrum. Two vectors are needed to represent the full dual polarity spectrum for the spore and each vector contains 350 elements.

remaining jitter in the calibration should be reduced. A portion of a mass spectrum and the vector elements that represent it are shown in Figure 17.

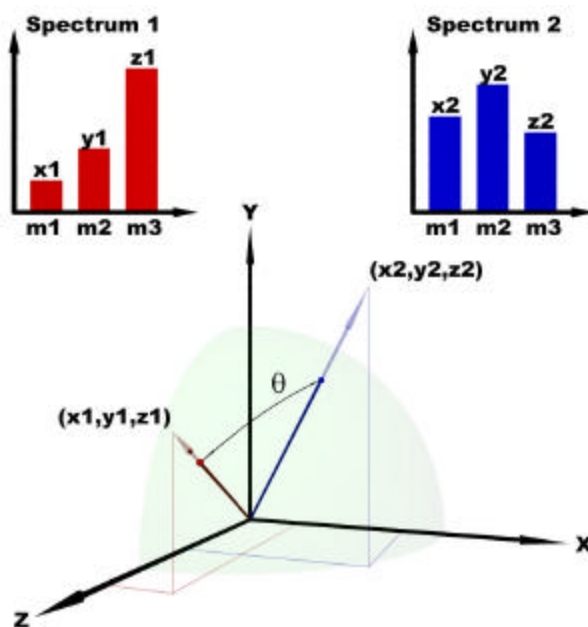
The main theoretical objections to this vector formation method are related to the imperfect peak finder. One of the chief problems with the simple approach used here is that when two large peaks are close together, the signal level between the peaks may not drop below the threshold value and so the two peaks will be identified as one. Obviously, any small peak that fails to cross the threshold will also be missed. In practice, this is not usually a significant problem. On a more practical note, however, the method is computationally intensive.

A second method to produce vectors was developed to be more efficient and does not rely on the peak finder (or a threshold value) at all. (To be precise, the method is more efficient primarily because the peak finder is not used.) It simply sums all of the data near an integer mass-to-charge ratio (whether there is an identified peak there or not) and uses the result as the value of the corresponding vector element. This is the method used in Chapter 5. It is more sensitive to small shifts in the baseline than the previous method so an improved baseline subtraction method was also implemented. The mass spectrum is broken into a number of sections (~10 per mass range of interest per polarity) and the data within each section is sorted by signal amplitude. The mean value of a subset of the sorted data is then treated as the value of the baseline at the center of each section (the value is similar to the median of the data in practice). A linear interpolation between the values is then used to determine the baseline for all the remaining data points of interest. Similar to the previous vector forming routine, data points that fall near half-integer

mass-to-charge ratios are not assigned entirely to one vector element or another, but are shared. The relatively simple code can be found in the appendix (A.5).

### 3.4 Clustering and Identification

Once the individual spectra in a data set have been reduced to vector form, the angle between vectors from different particles can be calculated and used as a quantitative measure of their similarity (Figure 18). Similar spectra produce similar vectors separated by a small angle. Spectra that have no peaks in common produce orthogonal vectors. Using this metric, it is often possible to group a large number of individual particle spectra into a fairly small number of clusters where each cluster contains only particles with relatively similar spectra. This facilitates the visualization of large data sets, as shown in chapter 4. More importantly, however, the ability to group similar spectra is an



**Figure 18.** The positive or negative spectra from two different particles can be represented as vectors in space. In the picture the vectors contain three elements, but in practice they usually contain 350 elements. The angle between vectors is a measure of their similarity. If the angle is small, the vectors (and more importantly the spectra) must be similar.

important step towards the automated identification of particles (section 3.4.3).

### 3.4.1 Structure of the Clustering Algorithm

The clustering algorithm used in this thesis is a refinement of earlier implementations of the ART2a neural network, which were written by other members of the BAMS group<sup>103</sup>. The current algorithm is named BART, which is an acronym for Bunched Adaptive Resonance Theory. A brief description of the overall structure and operation of BART is given in this section. Individual components of the algorithm are described in more detail in the next section (3.4.2).

The goal of BART is to produce clusters of particles with similar mass spectra. Each cluster is represented by a pair of synthetic vectors, which are referred to as neurons (one neuron represents all the positive ion half spectra and one neuron represents all the negative ion half spectra as explained later). BART accomplishes its goal using an iterative routine. Each iteration begins with a list of all the dual polarity mass spectra to be analyzed (with each spectrum represented by two vectors as described in 3.3). The spectrum list is randomly shuffled and then the vectors corresponding to the spectrum at the top of the list are compared with any neuron pairs that exist. If the spectrum's vector pair matches one or more of the neuron pairs, the spectrum is added to the cluster represented by the most similar pair of neurons. The neurons representing that cluster are then updated using a weighted average. (The definition of a match and the exact process of updating are described in section 3.4.2.) If the spectrum can't be added to any of the existing clusters (or if no clusters exist) it becomes the first member of a new cluster. The vector pair representing the second spectrum on the list is next compared to all of the



existing neuron pairs and the process is repeated until all of the spectra have been assigned to clusters.

In the next stage of the iteration, BART compares the clusters just produced to see if any of them are so similar that they should be combined, in which case it combines them. It also erases any empty clusters (i.e. neurons to which no particles have been assigned). At this point one iteration cycle is basically over. BART compares the clusters of particles formed in the current iteration to the clusters formed in the previous iteration (assuming that there was one) and, if identical clusters are found, the routine stops. If different clusters of particles were produced, the neurons are retained, but the actual groups of particles are erased. The master list of spectra is reformed and reshuffled, and the next iteration begins. The iterations continue until identical clusters are obtained or until some maximum number of iterations have been performed (usually ~40).

### **3.4.2 Matching Spectra, Updating Neurons and Merging Clusters**

Clearly it is important to define what conditions constitute a match. When the algorithm compares pairs of vectors (neurons are a special type of vector), the angle between the two positive ion vectors is calculated and then the angle between the two negative ion vectors is calculated. The angle that is larger is used as the “score” for that comparison. If the score is less than a specified value, the particles are considered a match. At the moment, the “specified value” is typically  $\sim 46^\circ$  (the cosine of the angle is referred to as the vigilance factor,  $\cos(46^\circ) \approx 0.7$ ). This particular value has no special significance; it is simply a value that has been used traditionally and that produces reasonable results. If the vector pair representing a spectrum matches the neuron pairs

from several clusters, the spectrum is associated with the cluster that produces the smallest score. Vectors representing positive ions are never compared or combined with vectors representing negative ions at any point in the entire BART algorithm.

When a new cluster is formed by BART, it is represented by the vector pair of the spectrum first put in to it (i.e. the first spectrum's vectors are the first neurons). When subsequent spectra are added to the cluster, the neurons must be updated using a weighted average. The effect of this average is simply to pull the neurons slightly toward the new vectors. Each new vector is multiplied by 0.05 and then added to the appropriate neuron, which has been multiplied by 0.95 (i.e.  $1.00 - 0.05$ ). The weight of 0.05 is referred to as the "learning rate" and was chosen for the same empirical reasons the vigilance factor was chosen above. It should be clear that the neurons are not generally equal to the normal, unweighted averages of the vectors of the particles in a cluster, but they will usually be similar, particularly if the cluster contains many particles.

Once all of the particles have been assigned to clusters, BART checks to see if any of clusters match and, if so, combines them. Clusters are compared one by one. The individual vector pairs in the smaller cluster are compared to the neuron pair of the larger cluster. If all of the vector pairs match the neurons, then the mean vectors from the small cluster are averaged with the neurons of the larger cluster, using the populations of the two clusters as weighting factors, to obtain new neurons for the combined cluster. The details of this procedure should be made clear by examining the code included in the appendix (A.6). These comparisons and combinations are not performed on the final iteration if the maximum number of iterations has been reached.

One of the most significant differences between BART and earlier implementations of the ART2a algorithm is that BART considers both polarities of each spectrum “at once”. Earlier methods would cluster spectra with similar positive spectra (and ignore similarities or differences between negative spectra) or would cluster similar negative spectra (and ignore differences in the positive spectra). Another significant difference is that BART joins similar clusters and deletes empty clusters at the end of each iteration. This step was altogether absent from the earlier ART2a implementations. This meant that once a neuron was formed, it could not be removed regardless of whether it was virtually identical to other neurons or was never associated with any particles. This caused the earlier implementations to produce unnecessarily large numbers of clusters.

BART and its predecessors are useful tools that have significant value in some applications, but they are certainly not perfect. One troublesome property is that they produce somewhat variable results. If the same set of data is reshuffled and fed into any one of the algorithms multiple times, at least slightly different clusters will generally result each time. In many cases this is easy to understand. There are many equally valid ways to group the individual spectra. Nonetheless, distinct groups of spectra may be broken apart while other separate and distinct groups of spectra may be combined. It is possible to exert some influence over these types of events by redefining the vigilance angle, but no angle will be perfect for all sets of data. It is important to note that the current vigilance angle is  $\sim 46^\circ$ . It may not be probable, but it appears theoretically possible, that two orthogonal vectors could be included in the same cluster (both  $\sim 45^\circ$  from the neuron, but separated from each other by  $90^\circ$ ). Clearly, care must be taken not to overestimate the significance of a particular clustering result. Nonetheless clustering

has proven itself to be a useful and powerful qualitative tool for the analysis of large sets of data. Besides that, the single most important application of the clustering routines is building spectral libraries for identification purposes. In this application, most of the shortcomings mentioned are not significant.

### 3.4.3 Identification

Clustering of data involves taking individual spectra and matching them to groups of similar spectra. It is a relatively small leap from clustering to identifying unknown spectra. Spectra from known samples are acquired and clustered to produce a spectral library. Once a library is made, unknown particles can be sampled and compared to the library using the angle between vectors as a measure of similarity just as was done for the clustering. Although improved identification of aerosol particles is a general goal of this thesis, the work described here did not in fact require the use or modification of the actual identification algorithms so they will not be described here.

Although significant and encouraging progress has been made (by other members of the BAMS group) in developing new identification algorithms based on the angular comparison of vectors, it is not clear the angle between vectors is really the best means to determine the similarity of mass spectra. The angle between vectors can be dominated by a few prominent peaks. Differences between small but important peaks in the mass spectra will have almost no effect on the angle even though these differences may have significant value for differentiating similar types of particles (e.g. different, but closely related *Bacillus* species). A new metric is needed to better measure the similarity of mass

spectra. A very closely related metric is derived in the next sections to quantify the variability of the spectra in a particular set of data.

### **3.5 Statistics and Quantification of Mass Spectral Variability**

In MALDI experiments, material is often ablated from a “large” area that may be hundreds of microns in diameter. In addition to this, spectra from many separate laser shots are often averaged. The end result of this is that a peak in an average MALDI spectrum may represent thousands or perhaps even millions of ions. In a BAMS system, on the other hand, single mass spectra generated from single  $1\ \mu\text{m}$  ( $10^{-12}$  g) particles must be analyzed without averaging. A peak in a BAMS mass spectrum may be generated by only a few tens of ions. It is an inevitable consequence that statistical fluctuations will be observed and contribute to the differences seen between spectra from “identical” particles. The fact that many peaks are present in a given spectrum makes it likely that the fluctuations of at least a few of these peaks will be significant. These fluctuations ultimately limit the consistency of the data that can be obtained and thus the ability of the system to differentiate very closely related particle types. The magnitude of these variations must be quantified. The following derivations explicitly account for some of the fundamentally statistical processes that govern the reproducibility of mass spectra. The metric ultimately obtained here is applied to quantify the variability of real data in chapter 5.

The basic calculations proceed as follows. The magnitude of the fluctuations in the area of a single peak is calculated in section 3.5.1. In section 3.5.2, the result from section 3.5.1 is applied so that whole spectra (with many peaks) or even sets of spectra can be

dealt with. In particular, the expected “distance” between an individual spectrum and its true mean value is considered. A new normalization scheme must be (and is) implemented to make this basic concept more useful. The scheme allows peak of vastly different sizes to be more fairly compared. Ultimately, the variability of an entire data set is quantified with a single number.

### 3.5.1 Statistical Fluctuations in a Single Peak

Assume that a certain type of particle produces spectra where an average of  $N$  ions are detected at a particular mass-to-charge ratio with a standard deviation of  $s_N$ .  $N$  includes only the ions detected and does not include ions that missed the detector, hit charge-depleted detector channels (in the case of an MCP) or were otherwise lost. The probability to detect exactly  $n$  ions given an average of  $N$  is denoted as  $P_M(n)$  (no particular functional form is assumed). Notice that these parameters and this distribution entail all of the physics of the desorption and ionization event and all of efficiencies (or inefficiencies) of the mass spectrometer up to the point of ion impact onto the detector.  $N$  describes the number of ions detected, but not the actual response of the detector. Assume that the response of the detector to a single ion has a distribution of signal values with a mean of  $r$  and a standard deviation of  $s_r$ . This does not include ions that hit the detector but fail to produce a detectable signal because of saturation or other factors. In the case of the MCP detectors in the ATOFMS system,  $s_r \gg 0.4r$ , but the exact value of  $r$  (i.e. the mean voltage pulse area for a single ion impact) is currently unknown in terms of the units in which the raw data is acquired.

The mean signal observed (i.e. the average peak area) will simply be  $Nr$  – assuming that  $r$  is independent of  $N$ . The standard deviation of the observed signal,  $\mathbf{s}_{Nr}$ , must be derived more carefully. The first step is to calculate the probability to obtain an observed signal within  $dy$  of a particular value  $y$  given the average of  $Nr$ . Once this probability is known, the formal definition of the standard deviation (i.e. the root-mean-square deviation from the mean) is applied to derive a more useful result.

Let  $P_n(y)dy$  denote the probability to observe a peak height within  $dy$  of  $y$ , given that exactly  $n$  ions are detected (not an average of  $N$ ). This distribution has a mean value of  $nr$  and a standard deviation of  $n^{1/2}\mathbf{s}_r$ . (This assumes that  $n$  is independent of  $r$ , which should be relatively accurate as long as the MCP detector is not heavily saturated.) The last result follows from the fact that the sum of a series of independent random variables is a new random variable with a mean equal to the sum of the means and a standard deviation equal to the sum in quadrature of the standard deviations.

Since the number of detected ions is not actually fixed,  $P_n(y)dy$  must be summed over all the possible values of  $n$  weighted by the probability to obtain  $n$ . The resulting probability to observe a peak signal within  $dy$  of  $y$  given an average signal of  $Nr$  is

$$P_{Nr}(y)dy = \sum_{n=0}^{\infty} P_N(n)P_n(y)dy.$$

**Eq. 10**

Applying the definition of the standard deviation leads to

$$\mathbf{s}_{Nr}^2 = \int_0^{\infty} (y - Nr)^2 P_{Nr}(y)dy = \int_0^{\infty} (y - Nr)^2 \left( \sum_{n=0}^{\infty} P_N(n)P_n(y) \right) dy.$$

**Eq. 11**

The integral can be pulled inside the summation while the  $(y-Nr)^2$  term is expanded, added to zero (i.e.  $-2nry+n^2r^2+2nry-n^2r^2$ ) and rearranged to obtain

$$\mathbf{s}_{Nr}^2 = \sum_{n=0}^{\infty} P_N(n) \int_0^{\infty} \left( (y-nr)^2 + 2nry - n^2r^2 - 2Nry + N^2r^2 \right) P_n(y) dy.$$

**Eq. 12**

When integrated, the first term in the integral results simply in the variance of  $P_n(y)$ . The second and fourth terms produce the mean multiplied by a constant. The third and fifth terms are just constants unaffected by the integration. The mean and variance of  $P_n(y)$  are already known so the integrated result can be written as

$$\mathbf{s}_{Nr}^2 = \sum_{n=0}^{\infty} P_N(n) \left( n\mathbf{s}_r^2 + (N-n)^2 r^2 \right).$$

**Eq. 13**

The mean and standard deviation of  $P_N(n)$  are known as well so the summation can be easily evaluated

$$\mathbf{s}_{Nr}^2 = \mathbf{s}_r^2 \left( \sum_{n=0}^{\infty} P_N(n)n \right) + r^2 \left( \sum_{n=0}^{\infty} P_N(n)(N-n)^2 \right) = N\mathbf{s}_r^2 + r^2\mathbf{s}_N^2, \text{ or}$$

$$\mathbf{s}_{Nr} = \sqrt{(N^{1/2}\mathbf{s}_r)^2 + (r\mathbf{s}_N)^2}.$$

**Eq. 14**

This result for the standard deviation of the observed signal,  $\mathbf{s}_{Nr}$ , is simply the sum in quadrature of the standard deviation of the signal if the number of ions detected was perfectly fixed ( $\mathbf{s}_N=0$ ), but the detector response varied, and the standard deviation of the signal if the detector were perfect ( $\mathbf{s}_r=0$ ), but the number of ions varied (assuming that  $N$  and  $r$  are independent). This result is not particularly surprising, but it is important to note that it was obtained without making any assumptions about the particular forms of  $P_N(n)$



and  $P_n(y)dy$ . The two distributions do not have to be Gaussian (or of any other standard type). To quickly summarize, a peak generated by an average of  $N$  detected ions will have a mean value of  $Nr$  (where  $r$  is the mean single ion detector response) and a standard deviation of  $\mathbf{s}_{Nr}$  given by Eq. 14. With this information in hand, a metric for the calculation of the variability of a data set can now be developed.

### 3.5.2 Quantification of Variability in a Data Set

From the earlier discussions it should be clear that it is useful to represent or think of mass spectra as vectors in space. For the purposes of clustering, it has historically proven useful to measure the similarity of spectra by calculating the angle between the vectors representing the spectra<sup>104</sup>. This measure, however, is less than ideal. In practice, differences between small peaks in the spectra are often very important for differentiating similar, but distinct, particle types. Unfortunately, these differences may have little effect on the angle. The angle can be dominated by a few large peaks that are shared by the similar particle types. A new metric is needed that appropriately weights the contributions from all of the potentially relevant peaks.

It is more convenient mathematically to consider a distance, rather than an angle between vectors. A raw distance measurement, however, suffers from many of the same limitations as the angle. Fortunately, this can largely be fixed by applying a new normalization scheme. A few calculations relevant to the determination of distances will be described next and then the normalization will be described afterward.

In order to calculate the mean-squared distance between an individual vector and its mean value (i.e. the mean vector for that particle type), the mean-squared difference

between the area of a peak in a given spectrum and the area of that same peak in the mean spectrum is calculated next. Let the difference between a peak area and its mean be described by  $P_d(y)$ .  $P_d(y)$  has a mean value of zero and a standard deviation of  $\mathbf{s}_d = \mathbf{s}_{Nr}$  (the value of  $\mathbf{s}_{Nr}$  was derived in the last section and given in Eq. 14). The mean value of the squared difference is thus

$$\overline{d^2} = \int_{-\infty}^{\infty} y^2 P_d(y) dy = \mathbf{s}_d^2 = (N^{1/2} \mathbf{s}_r)^2 + (r \mathbf{s}_N)^2.$$

**Eq. 15**

The integral is equivalent to the variance of  $P_d(y)$  since the mean of  $P_d(y)$  is known to be zero. The important point to notice is simply that the mean squared difference depends upon the mean number of detected ions,  $N$ .

The mean squared distance between an individual vector and its corresponding mean vector can be (and is) now determined. The mean squared distance is simply equal to the sum of a series of terms like Eq. 15, where each term represents a different vector element (this is just the Pythagorean theorem,  $D^2 = \Delta x^2 + \Delta y^2 + \Delta z^2 + \dots$ ). As already mentioned, this distance metric isn't much better than the corresponding angle at this stage. Big peaks still dominate small peaks because Eq. 15 depends on  $N$  ( $\sigma_N$  is likely a monotonically increasing function of  $N$ ). A method is needed to normalize the data so that equation Eq. 15 no longer depends so strongly on the number of detected ions.

To determine the appropriate normalization scheme an idealized system is considered. In such a system, it may not be unreasonable to expect a binomial distribution of detected ions. If each particle contained the same number of parent molecules, each one of the molecules had the same probability of being ionized and detected, and the probability of detecting any given molecule was independent of the probabilities for all other

molecules, the number of detected ions would in fact be described perfectly by a binomial distribution. Since the number of ions detected is almost certainly a very small fraction of the total number of molecules, even in an idealized case, it should be safe to approximate the binomial distribution with a Poisson distribution. In this limit,  $\mathbf{s}_N = N^{1/2}$  and Eq. 15 can be rewritten as

$$\overline{d^2} = Nr \left( \frac{\mathbf{s}_r^2 + r^2}{r} \right).$$

**Eq. 16**

$Nr$  is simply the mean peak signal (i.e. area). A potential normalization scheme is now apparent. When a set of data is collected, the mean value of any peak signal can easily be found. The value of each peak signal in each individual spectrum will simply be divided by the square root of its average value (or more precisely, each element of each vector will be divided by the square root of its mean value). It is not difficult to show that after this normalization is employed, the mean squared difference between a peak and its mean value will be

$$\overline{d^2} = \left( \frac{\mathbf{s}_r^2 + r^2}{r} \right).$$

**Eq. 17**

As was desired, this value no longer depends on  $N$ . If the vectors under consideration contain  $m$  elements, the mean squared distance between an individual spectrum and the mean spectrum will simply be

$$\overline{D^2} = m \left( \frac{\mathbf{s}_r^2 + r^2}{r} \right).$$

**Eq. 18**

The root mean square distance between a normalized spectrum and its mean value will simply be the square root of Eq. 18,

$$D_{rms} = \sqrt{m \left( \frac{\mathbf{s}_r^2 + r^2}{r} \right)}.$$

**Eq. 19**

This is independent of the value of  $N$  (for any vector element) and is thus the same for any type of particle that behaves in a manner consistent with the assumptions made. This is a very appealing property. It doesn't matter what type of particle is looked at (so long as only one particle type is examined at a time), or what laser power is used. If the assumptions remain valid, the same RMS distance will be obtained. In reality, most particles will probably not produce a Poissonian distribution of detected ions. It is likely, for example, that the probability to ionize and detect a specific ion is correlated with the probabilities for other ions. The assumption that  $\mathbf{s}_N = N^{1/2}$  will thus generally be an underestimate of the true variability in the number of ions detected. If a real set of data is collected, normalized and used to directly calculate the rms distance, the result will generally be bigger than the theoretical result in Eq. 19. This is, however, precisely what makes the metric useful: the more variable the set of data, the larger the rms distance will be. The rms distance can be calculated for any set of data, but Eq. 19 is only relevant to data sets containing a single type of particle.

Another important property of this metric (at least in the idealized limit) is that differences between peak areas are weighted by their statistical significance; big peaks do not necessarily have any greater influence on the results than small peaks. Once the spectra are normalized, every peak has the same standard deviation (in the ideal limit). When the difference between peaks areas or the distance between spectra is found, the

result is really just a measure of how many standard deviations separate pairs of peaks on average.

Because all of the elements in a vector representation of a spectrum are given essentially equal weight in determining the ultimate value of the rms distance, care must be taken when calculating an rms distance directly from a real set of data. The standard practice when clustering spectra is to produce 350 element vectors (one for the positive peaks and one for the negative peaks). Most of the elements in these vectors are zero or near zero. They don't contain any information about the ions that are actually present in the aerosol particle being analyzed; they simply represent noise. This means that the portions of the spectrum simply representing random chemical noise can dominate portions of the spectrum representing true signal. The solution is simply to choose carefully which mass-to-charge ratios are represented by elements in the vectors. Examination of an average spectrum is usually all that is required to determine which vector elements represent real ions (even if the ion peaks are very small) and which elements simply represent various types of background noise. In chapter 5, it is simply the ten largest positive peaks (i.e. vector elements) and the ten largest negative peaks that are included in the calculations. The ideal number will likely vary from experiment to experiment. Once again, it also should be noted that this particular metric should only be applied to a homogenous set of data (i.e. data from a single particle type).

At this point, all of the tools required to analyze the data from real experiments have been described. Chapter 4 describes the first real experiments performed. All of the data is smoothed, calibrated and so on. At the end of chapter 4, the clustering routine from

section 3.4 (i.e. BART) is used. In chapter 5, similar initial data processing is used and the variability metric just described is applied.

## Chapter 4. Spectral Consistency and Laser Power Dependence

Chapter 2 described the ATOFMS instrument and a number of experimental details. Chapter 3 described the algorithms used to process the raw time-of-flight data. This chapter shows the results of their use. The material covered in this chapter constitutes the bulk of the first paper published by the BAMS group<sup>105</sup>. That paper is believed to be the first published report showing both positive and negative mass spectra collected from the same individual spore. The fact that so much information can be obtained from such a small sample ( $\sim 10^{-12}$  g) in so short a period of time ( $\sim 1$  s) is truly impressive.

A useful BAMS system, however, must do much more than simply produce mass spectra from single particles. It must be able to identify particles on the basis of their spectra. Unless the spectra produced by a given type of particle are consistent and distinct from those of another particle type, the two types will not be identified or differentiated efficiently. Significant spectral variability is a shortcoming of the data produced by the current systems. This variability has many potential sources. Some of it may result from natural particle variations. More importantly, some of it may result from the instrument itself. If true, this must be prevented. As will be described in detail, the DI laser and in particular its beam profile are easily identified as potential sources of variability. The first experiments described here were performed using the standard DI laser to assess, as best as possible, its actual contribution to the observed spectral variability. (An improved laser setup is described and used in the next chapter.)

The laser pulse used to desorb and ionize molecules from aerosol particles is clearly one of the most important variables in BAMS. It is, after all, the source of energy that actually desorbs and ionizes the molecules that are analyzed to produce a mass spectrum.

It is widely known that MALDI ion yields and ion velocities are fluence dependent<sup>106-108</sup>. Furthermore, spectra from single non-biological particles have been observed to change dramatically with laser energy<sup>78, 109</sup>. Other groups have used single particle mass spectrometers to look at biological particles<sup>73-76</sup>, but the experiments presented here are believed to be the first attempt to systematically correlate single particle mass spectra from bacterial spores with laser energy and beam profile measurements. (A related paper was just recently published by the Prather group concerning their DI laser profile and the variability of spectra produced from 2,4-dihydroxybenzoic acid particles<sup>110</sup>.)

#### **4.1 Desorption/Ionization Laser**

A Big Sky Laser Technologies Ultra was used as the DI laser for the experiments described here. This is the “standard” Q-switched, frequency-quadrupled Nd:YAG laser described in chapter 2. The laser pulse energies used range from 0.2 - 1.0 mJ and were obtained by adjusting the laser’s flash lamp power. After proper focusing and alignment, the laser beam diameter was roughly 400  $\mu\text{m}$  (FWHM) at the target plane (the plane perpendicular to the laser beam where particles are most likely to be hit).

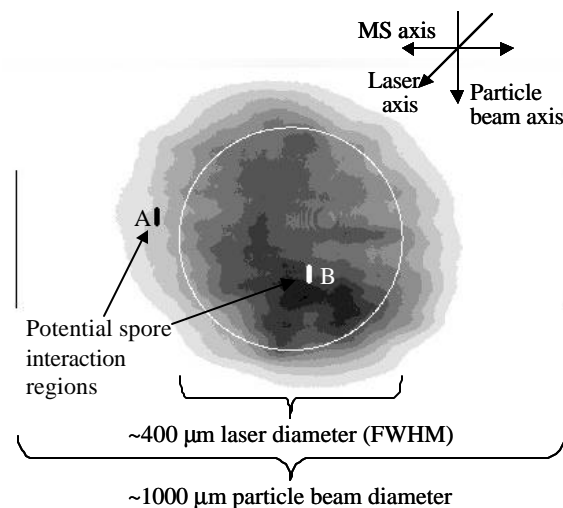
Terms describing laser energy, power, fluence and intensity are often used loosely in practice so it is worthwhile to clarify their meaning here. Intensity (measured in  $\text{W}/\text{cm}^2$  or  $\text{nW}/\mu\text{m}^2$ ) is used here to describe the rate at which energy is incident on an area at a particular point in time (the strict definition is power per solid angle). Fluence (in  $\text{J}/\text{cm}^2$  or  $\text{nJ}/\mu\text{m}^2$ ) describes the net energy incident on an area and is equivalent to the “intensity” integrated over the temporal duration of the 6 ns laser pulse. The total pulse



energy (in J or mJ) is the fluence integrated over the spatial cross-section of the laser beam.

The total pulse energy can be recorded for every analyzed particle, as was done here, but this quantity has limited value. The area over which the laser profile has a fluence greater than or equal to half its peak fluence is  $\sim 1.3 \times 10^5 \mu\text{m}^2$ . The cross sectional area of a spore is less than  $1 \mu\text{m}^2$ . Over the course of a 6 ns laser pulse the spore does move slightly, but even a spore traveling at 400 m/s traverses only  $2.4 \mu\text{m}$  and samples a very small fraction of the laser pulse. It is, therefore, the *local* intensity and *local* fluence that determine the maximum energy a spore may absorb.

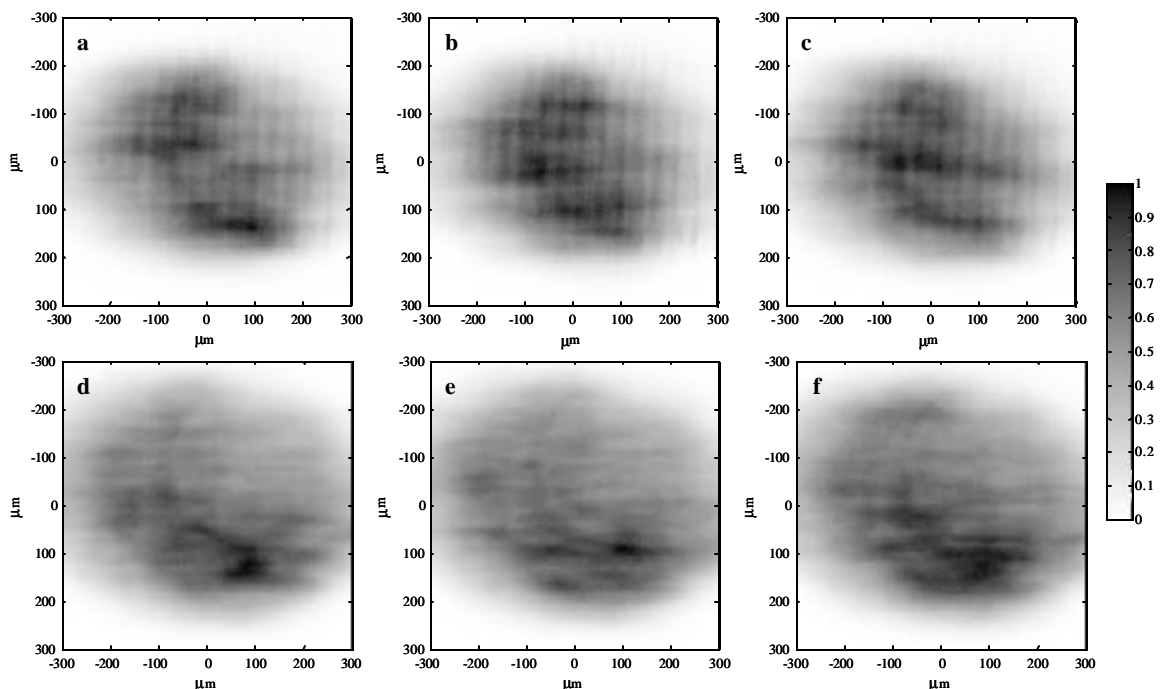
Figure 19 contains an image of the standard laser's profile (obtained with a Coherent LaserCam IID camera) that is clearly not uniform in fluence or intensity. This is significant because preliminary experiments indicate that the focused stream of aerosol



**Figure 19.** The profile of a single laser pulse is shown with two examples of regions (A and B) in which spores could interact with the pulse. The dimensions of the regions, which are elongated because of the spore velocity, have been magnified by a factor of 10 (area $\times 100$ ). The width of the focused particle stream is roughly 1 mm, as indicated by the vertical lines, so successive spores will interact with different regions of the laser pulse. The amount of area enclosed by the large circle is equal to the amount of area in the profile where the fluence is greater than or equal to half of the maximum fluence. Neighboring shades on the stepped linear gray scale represent fluences differing by 10% (of the maximum).

particles is more than 1 mm wide at the point where it intersects the laser (Figure 14).

The result of this is that successive aerosol particles interact with different portions of the laser beam and absorb different amounts of energy even if the pulse energy is fixed. A spore located near region B in Figure 19, for example, would encounter higher fluences and absorb more energy than an identical spore located near region A. Unfortunately, it is not currently possible to determine precisely where a given particle interacts with a laser pulse. As a result, it is not possible to determine the local fluence that a particle encounters or how much energy a given particle absorbs. Nonetheless, it is quite possible to control and measure the average fluence (i.e. the pulse energy) and use statistical arguments to infer the fluence dependence. It should be noted that simply focusing the aerosol particles more tightly would not solve the problem. The profile changes from shot to shot so even if particles always interacted at the same position within the profile, the

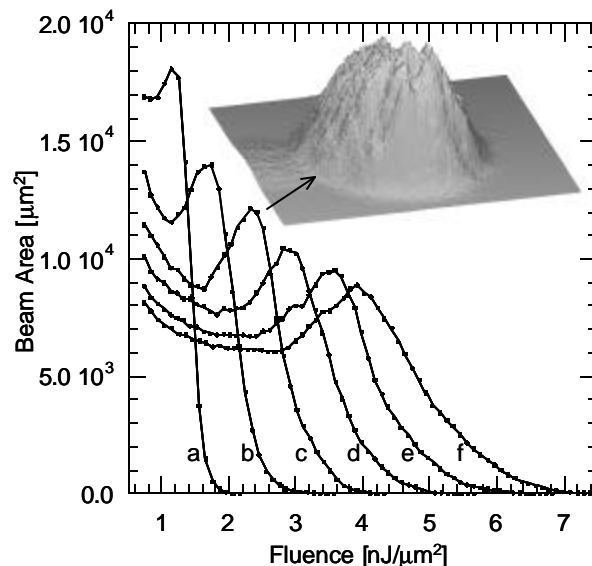


**Figure 20.** a-c show individual pulse profiles at  $\sim 0.2$  mJ. d-f show profiles at  $\sim 1.0$  mJ. The profiles have all been scaled to the same range to maximize contrast and to remove the effects of any pulse energy variations. The true average fluence distributions are shown below. The profiles clearly change from shot to shot. (Vertical fringe in a-c result from filtering optics and are not actual features of the profile.)

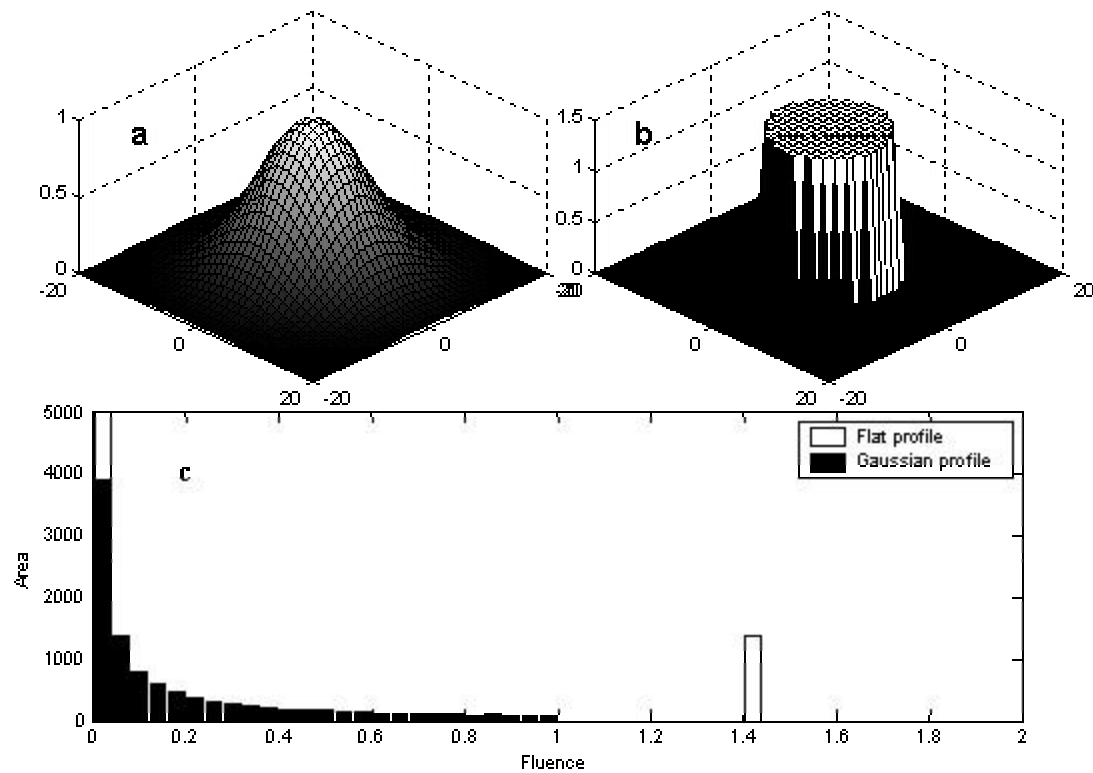
range of fluences encountered would still vary (Figure 20).

Figure 21 shows fluence distributions for the six laser settings used (0.21, 0.34, 0.50, 0.66, 0.84 and 1.02 mJ average pulse energies) and a different view of the profile shown in Figure 19. The units of  $\text{nJ}/\mu\text{m}^2$  ( $1 \text{ nJ}/\mu\text{m}^2 = 0.1 \text{ J}/\text{cm}^2$ ) are convenient since the physical cross-section of a spore is on the order of  $1 \mu\text{m}^2$ . The distributions in Figure 21 are useful, but not commonly presented (they are not line-outs or 1-D projections of the profile images), so a short description of their generation and significance is in order.

For each of the six laser settings, multiple images of the laser beam cross-section slightly in front of, at, and slightly behind the target plane were acquired. For each image, the absolute fluence was calculated for each pixel and a histogram of the values was produced. The number of pixels falling in each fluence bin was multiplied by the area per pixel ( $\sim 10.7 \mu\text{m}^2$ ) to find the total area represented by the bin in the laser cross-section. Histograms from the 49 images collected at each laser setting were then combined to



**Figure 21.** Laser fluence distributions produced by the six laser energy settings used. Distributions correspond to average laser pulse energies of 0.21 (a), 0.34 (b), 0.50 (c), 0.66 (d), 0.84 (e), and 1.02 mJ (f) respectively. A broad range of fluences is present in all laser pulses. This is expected to cause variability in the patterns of ions produced by the DI laser from shot to shot. The inset shows a typical 0.5 mJ, 266nm DI laser pulse profile (single shot). The diameter of the profile is  $\sim 400 \mu\text{m}$  (FWHM).



**Figure 22.** Laser beam profiles for an ideal Gaussian (a) and flattop (b) are shown. The fluence distributions for these profiles are shown in c. Only two distinct fluences are present in the flattop (the value at the top, and the background of zero). A range of fluences is present in the Gaussian profile with increasing area present at lower and lower fluences.

produce the results in Figure 21. The curves show the range of fluences that an individual aerosol particle may encounter. To help properly interpret these distributions, the fluence distributions for a perfect Gaussian and a perfect flattop profile are shown in Figure 22. The fluence distributions produced by the standard laser are not entirely consistent with those of a true Gaussian (some of the fluences are over represented in the cross section of the standard laser). Nonetheless, the standard laser's profiles will be referred to as roughly Gaussian for convenience.

If the simplifying assumption is made that the aerosol particle distribution is uniform across the laser pulse (recall that the particle beam diameter is much larger than the laser beam diameter), the area in the beam at a given fluence should be roughly proportional to

the probability of a spore interacting with that fluence. By including data from multiple laser shots and three spatial planes the effects of shot-to-shot energy and profile variations are reflected in the curves for particles whose positions may vary in all three spatial dimensions.

As expected, a broad distribution of fluences are observed for all the laser settings used. At the highest setting, the fluence ranges from zero to more than  $7 \text{ nJ}/\mu\text{m}^2$ . The broad distributions result primarily from the laser profile, but pulse-to-pulse energy variations also contribute. The distribution of laser pulse energies (measured while actively collecting spectra at a fixed laser energy setting) is approximately Gaussian with a standard deviation of 14% at the lowest setting, which shrinks to 4% at the highest. Ignoring profile variations for the moment, these pulse energy variations could cause the fluence to vary from its mean value by  $\pm 14\%$ . The profile, however, causes the fluence to range from zero to perhaps several times the mean. Any effects due to shot-to-shot pulse energy variations (at fixed laser setting) will be smaller than those due to the laser profile.

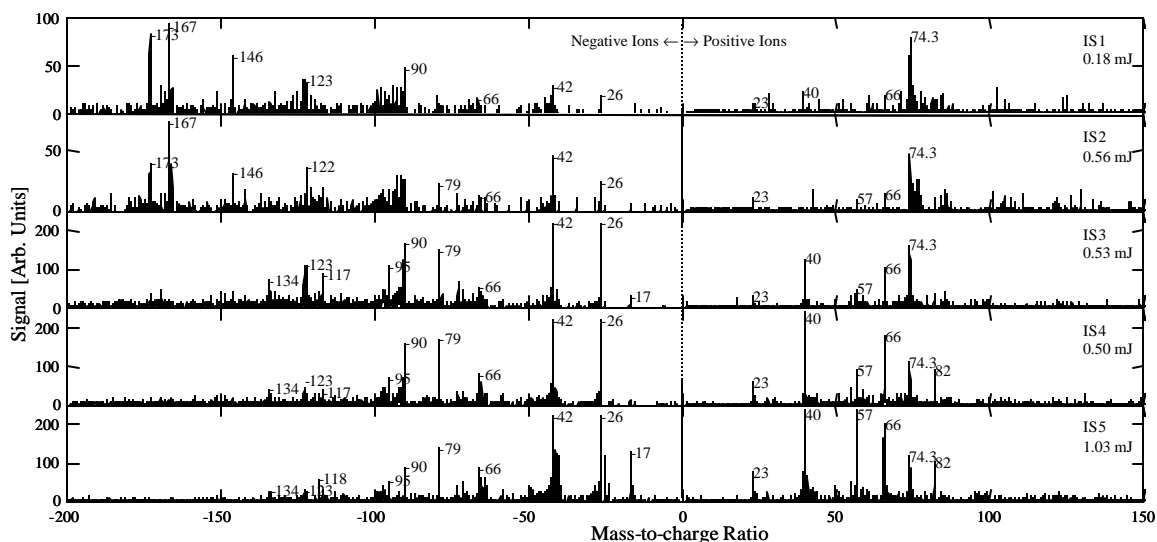
A potential source of error in the calculated fluence values is the camera system and associated optics. Inhomogeneities in the CCD response have not yet been fully quantified. Furthermore, the system produces images of a particular plane in space, but it is very difficult to ensure that this plane is exactly coincident with the center of the aerosol particle stream (i.e. the target plane). The focusing error may be 1-2 mm thus the errors in the calculated fluences may be 5-10%.

## 4.2 Data Collection and Initial Processing

*B.at.* spores were aerosolized using a Collision nebulizer. 1000 dual-polarity mass spectra from individual spores were collected at each of the six laser settings mentioned earlier. The data was collected without modification of the ATOFMS or associated laser systems so the range and particular values of the average laser energy were limited. Laser settings below 0.2 mJ did not produce spectra at a practical rate. Furthermore, the majority of the spectra that were produced had very few ion peaks, which made them difficult to calibrate accurately and ill-suited for identification purposes. Energies above ~1.0 mJ did not produce significant spectral changes compared to spectra taken at 1.0 mJ except for increasing fragmentation. The raw spectral data was smoothed, corrected for baseline and “autocalibrated” as described in chapter 3.

## 4.3 *B.at.* Spore Mass Spectra and Their Energy Dependence

Figure 23 shows individual dual-polarity mass spectra obtained from single spores at



**Figure 23.** Individual spectra (IS) from single spores collected with three different laser energy settings. The vertical scale is arbitrary but consistent from spectrum to spectrum in all figures. Although the three middle spectra are taken with the same approximate pulse energy, the first (IS2) resembles the low energy spectrum (IS1), the last (IS4) resembles the high energy spectrum (IS5) and the middle (IS3) resembles neither. It is argued that the similar spectra may have been generated with similar fluences even though the total pulse energies were different.

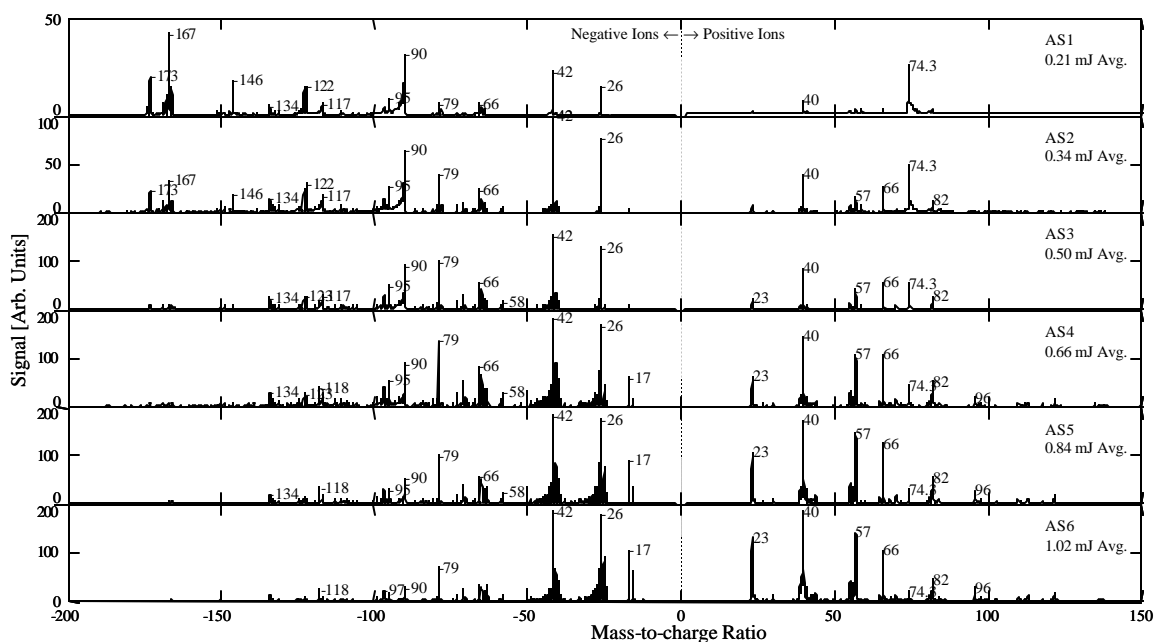
three (of the six) DI laser energy settings. Since this is the first time true dual polarity mass spectra are presented, it should be noted that each full spectrum in Figure 23 consists of a “half” spectrum from positive ions and a “half” spectrum from negative ions. Negative ions have negative mass-to-charge ratios ( $m/q < 0$ ) and positive ions have positive mass-to-charge ratios (i.e. the “charge” or “ $q$ ” is the actual net charge on the ion and not just the magnitude). There are of course separate microchannel plate detectors for the positive and negative ions and thus the vertical scaling (gain) may be somewhat different for the two halves of each spectrum. Nonetheless, the scales are consistent from spectrum to spectrum in all of the figures in this chapter (i.e. if two positive peaks have roughly the same area, they represent roughly the same number of ions, but this may be a different number than represented by similarly sized negative peaks).

The spectra IS1, IS3 and IS5 in Figure 23 are more or less typical of the spectra collected at the laser energies used to generate them (0.21, 0.50 and 1.02 mJ respectively). Although the three middle spectra were taken with the same approximate pulse energy, the 0.56 mJ spectrum (IS2) resembles the 0.18 mJ spectrum (IS1), the 0.50 mJ spectrum (IS4) resembles the 1.03 mJ spectrum (IS5) and the 0.53 mJ spectrum (IS3) is unique. It is apparent that IS1, collected with the lowest pulse energy, has the most prominent “high mass” peaks (defined here as  $|m/q| > 150$  for negative ions) and the least prominent low mass peaks. IS5, collected with the highest pulse energy, has the least prominent high mass peaks and the most prominent low mass peaks. All of this can be consistently explained as a result of the laser profile once it is demonstrated below that differences in absorbed energy produce these kinds of variations. Before that is done, however, some of the ion peaks will be identified.

Most of the prominent peaks seen in the spectra shown in Figure 23 have been tentatively identified (by other members of the BAMS group). The identifications are based on spectra acquired from DPA, DNA and various amino acids as well as *Bacillus* grown in isotopically labeled growth media (an article detailing the experiments is in preparation). The peak at  $m/q=-173$  is ascribed to arginine. A metabolic precursor of DPA, 2,3 dihydrodipicolinate, was assigned to the peak at  $m/q=-169$ . The peaks at  $m/q=-167$ ,  $-166$  and  $-122$  are believed to represent the molecular ion of DPA and two of its fragments. DPA is found almost solely in spore forming bacteria so these peaks are particularly useful for identification purposes. The peak at  $m/q=-146$  is attributed to glutamic acid, which is known to be common in *Bacillus* spores. The peak at  $m/q=-134$  is thought to be related to aspartic acid. Peaks at  $m/q=-97$  and  $-79$  are attributed to  $\text{H}_2\text{PO}_4^-$  and  $\text{PO}_3^-$  from phosphates found, for example, in nucleic acids and cell membranes. Peaks at  $m/q=-42$ ,  $-26$  and  $+23$  are identified as  $\text{CNO}^-$ ,  $\text{CN}^-$  and  $\text{Na}^+$  respectively. The DPA in spores is usually associated with calcium. Peaks at  $m/q=+40$ ,  $+57$ ,  $+66$  and  $+82$  are attributed to  $\text{Ca}^+$ ,  $\text{CaOH}^+$ ,  $\text{CaCN}^+$  and  $\text{CaCNO}^+$ , respectively. The prominent peaks at  $m/q=-90$  and  $+74.3$  remain to be identified. (The mass of the latter peak was the only prominent peak mass to vary significantly from an integer mass-to-charge ratio so an extra decimal value was kept.) Several of the peaks discussed here are similar to peaks that were identified in single particle *B. subtilis* spectra obtained by Gieray et al.<sup>75</sup>. That study, however, utilized significantly higher fluences ( $\sim 20 \text{ nJ}/\mu\text{m}^2$ ) at a different wavelength (308nm), which most likely explains some of the differences between those spectra and the ones presented here.



As explained earlier, profile variations produce variations in the energy absorbed by spores, but it remains to be proven that these absorbed energy variations actually have an effect on the mass spectra. (Obviously large differences in absorbed energy must have some effect on the mass spectra, but this doesn't mean that differences on the scale produced here will.) Since the energy absorbed by a single spore cannot be determined, the average energy absorbed and the average spectrum produced by a collection of spores must be considered instead. Figure 24 shows averages of 1000 spectra collected at each of the six laser settings used. Variations due to factors other than the laser setting are averaged out and cannot account for differences between spectra (ignoring very small statistical fluctuations). Figure 21 shows that each laser energy setting produces a different average fluence. This means that spores must absorb different amounts of energy of



**Figure 24.** Averages of the 1000 spectra (AS) collected at each laser energy setting. The average pulse energy is labeled on each spectrum. The highest mass negative ion peaks are in AS1 and steadily decrease in AS2-AS6. The highest mass positive ion peaks, however, steadily increase and are greatest in AS6. The spectral differences result from changes in the average fluence produced by each laser setting. This leads to the conclusion that the laser profile must introduce spectral variability from shot to shot since successive spores will interact with different regions (and different fluences) in the non-uniform profile and absorb different amounts of energy.

energy, on average, at each of the laser settings. Since the spectra in Figure 24 are clearly different, the energy absorbed by a spore must affect the spectrum that it produces. This proves that the laser profile will produce variations in the mass spectra.

Before the magnitude of the profile induced variations is further quantified, it is important to note the general trends in the average spectra of Figure 24. As the laser energy is increased, the amplitudes of high mass negative ion peaks (in particular  $m/q = -173$  and  $-167$ ) are reduced and the negative low mass peaks become more prominent. At low energy, the positive data is dominated by the peak at 74.3. As the pulse energy is increased, the peak initially grows and then declines. Most of the other peaks in the positive part of the spectrum grow monotonically with energy. In fact, the highest mass positive ion peaks are seen most clearly at the highest pulse energies (although they are much smaller than the low mass positive ion peaks). The trends in the average spectra are consistent with the changes observed above between the individual spectra IS1, IS3 and IS5.

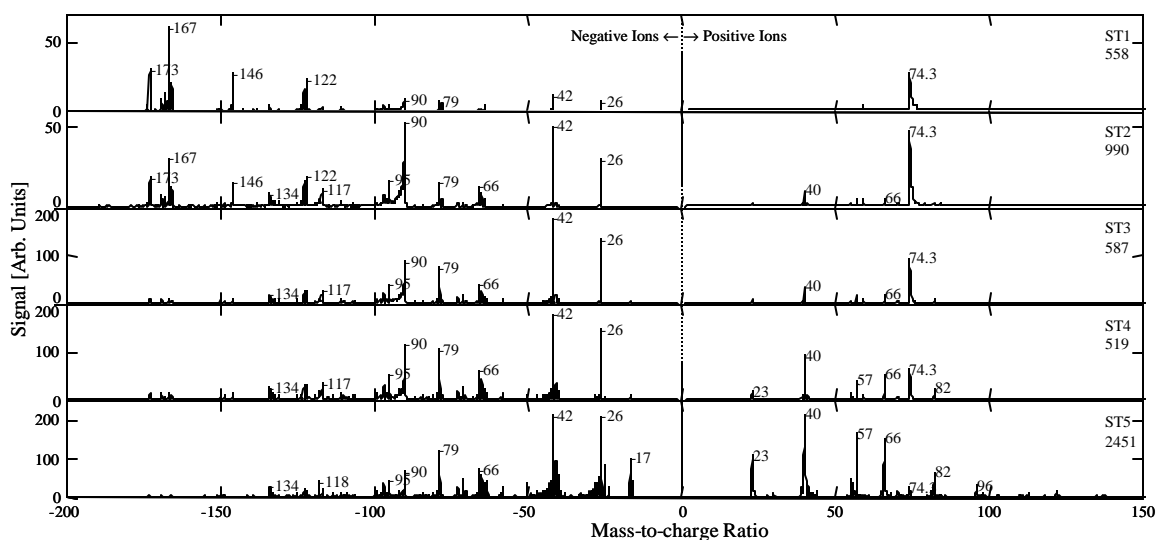
#### **4.4 The Magnitude of the Profile Induced Variations**

At this point, it is clear that profile related fluence variations must produce variations in the mass spectra, but the magnitude of these variations has not yet been determined. It is readily shown that the range of profile-induced spectral variations is at least as large as the difference between the average spectra AS1 and AS6 (and could be even larger). The line of reasoning used to reach this conclusion is based on the hypothesis that spores that interact with the same fluence must, on average, produce the same mass spectra (even if the laser setting is changed or there are other sources of variability). Figure 21 reveals that the ranges of fluence produced by the six laser settings are broad and overlap

significantly. Note in particular that the highest laser setting produces a range of fluences that covers and exceeds the range of fluences produced by any of the lower laser settings. If a large number of mass spectra were collected using the highest laser setting, it would be possible, in principle, to select a subset of the spectra that were produced by the same distribution of fluences that led to AS1 and, therefore, when averaged would produce a spectrum identical to AS1. It would in fact be possible to reproduce any of the average spectra in Figure 24. Similarly, subsets of data collected at the second highest laser energy setting could reproduce spectra AS1 through AS5 and so on. Since averaged subsets of data can produce variations of this magnitude, the variations between individual spectra must be at least as great.

Finally, it can be argued that spectral variations due to natural variations between particles (or other potential factors) are not likely to be significantly larger than the profile induced variations. It will be shown that when similar spectra are grouped together, the resulting groups appear and behave as though the spectra in a given group were generated at a similar fluence level. If natural variations (or other random variations) were significantly larger than profile induced variations, the spectra would not be expected to group together in this manner. It must be admitted, however, that these arguments are not entirely rigorous. Independent of the profile arguments, however, the clustering of similar spectra is important for automated identification algorithms so this topic merits discussion here.

The clustering algorithm employed is a modified ART2a neural network<sup>103</sup> named BART. It was described in the last chapter so it will not be described again here. The data collected at all six laser energy settings was combined and then processed with BART. Although BART is an improvement over earlier implementations of the ART2a neural network, it still produced an unwieldy number of clusters that were further concatenated or eliminated using a simple algorithm. This resulted in a total of five “super” clusters of spectra containing 5105 of the initial 6000 spectra (85%). The 895 spectra that were excluded had atypical features resulting, for example, from calibration errors (introduced by the automated calibration routine), impurities in the spore solution, or fragmented spores. The individual spectra in each cluster were averaged to obtain the spectra shown in Figure 25. Since individual spectra are only averaged with similar individual spectra, Figure 25 presents a better picture of the range of variability in the data than Figure 24

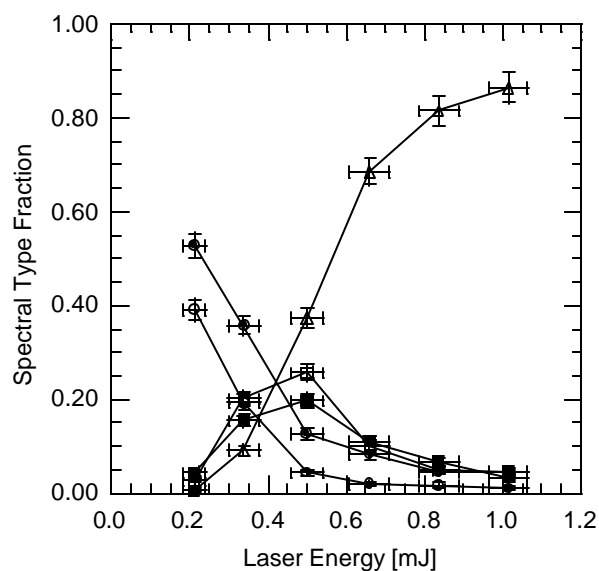


**Figure 25.** Averages from five clusters of spectra are shown. Each cluster is a group of similar individual spectra and so its average spectrum can be thought of as a spectral type (ST). The spectral types are sorted by the average of the laser pulse energies measured for every laser shot used to generate the spectra in each cluster. This produces an arrangement that is clearly consistent with the power dependent trends established in Figure 24. This is consistent with the belief that each spectral type is largely formed from spectra generated at a particular fluence. The number of spectra in each cluster is indicated below the spectral type label.

(where very different spectra were averaged together because they were produced at the same laser setting). It is appropriate, therefore, to treat the average cluster spectra as “spectral types”. This is important because an automated identification routine that compares an unknown spectrum to previously determined spectral types would be more likely to find a close match than another routine that compares an unknown spectrum with averages like those in Figure 24.

If most spectral variability truly resulted from fluence variations in the laser profile, the spectra that happened to be generated at a fixed fluence would show little variability and should cluster together (i.e. each spectral type would represent a specific range of fluences). Spectral types that represent high fluences would be comprised largely of spectra collected with high laser pulse energies (since these pulses contain the most area at high fluence). Clusters that represent low fluences would be comprised largely of spectra collected with low laser energies. The averages of the laser pulse energies individually measured for every spectrum in each cluster would, therefore, provide some indication of the fluence level represented by each spectral type. These average energies were calculated and the spectral types in Figure 25 are presented in order of this value. ST1 has the lowest average pulse energy while ST5 has the highest. It is reassuring, therefore, that the progression of spectral types is consistent with the power dependent trends identified in Figure 24. ST1 clearly looks like it is generated at the lowest fluence, ST2 at a slightly higher fluence and so on.

Figure 26 shows the distribution of spectral types collected at each laser setting. In general, the trends in the data are consistent with the expected power dependent behavior. The “low fluence” spectral types are primarily produced with low pulse energies and the “high fluence” spectral types are primarily produced with high pulse energies. An important observation is that spectra of all spectral types are produced at all laser energy settings, just at different ratios. This implies that all of the spectral types can be produced with fluences below  $\sim 2 \text{ nJ}/\mu\text{m}^2$  since the lowest energy pulses contain almost no area at a greater fluence. Fluences much above  $2 \text{ nJ}/\mu\text{m}^2$  evidently produce spectra that generally fall into spectral type 5 (ST5). The fact that ST5 has nearly two and a half times as many members as any of the other types is consistent therefore with the fact that most of the area in the profiles is well above  $2 \text{ nJ}/\mu\text{m}^2$ . At this point, it is not easy to explain how most natural variations (or other random sources of variability) would produce this type

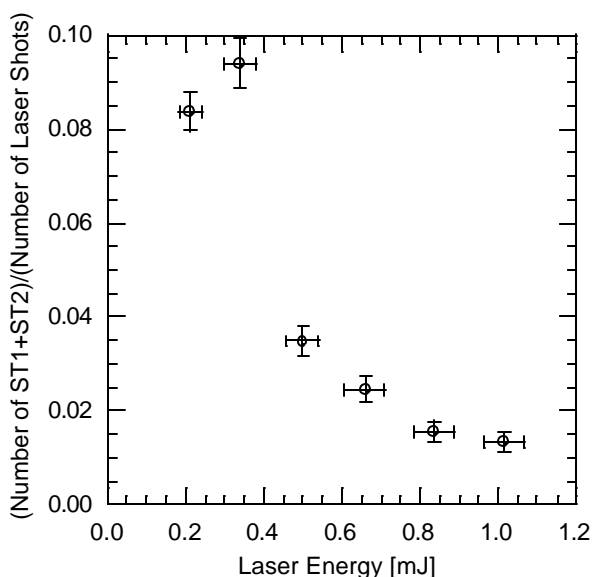


**Figure 26.** The relationship between spectral type and laser energy. Spectral types believed to represent low fluence spectra (ST1 ○ and ST2 ●) are produced primarily from low energy pulses. ST3 (□) and ST4 (■) appear to represent higher fluences and are largely produced by higher energy pulses. The highest energy pulses, however, overwhelmingly produce spectral type ST5 (△). Nonetheless, it is important to note that all of the spectral types are produced at all of the laser energy settings.

of behavior. It would seem, therefore, that natural variations are not likely to be significantly larger than profile induced spectral variations. Nonetheless, this possibility cannot be ruled out completely.

Before concluding, it should be noted that all spectral types, ST1 through ST5, are characteristic of bacterial spores and are distinct from many other aerosols <sup>111</sup>. All of the spectral types have value for identification purposes, but ST1 and ST2 are preferred because of their more prominent high mass peaks (e.g. those indicative of DPA). Figure 26 could easily be interpreted, therefore, as implying that the lowest laser energy is optimal since it produces the largest fraction of these two spectral types.

The hit rate, however, is an important and practical factor that must also be considered; not every particle that is fired upon by the DI laser produces a spectrum.



**Figure 27.** Spectral types ST1 and ST2 contain prominent peaks at  $m/q = -167$  and  $-173$  that are believed to be particularly useful for identification purposes. Figure 26 shows that the lowest laser energy (0.21 mJ) produces the greatest fraction of ST1 and ST2 but this does not take into account the rate at which spectra are produced. When the number of laser shots required to produce the “good spectra” is factored in, the laser setting of 0.33 mJ per pulse is found to be optimal. This optimal setting produces the greatest number of good spectra in a fixed amount of time and therefore helps maximize the sensitivity of the BAMS system.

Figure 27 shows the number of type ST1 and ST2 spectra obtained divided by the number of DI laser shots fired versus laser energy. The rate at which the laser fires is roughly constant, thus the ratio is approximately equivalent to the rate (in time) at which “good” spectra are acquired and can be considered a figure of merit. The plot in Figure 27 reveals that the second lowest pulse energy (0.33 mJ) is in fact optimal in this sense. Of course it is not the pulse energy itself that is of ultimate importance. Rather it is the fluences produced and the areas over which they are spread.

#### **4.4 Conclusions**

It has been shown that a non-uniform DI laser profile causes significant variability in the spectra generated from single *Bacillus* spores. This variability is clearly undesirable in BAMS or any other system that must identify single particle mass spectra with high confidence. Most single-particle mass spectrometers used today employ lasers with similar profiles so this is an important observation. In some applications the identification of spectral types may help deal with the resulting spectral variability, but it is generally better to reduce the variability directly. The next chapter will describe how a flattened profile was produced and how the flattened profile helped reduce the spectral variability.



## Chapter 5. Laser Profile Modifications, Fluence Thresholds and Improved Spectral Consistency

In the previous chapter, the Gaussian DI laser profile was shown to be an important source of variability in the mass spectra of individual aerosol particles. The range of fluences produced by any non-uniform laser pulse causes imperfectly collimated aerosol particles to absorb varying amounts of energy since successive particles can interact with different portions of the laser profile. Variations in the energy absorbed cause variations in the spectra produced. It is clearly desirable, therefore, to produce a more uniform profile.

Section 5.1 describes how the DI laser system was modified to obtain a nearly “flattop” profile. Using this profile, fluence dependent changes in mass spectra are more easily identified. The reason for this is simply that the range of fluences that the aerosol particles can interact with at a given laser energy is greatly reduced. The new profile also makes it possible to more easily and accurately determine the fluence thresholds for desorption and ionization (i.e. the minimum fluences required to successfully desorb and ionize significant numbers of molecules from specific types of particles). This is a fundamental property of each particle type and is measured here for bacterial spores (section 5.2) and several other materials (section 5.3).

The fluence threshold must be known, in fact, to properly quantify and compare the variability of data resulting from the original and modified profiles. In experiments described below (section 5.4), the standard DI optical system was modified to produce laser pulses with “Gaussian” and “flattop” profiles such that the cross sectional area in the profile where the fluence was above the threshold as well as the total amount of energy contained in the same area were fixed. In this manner it was ensured that the

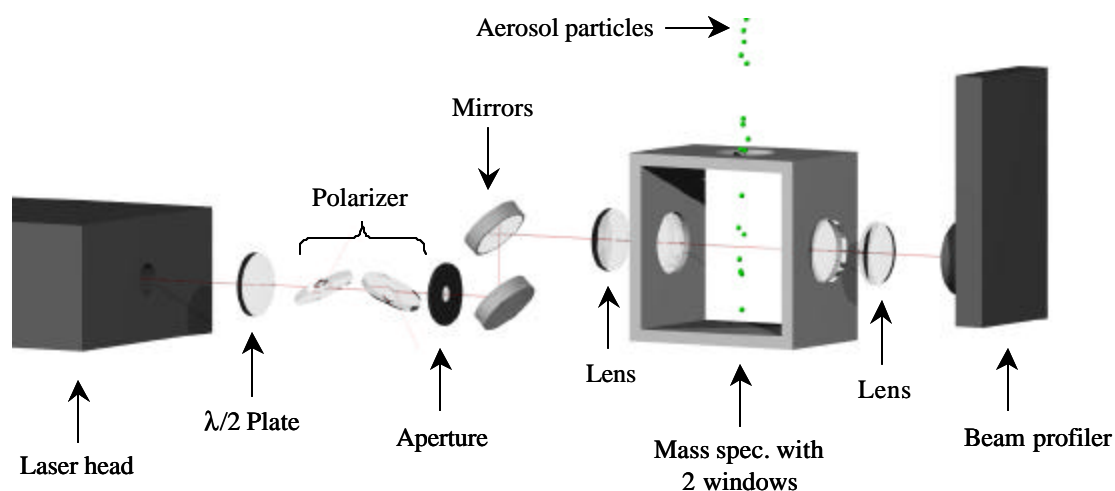
pulses produced the same “effective” average fluences. Sets of data collected with the two profile types are compared to show that the variability of the data is reduced by the improved laser profile.

The thresholds are clearly relevant for understanding and predicting the performance of the BAMS system. They are also relevant to other types of instruments. Several bioaerosol detectors look at the fluorescence from biological particles<sup>46, 50-52, 54, 112</sup>. More intense excitation generally leads to more intense fluorescence<sup>113</sup>, which is beneficial, but if the excitation is too intense the aerosol particles will be damaged potentially interfering with more selective types of analysis in subsequent stages of the instruments. In such instruments, the fluence thresholds calculated here serve as absolute upper limits on the excitation fluences that should be used (some damage is likely at fluences well below these levels however).

## **5.1 Flattopping and Characterizing the DI Laser Profile**

As in the last chapter, the DI laser used here is a Q-switched, frequency-quadrupled Nd:YAG laser (Ultra CFR, Big Sky Laser Technologies, Inc.) that produces pulses with a wavelength of 266 nm, a pulse length of less than 6 ns and a roughly Gaussian beam profile. The Ultra used in this chapter, however, has a new, more efficient fourth harmonic crystal that leads to higher pulse energies (>7 mJ/pulse) with reduced high spatial frequency profile variations. Nonetheless, the new laser’s unmodified beam profile is still closer to a Gaussian than an ideal flattop or top hat.

A number of methods have been used to produce more uniform laser profiles<sup>114-117</sup>. Since the Ultra can provide more energy than needed, a relatively simple method was used to improve the profile here at the cost of a significant fraction of the laser pulse energy. (Refractive optical devices in particular are much more efficient and may be of use in future instruments, but they were unnecessary for the experiments here.) Figure 28 shows the optical setup used to obtain and characterize a nearly flattop laser profile. A half-wave plate followed by a UV thin film polarizer allows the pulse energy to be continuously adjusted while the Ultra runs at full power. Rotation of the waveplate causes rotation of the linear laser polarization. The horizontally polarized component of each laser pulse (controlled by the waveplate) is reflected out of the main beam path by the polarizer thus enabling a controllable amount of energy to be transmitted through the polarizer. Running the laser at full power was critical since it produced a profile with a relatively flat central portion, minimized the pulse-to-pulse energy fluctuations and also

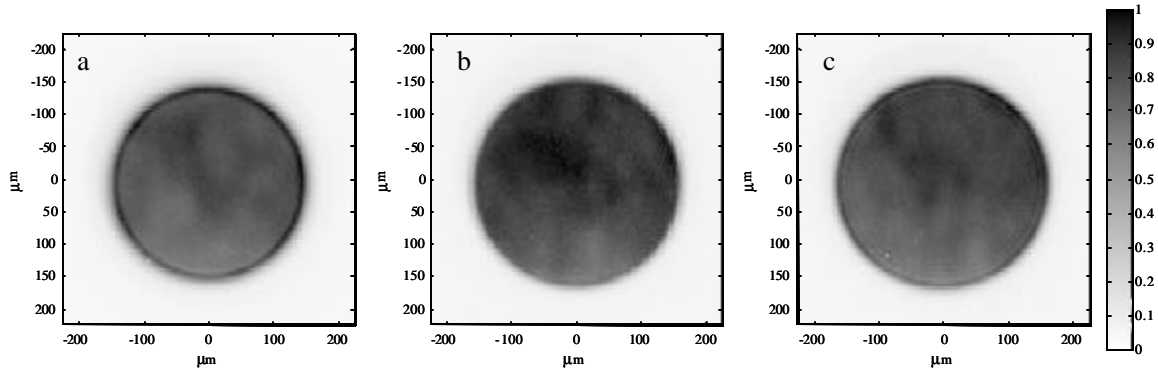


**Figure 28.** The layout of the optical system used to produce and characterize the flattened laser beam profile is shown (not to scale). The half-wave plate and thin film UV polarizer allow the laser pulse energy to be adjusted continuously while running the laser at constant power. The 1 mm diameter aperture clips the low fluence wings off the laser pulses. The first lens forms an image of the apertured pulse at the center of the mass spectrometer (reduced in size by a factor of  $\sim 3$ ). The second lens, in turn, images the profile at the middle of the mass spectrometer onto a LaserCam IIID camera with a UV profiling attachment.

minimized the timing jitter between the Q-switch trigger and the actual laser pulse. After the polarizer, the low intensity wings around the central flat portion of each pulse were removed with a 1mm circular aperture. At this point, the laser pulses had the desired profile, but they still had to be properly imaged onto the target plane. (As before, the target plane is defined as the plane normal to the laser beam that contains the axis of the particle stream where particles are most likely to interact with the laser.) The clipped pulses were reflected off two mirrors (used to align the beam through the chamber), passed through a ~130 mm focal length lens and window, and finally arrived in the ion source region of the mass spectrometer.

The distances between the aperture, lens and target plane were set such that a properly focused, one-third sized image of the aperture was obtained at the target plane. This ensured that the uniform profile at the aperture was reproduced as closely as possible at the target plane. Improper focusing would have caused diffraction rings to be seen on the profile. To verify that no rings were present, a second lens imaged the laser profile at the target plane (or nearby parallel planes) onto a Coherent LaserCam IIID camera with a BIP-12F UV profiling attachment (located well outside the mass spectrometer). This imaging system was carefully focused and calibrated using a procedure that will not be described here.

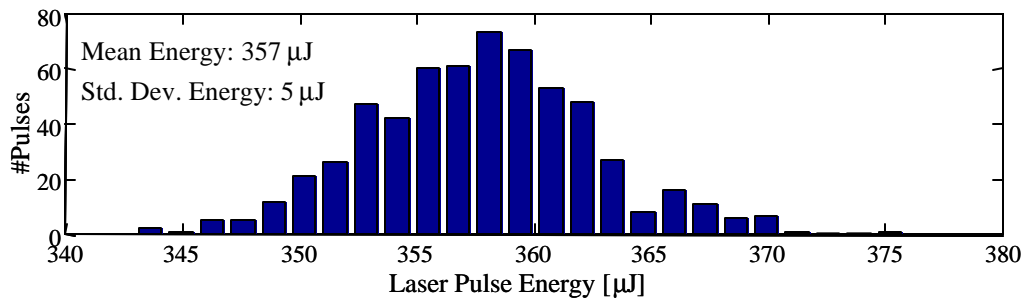
Figure 29 shows three images of the “flattopped” profile at three different planes within the ion source region of the mass spectrometer. At the target plane (b), the beam is approximately 320  $\mu\text{m}$  in diameter and relatively uniform. The absence of diffraction rings indicates proper focusing. On either side of the target plane (a and c), rings caused by diffraction at the aperture can easily be seen. (These rings cannot be avoided without



**Figure 29.** Three images of the flattop profile near the middle of the mass spectrometer. The left image (a) is 1mm before the target plane (i.e. 1mm toward the laser). The middle image (b) is at the target plane. The right image (c) is 1mm after the target plane. Note the diffraction rings in the images on either side of the middle and the absence of diffraction rings at the middle, indicating proper focusing at the target plane. Each image has been scaled to have a maximum value of one (in arbitrary units of fluence) to maximize contrast.

softening the edges on the profile). Nonetheless, the modified profile, in any of these planes, remains far more uniform than the unmodified profile (see Figure 43 below).

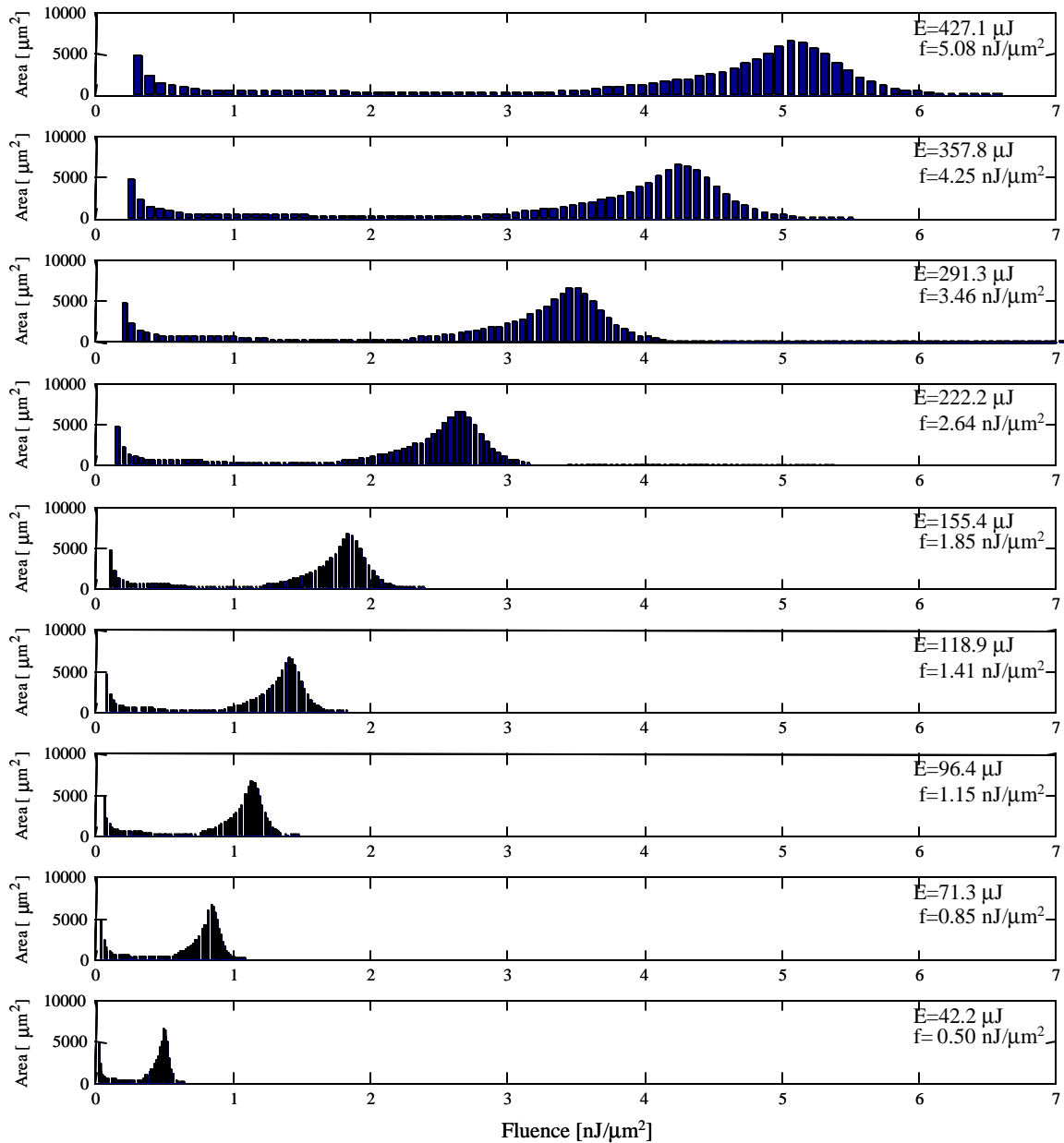
It is important to point out that even if the profile were perfectly flat, successive aerosol particles would interact with different fluences if the total laser pulse energy varied from shot-to-shot. Fortunately, it is very easy to measure the total pulse energy and in practice the energy used to create every mass spectrum is measured and stored. Figure 30 shows the spread of pulse energies typically observed while collecting data at a fixed laser setting. Since the laser is always run at full power and the pulse energy is selected



**Figure 30.** A typical histogram of the laser pulse energies observed while collecting a set of mass spectra. The time between laser pulses varies greatly since the aerosol particles arrive at random times. This is expected to cause larger pulse energy fluctuations than operation at a uniform rate. Nonetheless, the normalized standard deviation is only  $\sim 1.4\%$  and it does not change with the average pulse energy since the laser always operates at full power and is attenuated externally.

with a waveplate, the standard deviation is relatively low and is fixed at  $\sim 1.4\%$ . This pulse energy variation is insignificant compared to the range of fluences that still exists within the profiles.

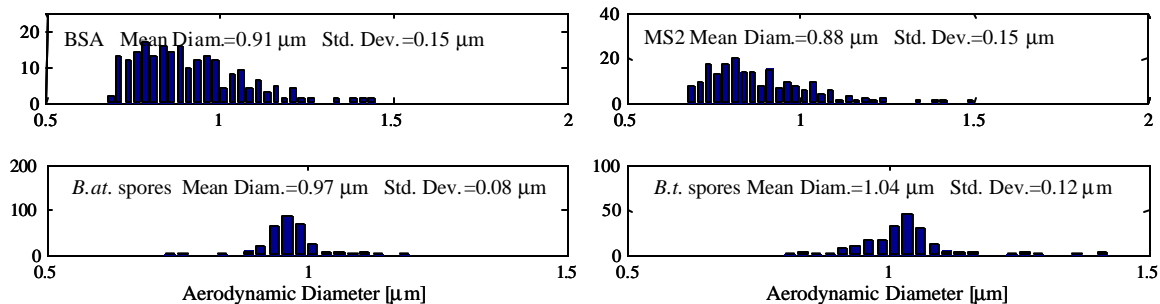
The fluence distributions (i.e. modified fluence histograms) that were introduced in the last chapter (Figure 21) are a useful tool to quantify the flatness of the “flattop” profiles produced here. Figure 31 shows fluence distributions for nine different average laser pulse energies used throughout the rest of this chapter. All of the distributions in Figure 31 are based on a set of 21 images of single laser pulses (with a single, nearly constant energy) at the target plane. A set of images based on a single laser pulse energy is sufficient to calculate the fluence distributions at other pulse energies because the waveplate and polarizer do not cause significant changes in the profile when adjusted. This was confirmed experimentally. It was also confirmed that data collected from the target plane alone provides a fair representation of the fluences encountered by particles scattered in all three spatial dimensions. A small tube limits the incoming particle stream diameter to a maximum of about 2.5 mm. When profiles taken 1 mm in front of the target plane, at the target plane and 1 mm behind the target plane are combined and equally weighted (which represents a worst case scenario), the resulting fluence distributions have virtually identical means and only slightly increased widths. The new profiles are not perfectly flat (i.e. perfectly uniform in fluence), but their fluence distributions are reasonably similar to the ideal results shown in Figure 22 and vastly superior to the original distributions shown Figure 21. The new profile will thus be referred to as a flattop for the remainder of this thesis.



**Figure 31.** “Flattop” fluence distributions are shown for nine different average laser pulse energies. Notice that a finite range of fluences accounts for the vast majority of the area at each energy and that these ranges are fairly distinct for different energies. This is similar to the ideal flattop result in Figure 22. The average laser pulse energies (E) are labeled along with the fluences corresponding to the peak of each distribution (f). The width (FWHM) of each distribution is ~15% of the labeled fluence.

## 5.2 DI Fluence Thresholds for *Bacillus* Spores

The flattop laser profile was used to collect sets of data at a series of laser energies from *B.at.* spores prepared in  $\frac{1}{4}\times$ TY, *B.at.* spores prepared in resuspension media (rs),



**Figure 32.** Typical size distributions for BSA, MS2, *B.at.* ( $\frac{1}{4} \times \text{TY}$ ) and *B.t.* particles aerosolized from solution using a Collison nebulizer. These distributions were measured by aerodynamic sizing in the ATOFMS system and include only particles that were successfully analyzed by the mass spectrometer.

*B.t.* spores prepared in rs, clumps of MS2 virions, clumps of BSA and clumps of DPA (all described in chapter 2). All of the samples were aerosolized using a Collison nebulizer (also described in chapter 2). Particle size distributions are shown in Figure 32. The details of the spore fluence threshold experiments are discussed in this section and the remaining threshold experiments are discussed in section 5.3.

The most extensive set of mass spectral data was obtained for *B.at.* ( $\frac{1}{4} \times \text{TY}$ ) spores so its collection, processing and analysis are described in detail. The collection and processing of data from the other samples were virtually identical, so only the results and a few important differences will be mentioned. For each particle type, the full data set was collected in one continuous period (~6 hrs) to ensure that the instrument performance remained as constant as possible except for intentional changes to the pulse energy or profile. Furthermore, duplicate sets of data were collected at the beginning and end of each experiment to ensure that there was no significant instrumental drift over the course of the experiment.

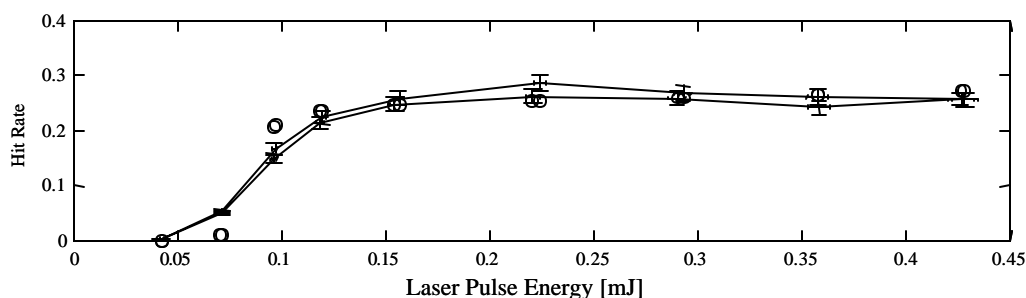
The basic experimental procedure was to collect spectra at a number of different laser pulse energies while recording the particle size distribution and hit rate. (As before, the hit rate is defined as the ratio of the number of particles that produce spectra to the



number of particles tracked.) The hit rate data and the laser profile data can be combined to determine the fluence threshold (as is described more fully below). It should be noted that MALDI experiments indicate that there may be a fluence threshold for the desorption of neutral material below the threshold for ionization in some experiments<sup>118, 119</sup>. Neutral material cannot be detected in the present instruments, so the thresholds reported here entail both desorption and ionization. The mass spectra themselves can be examined to determine the effects of the laser fluence on ion peak sizes and ratios.

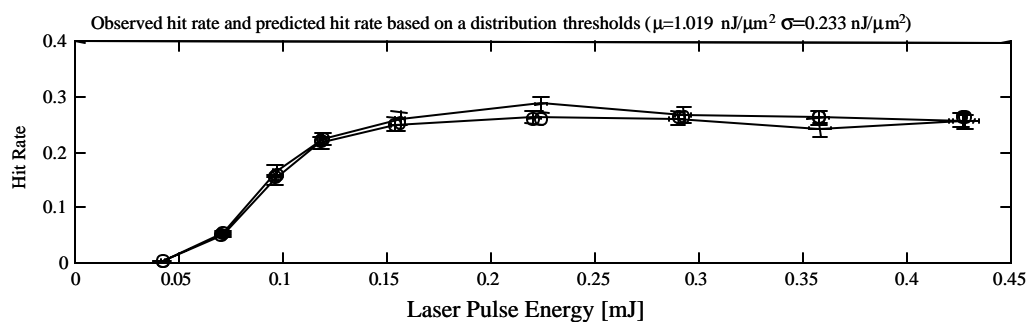
In the case of *B.at.* (1/4×TY), the laser energy was initially set to a value (~430 μJ) where the average fluence (~5 nJ/μm<sup>2</sup>) was well above the expected threshold fluence. From this starting point, the energy was stepped downward until the hit rate fell nearly to zero and then stepped back up (retracing the earlier steps) to the starting energy. Each time the laser energy was changed, ≥300 spectra were collected (except at the very lowest energy where only a few spectra were collected because of time constraints). The data points with error bars in Figure 33 shows the hit rates observed. The vertical error bars result from counting statistics. The small horizontal error bars represent the standard deviations of the laser pulse energies measured at each setting. The hit rates are expected to be significantly less than unity because it is possible for imperfectly focused particles to pass through both tracking lasers but pass well to either side of the ~320 μm DI laser beam. Inaccuracies in the timing system also cause the DI laser to occasionally fire at aerosol particles before (or after) they reach the proper vertical range.

The hit rate was observed to be relatively constant at the highest laser energies (>0.2 mJ). The simplest explanation for this is to assume that virtually every spore that interacted with a laser pulse, at one of these energies, produced a mass spectrum. Figure 31 shows that the vast majority of the beam area in such laser pulses had a fluence of 2 nJ/μm<sup>2</sup> or higher. As a consequence, it is relatively clear that 2 nJ/μm<sup>2</sup> is more than enough fluence to produce a mass spectrum from a *B.at.* spore. It is more informative, to examine the lower pulse energies when asking what the minimum fluence required to desorb and ionize significant numbers of molecules from a spore is. Another relevant question is whether all spores have the same fluence threshold or different thresholds. If the laser profile were perfectly flat, the pulse energy were perfectly constant (at a given setting) and all spores had exactly the same fluence threshold, the hit rate would be zero until the laser fluence exceeded the threshold and then the hit rate would jump almost instantly to its maximum value and remain there. This is clearly not what was observed. Although the pulse energy is consistent enough that its shot-to-shot variations can be safely ignored, the profile is less than ideal and the exact properties of the spore population are unknown so it is not immediately clear which factor is responsible for the gradual change of hit rate between pulse energies of 0.05 and 0.2 mJ.



**Figure 33.** The observed hit rates (solid line with error bars) and the calculated hit rates (○) for *B.at.* (¼×TY) based on a single fluence threshold of 0.95 nJ/μm<sup>2</sup>. The calculated “curve” rises more rapidly from a low hit rate to a high hit rate than the observed results. This indicates that different spores have at least slightly different thresholds.

It is unlikely that variations in the profile alone are capable of fully accounting for the gradual change. If the assumption is made, for the moment, that all spores do have the same fluence threshold and that the spores are equally likely to encounter any portion of the laser profile, the hit rate at a given laser energy should be proportional to the area in the profile above the fluence threshold. If for example, a certain fraction of the area in the profile is initially above the threshold and then the laser energy is increased such that the area above threshold is doubled, the probability that a spore will interact with that area must double as well. This means that the observed hit rate will double. (The hit rate includes only particles that produce spectra; this is usually different than the number of particles that intercept the laser.) If no part of the profile is above threshold, the hit rate should clearly be zero. Figure 33 includes calculated hit rates based on a single uniform threshold of  $0.95 \text{ nJ}/\mu\text{m}^2$  (○). It is easily seen that the slope of the calculated hit rate “curve” is steeper than the observed results. The most likely explanation is that different spores have different thresholds. This is a reasonable hypothesis because spores are known to have a range of sizes (Figure 32) and chemical properties. If the amount of DPA in the spores varied, for example, it would affect the amount of energy absorbed (DPA absorbs at 266 nm) and hence the probability to produce ions and a spectrum. The



**Figure 34.** The observed hit rates for *B.at.* spores ( $\frac{1}{4}\times\text{TY}$ ) (solid line with error bars) and the calculated hit rates (○) based on a distribution of fluence thresholds with a mean of  $1.0 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.23 \text{ nJ}/\mu\text{m}^2$ .

DPA content of *B. subtilis* W23 spores has been observed to vary by 30%<sup>120</sup>.

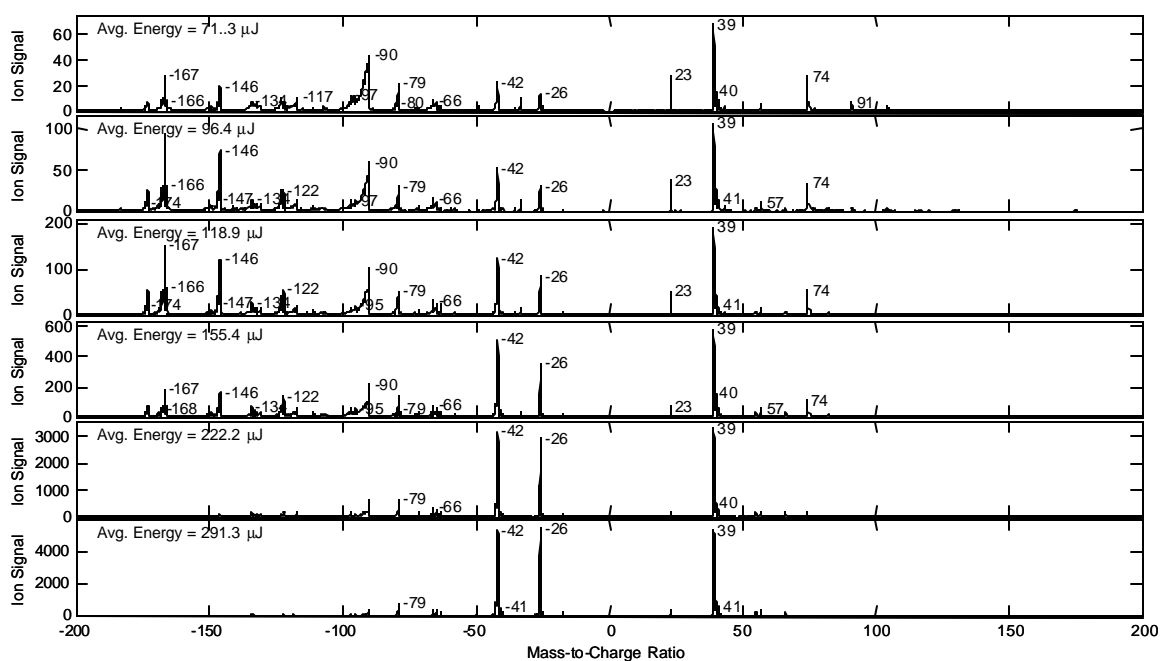
If the assumption is made that the spore population has a normal (i.e. Gaussian) distribution of thresholds characterized by a specific mean and standard deviation, these two parameters can be determined by what basically amounts to least squares curve fitting. The procedure also requires a third parameter, which is a simple proportionality constant to convert from beam area to hit rate, but this parameter is determined primarily by instrument settings and does not reveal fundamental spore properties. The end result of the fitting procedure is that *B.at.* spores grown in 1/4xTY media are found to have a distribution of DI fluence thresholds with a mean of 1.0 nJ/μm<sup>2</sup> and a standard deviation of 0.23 nJ/μm<sup>2</sup>. As shown in Figure 34, this “model” does a far better job of reproducing the observed hit rate curve than the previous model which assumed a single uniform spore threshold fluence.

The width of the fluence threshold distribution cannot be quantitatively predicted on the basis of currently measured spore properties, but its magnitude does seem compatible with the scale of inhomogeneities expected in the spore population. For a typical set of data, the mean aerodynamic diameter of analyzed spores is roughly ~1.0 μm with a standard deviation of 0.1 μm. This range of sizes will lead to a range of energies absorbed by successive spores. The complex index of refraction of *B. subtilis* spores in water at 265 nm is  $n=1.550+0.0138i$ <sup>121</sup>. Based on this value, a freely available Mie scattering algorithm<sup>122</sup> indicates that a 0.9 μm (physical) diameter spore would absorb 0.34 nJ of laser energy or 0.89 nJ/μm<sup>3</sup> (assuming a fluence of 1 nJ/μm<sup>2</sup> and a uniform spherical particle). A 1.1 μm diameter spore would absorb 0.56 nJ or 0.80 nJ/μm<sup>3</sup>. It is almost certain, therefore, that these particles will have different thresholds even though it

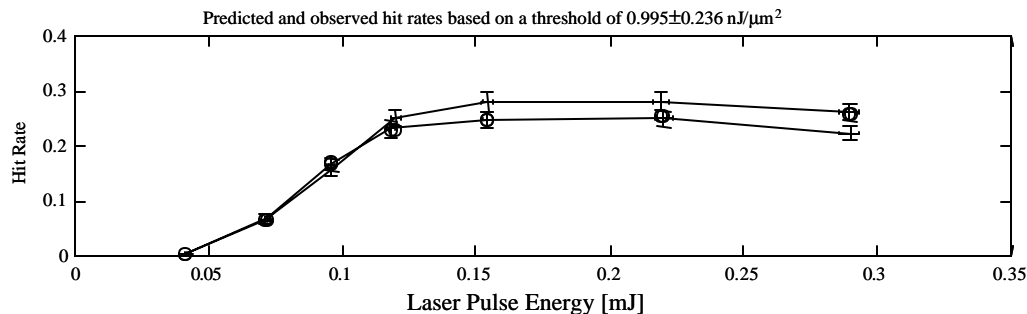
is implicitly assumed that they have identical chemical compositions (i.e. the same bulk index of refraction was used in both calculations). The fact that the chemical compositions do vary<sup>120</sup> will cause additional variations in the fluence threshold.

Before considering the thresholds from other particle types, it is worthwhile to briefly examine the *B.at.* spore spectra and their variation with fluence. Figure 35 shows average mass spectra collected at several laser energies. Each average contains a total of 600 spectra (300 from each of the two data sets collected at the labeled laser energy). The general appearance of the spectra is very similar to those shown in the last chapter, but there are distinct differences. The positive ion spectra now have a prominent peak at 39, which is presumably potassium. (There is a small peak at  $m/q=41$  consistent with the expected isotope.) Potassium is the only element likely to be encountered (in any significant amount) with an ionization potential less than the single 266nm photon energy (4.34 eV I.P. versus 4.66 eV photon energy). This means that it is likely to be detected very efficiently. The prominent new peak may thus be related to a very small impurity left over from the preparation procedure. Although there appear to be significantly fewer positive ion peaks compared to the data in chapter 4, this is largely explained by the fact that the newer ATOMFS instrument was used to collect this data. The newer instrument has a ~12-bit data acquisition system and a larger dynamic range. In the past large peaks were severely clipped making the smaller peaks appear more prominent. Most if not all of the peaks previously observed are still present; they are just difficult to see when compared to the potassium peak.

The *B.at.* spores dealt with up to this point were prepared in  $\frac{1}{4}\times$ TY media. To explore the effects of the growth media on the mass spectra and fluence threshold, a new sample of *B.at.* spores was prepared in resuspension (rs) media. As before, the data collected was started at high pulse energy, the energy was stepped downward until the hit rate fell nearly to zero and then stepped back up to the starting energy. With a better understanding of the expected behavior, however, it was possible to start at a lower energy ( $\sim 290 \mu\text{J}$ ) and take fewer spectra at each setting ( $\sim 200$ ). The observed and fit hit rates are shown in Figure 36. *B.at.* spores prepared in resuspension media are calculated to have a distribution of fluence thresholds with a mean of  $1.0 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.24 \text{ nJ}/\mu\text{m}^2$ . These values agree very well with the results obtained for *B.at.* ( $\frac{1}{4}\times$ TY). Before proceeding, it should be pointed out that there appears to be a slight, but systematic drift in the observed hit rate with time. The hit rates for the first four sets of



**Figure 35.** Average *B.at.* spore ( $\frac{1}{4}\times$ TY) spectra from the six lowest average energy settings where full sets of data were acquired. Each average contains 300 spectra from the first set of data and 300 spectra from the second set of spectra collected at each laser pulse energy setting. Higher energies produce spectra similar to the  $\sim 290\mu\text{J}$  spectrum. Note that the vertical range is different for different spectra.

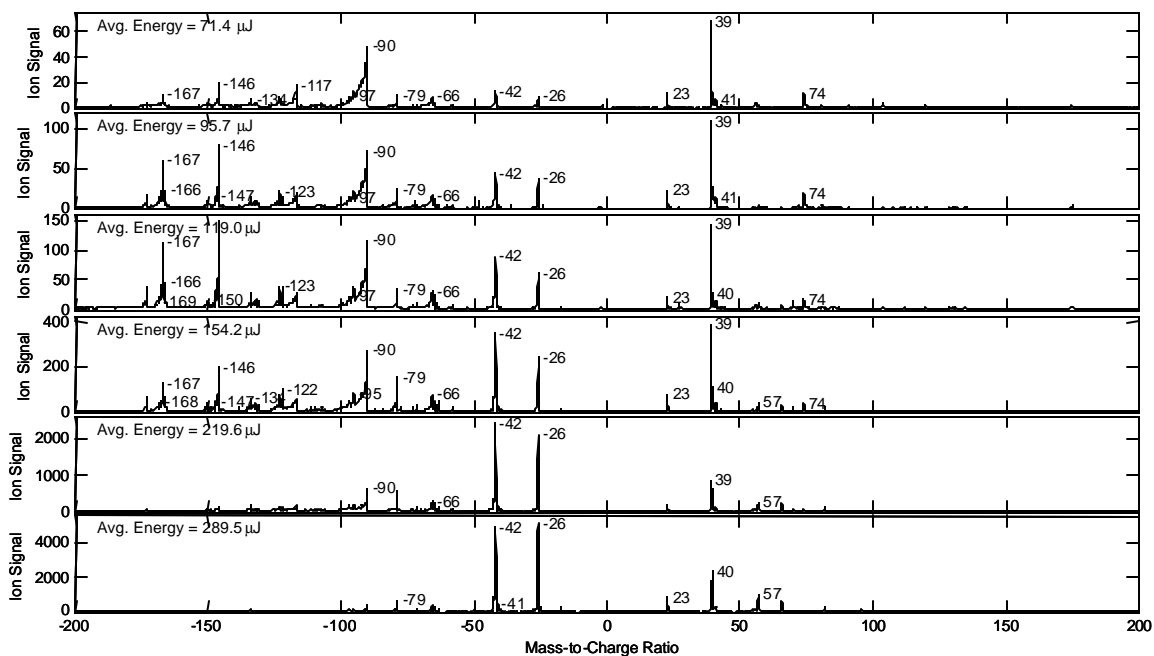


**Figure 36** The observed hit rates (solid line error bars) and the calculated hit rates (○) for *B.at.* spores prepared in resuspension media. The results indicate a distribution of fluence thresholds with a mean of  $1.0 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.24 \text{ nJ}/\mu\text{m}^2$ .

data (starting at the highest pulse energy) are consistently higher than the rates for the last four sets of data taken at the same pulse energies. Although it is conceivable that this is simply a statistical fluctuation, the drift most likely indicates that the nozzle or skimmers were slowly becoming clogged. These were not cleaned while the data was being collected since it was feared that the cleaning procedure might affect the subsequent hit rate data in an unpredictable way. Fortunately the drift does not appear to have had a significant effect on the threshold results.

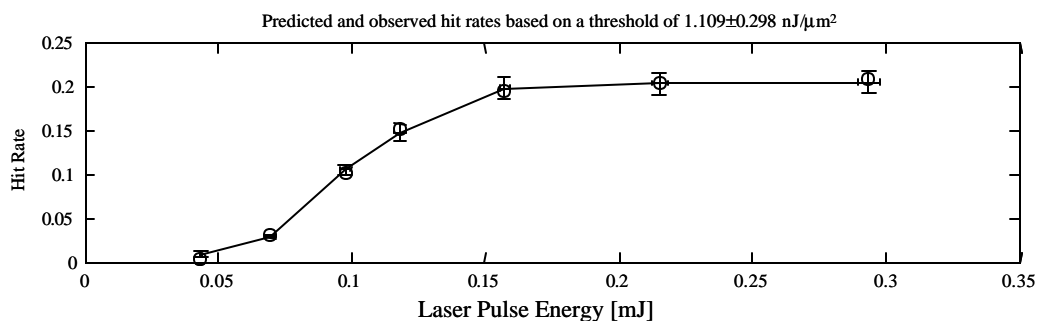
Figure 37 shows several average *B.at.* (rs) spectra. It is encouraging to see that the spectra are very similar to those shown in Figure 35 although the spores are grown in different media. One significant difference is that the 74 peak is consistently smaller in this media. This peak is important since it can be used to help differentiate certain *Bacillus* species (as shown in a separate publication<sup>11</sup>). The peak's origin is still not known. Since the change in growth media has not produced significant changes in the spectra, there is some reason to hope that additional small changes in the preparation should generally produce only small changes in the spectra as well. This will of course have to be confirmed experimentally.

The results of experiments with *B.t.* spores are now considered. Sets of *B.t.* spore data were collected just as for *B.at.* (rs). The same procedure described previously was also used to calculate the threshold distribution's mean and standard deviation. The observed and fit hit rate curves are shown in Figure 38. *B.t.* spores grown in resuspension media are calculated to have a distribution of fluence thresholds with a mean of  $1.1 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.3 \text{ nJ}/\mu\text{m}^2$ . Although it may not be an experimentally significant difference, the *B.t.* spores have a slightly higher mean threshold than the *B.at.* spores analyzed earlier. The higher threshold may be related to the fact that the average aerodynamic diameter of the *B.at.* spores was  $0.98 \mu\text{m}$  while the value for *B.t.* was  $1.05 \mu\text{m}$  (averages include only particles that produced spectra). The threshold distribution is also broader for the *B.t.* spores. This is not surprising though since microbiologists in the BAMS group have observed that *B.t.* does not sporulate as uniformly as *B.at.* This may also explain the fact that the distribution of diameters was much broader for the *B.t.*



**Figure 37.** Average *B.at.* spore (rs) spectra. Each average contains 200 spectra from the first set of data and 200 spectra from the second set of spectra collected at each of the labeled laser pulse energy settings.

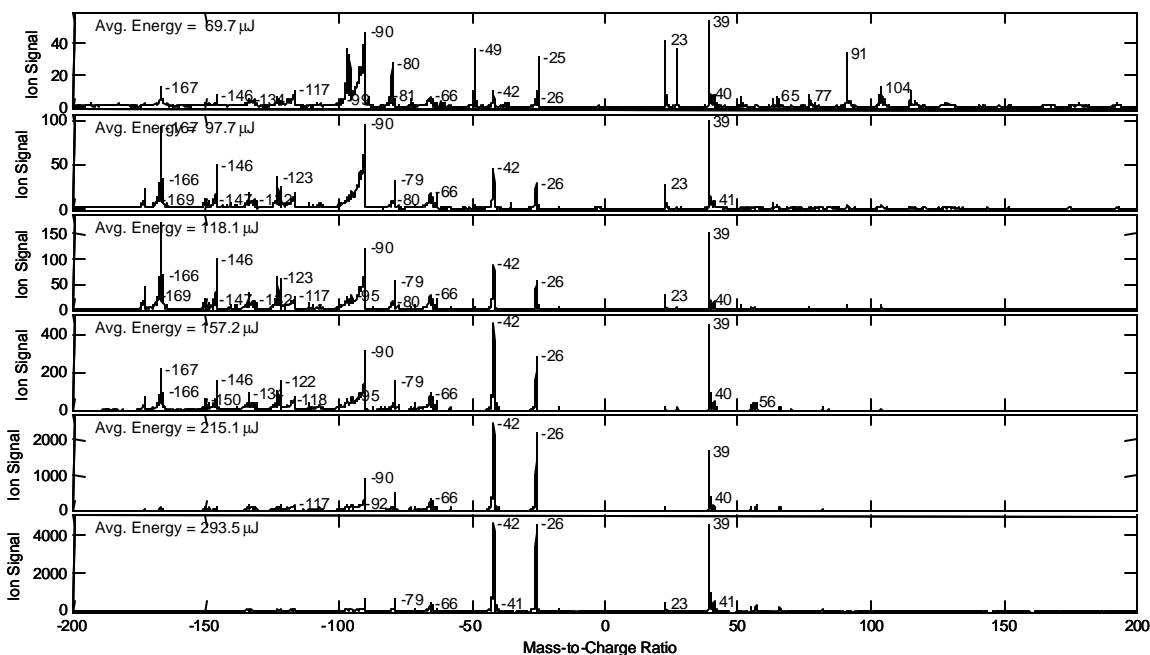




**Figure 38.** The observed hit rates (solid line with error bars) and the fit hit rates (○) for *B.t.* spores (rs) indicating a distribution of fluence thresholds with a mean of  $1.1 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.3 \text{ nJ}/\mu\text{m}^2$ .

spores than for *B.at.* spores.

Average spectra are shown in Figure 39. At the lowest energy in particular, there appear to be new or at least much larger peaks than seen in the *B.at.* spectra. There are significant peaks at  $m/z = -97, -96, -80, -49, -25$  and  $+91$ , for example. If the spectra at the lowest energy are clustered using BART, however, something very interesting happens. Roughly speaking, there are two prominent and distinct types of clusters that result. One set of clusters produce average spectra that more closely resemble the *B.at.* spectra. These clusters almost certainly represent real *B.t.* spores. The other set of clusters looks very different and has large peaks at  $m/z = -97, -96, -80, -49, -25$  and  $+91$ . It is very unlikely that these clusters represent spores. The particles are most likely impurities from the growth media or other sources. At any rate, this second set is the source of the prominent new peaks in the overall average. Although it is probably not clear in these average spectra, differences in the areas of peaks at  $m/z = -173$  and  $+74$  are sufficient to differentiate the two species (*B.at.* and *B.t.*) quite well in practice<sup>111</sup>.



**Figure 39.** Average spectra from *B.t.* (rs) spores. Each is the average of 200 spectra.

Since the two species of *Bacillus* spores considered here have basically the same size and basically the same chemical composition, it only makes sense that they have similar thresholds ( $\sim 1 \text{ nJ}/\mu\text{m}^2$ ). It is worthwhile to explore the significance of this particular value in a bit more detail. Based on the complex index of refraction mentioned earlier, a  $1 \mu\text{m}$  spore is expected to absorb  $0.47 \text{ nJ}$  (at a fluence of  $1 \text{ nJ}/\mu\text{m}^2$ ). The heat capacity of a spore is unknown, but it can be approximated very crudely using the heat capacity of tryptophan, which is  $238.1 \text{ J/mol K}$ . The mass of tryptophan is  $204.2 \text{ Daltons}$ . The mass of a  $1 \mu\text{m}$  spore is  $7.6 \times 10^{-13} \text{ g}$  (assuming a perfectly spherical particle and a density of  $1.45 \text{ g/cm}^3$ <sup>123</sup>). The spore will thus heat by  $\sim 530 \text{ K}$  upon interaction with the laser pulse. It is interesting to note, that this is very similar to the temperature of the pyrolysis tube ( $550^\circ\text{C}$ ) in the BLOCK II CBMS<sup>62</sup>. It is also of the same order of magnitude as the peak predicted surface temperature of a common matrix (DHB) in a UV MALDI experiment<sup>124</sup>.

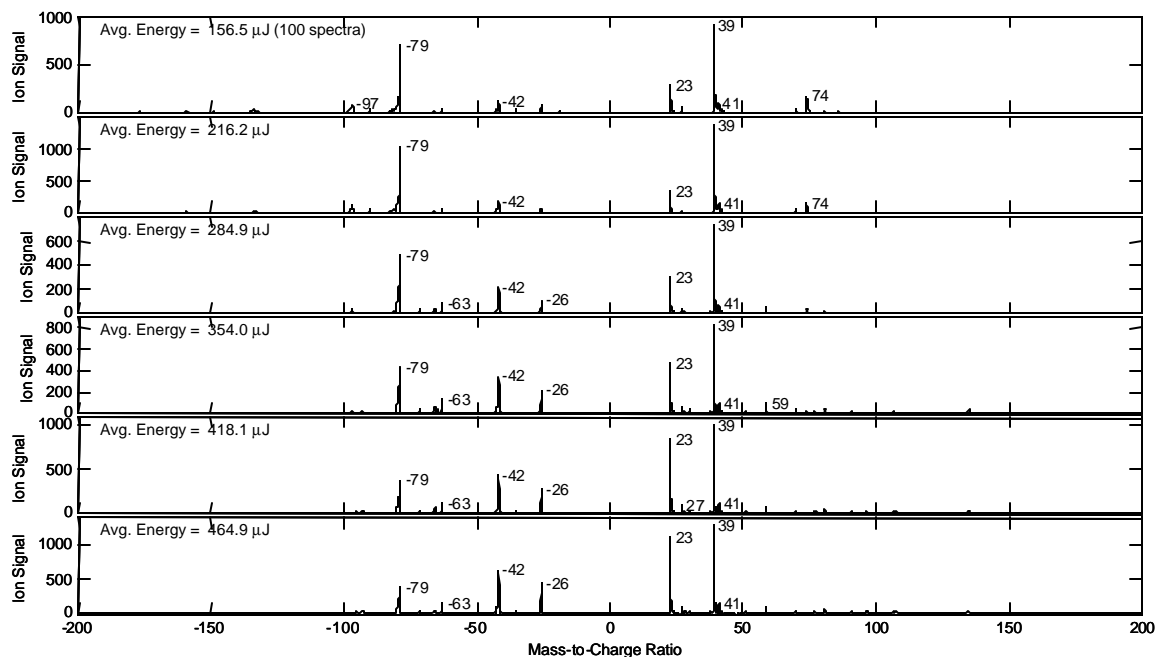
### 5.3 DI Fluence Thresholds for BSA, MS2 and DPA

The same data collection process used for the spores was repeated for clumps of MS2 virions and clumps of BSA. The quality of the BSA data was somewhat less than that for the other particle types because a low hit rate limited the amount of data that could be collected. Nonetheless, MS2 particles were observed to have a distribution of fluence thresholds with a mean of  $3.2 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.8 \text{ nJ}/\mu\text{m}^2$ . BSA particles were observed to have a distribution of fluence thresholds with a mean of  $2.7 \text{ nJ}/\mu\text{m}^2$  and a standard deviation of  $0.4 \text{ nJ}/\mu\text{m}^2$ . Both of these values are significantly higher than the values found for spores. The average size of hit particles was roughly  $0.9 \mu\text{m}$  for both MS2 and BSA, which is smaller than the mean spore diameter observed.

It would be very interesting to measure the UV absorption cross section of the particles. If a particle absorbs only weakly at 266nm, it would generally be expected to have a higher fluence threshold than another particle that absorbs strongly at 266nm. Spores may very well absorb more energy than either of these two types of particles. Unfortunately, the BAMS group does not have the ability to measure the absorption of single particles (note that absorption is only one component of the total extinction, which is more easily measured). Nonetheless, BSA is a large protein molecule, and MS2 is composed primarily of protein (and some RNA). The relative fractions of aromatic amino

	MS2	BSA	Ricin	Botox (A)
Phenylalanine	736	30	18	72
Tryptophan	372	3	10	14
Tyrosine	736	21	23	75
Total AA	23793	607	529	1296
Fraction	0.0775	0.0890	0.0964	0.1242

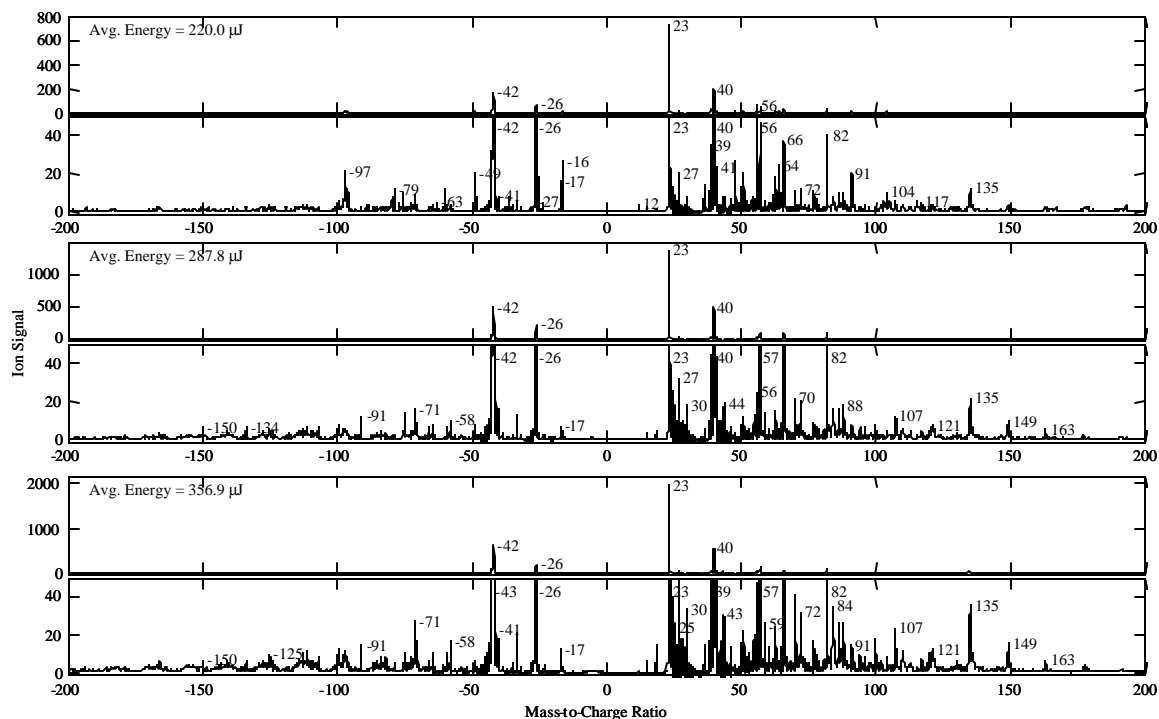
**Table 2.** Amino acid (AA) composition of simulants and toxins. MS2 is of course not a protein, but each virion contains 1 copy of the maturation or assembly protein and 180 copies of the coat protein. Phenylalanine, tryptophan and tyrosine are known to absorb at 266nm.



**Figure 40.** Average mass spectra from MS2 collected at six different laser energies.

acids (which are known to absorb at 266nm) contained by BSA and MS2 can thus be used as crude substitutes for the absorption cross-sections. Table 2 shows the number of phenylalanine, tryptophan and tyrosine amino acids contained by MS2, BSA and the toxins for which BSA serves as a surrogate. BSA contains a somewhat larger fraction of the amino acids of interest than MS2 and was observed to have the lower threshold. Although it is highly speculative, it is tempting to hypothesize that that ricin and botulinum toxin would have even lower thresholds.

Average spectra are shown in Figure 40 for MS2 and Figure 41 for BSA. A number of familiar peaks are present. In MS2, peaks at  $m/q = -97$ ,  $-79$ ,  $-42$  and  $-26$  are presumably  $\text{H}_2\text{PO}_4^-$ ,  $\text{PO}_3^-$ ,  $\text{CNO}^-$  and  $\text{CN}^-$ . Peaks at  $m/q = +23$  and  $+39$  are very likely  $\text{Na}^+$  and  $\text{K}^+$ . Some of the same peaks can be seen easily in the BSA spectra. Spectra with expanded vertical scales have been included in Figure 41 to reveals a number of additional small peaks. BSA appears to have peaks at  $m/q = -97$  and  $-79$  like those in MS2. Peaks at  $m/q = +39$ ,  $+57$ ,  $+66$  and  $+82$  are also seen and are presumably  $\text{K}$ ,  $\text{CaOH}^+$ ,  $\text{CaCN}^+$  and



**Figure 41.** BSA average spectra (each is the average of ~200 individual spectra). Because of the lack of features in the normally scaled spectra, spectra with expanded vertical scales have been included to reveal small peaks (most of these are too small to be useful in single spectra however).

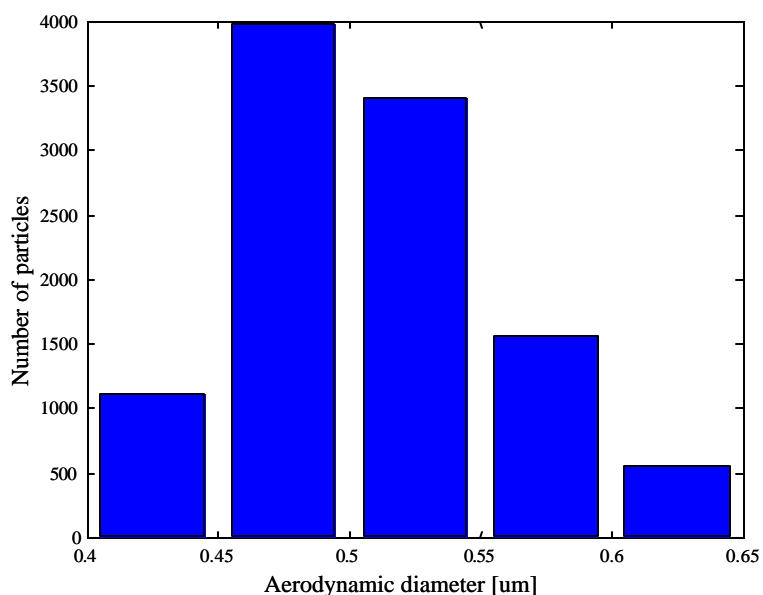
$\text{CaCNO}^+$ . A number of these elements (e.g. Na, K, and Ca) are not usually found in pure protein, so it is likely that the impurities in the BSA sample have a significant effect on the observed spectra. It is not clear what effect the impurities may have on the fluence threshold.

The desorption and ionization processes that lead to the observed mass spectra are not fully understood for any of the particle types and are likely to be complex. In some of the particle types, certain molecules may function as MALDI-like matrices for other molecules, but even classical MALDI is not well understood. Nonetheless, classical MALDI probably provides the best experimental analog for the current experiments so what is known about MALDI will briefly be summarized here. At the same time, significant differences between typical MALDI experiments and the present experiments will be pointed out. The production of ions in UV MALDI is thought to be the result of a

two step process<sup>124</sup>. In the first step, the laser either directly or indirectly generates “primary ions” during or shortly after the laser pulse. Knochenmuss claims that exciton pooling may be an important mechanism in the formation of the primary ions since matrix ionization potentials tend to be more than twice a “typical” photon’s energy<sup>124</sup>. At 266 nm, however, the photon energy is substantial and many molecules have ionization potentials less than 9.3 eV (twice the 266 nm photon energy); examples include virtually all of the amino acids<sup>125</sup> and several elements such as potassium, sodium and calcium (which are responsible for significant peaks in the present mass spectra). Direct photoionization may thus play a larger role here than in typical MALDI experiments. At any rate, in the classical MALDI model, deposition of energy by the laser causes desorption of the matrix and the formation of a plume. In the plume, reactions take place between the primary ions and other molecules producing whatever secondary ions may be energetically favored. These secondary reactions may include proton transfer, cation transfer and electron transfer or capture<sup>126</sup>. The importance of these secondary reactions is less clear in the case of small, isolated particles. The density of the plume from a 1  $\mu\text{m}$  diameter particle almost certainly drops much faster than the density of the plume in a typical MALDI experiment where expansion occurs from a large 2-dimensional substrate. The possibility that this effect may be significant is supported by the fact that the laser spot size, and thus the effective sample size, is known to be an important variable in MALDI<sup>119</sup>. It is probable that MALDI-like processes have at least some role in determining the mass spectra produced by the single particles here, but there are clearly important differences between the current experiments and typical MALDI experiments.

An experiment was performed using clumps of DPA in an attempt to directly observe a particle-size dependent fluence threshold. Particles with a significant range of sizes were generated, sampled and analyzed by the instrument at a series of laser energies. The collected data was then sorted into size bins. Figure 42 shows the exact size bins that were employed and the total number of particles analyzed in each bin (combining all laser energies). Once the particles were sorted, the fluence threshold for each size bin was calculated just as was done for the previous particle types. A 0.6  $\mu\text{m}$  DPA particle has  $\sim 2.4$  times the volume of a 0.45  $\mu\text{m}$  particle, but no significant differences were observed in the threshold fluences. The mean threshold for all sizes is approximately  $0.59 \text{ nJ}/\mu\text{m}^2$ . The standard deviation of the threshold distribution varies somewhat from size to size, but  $0.07 \text{ nJ}/\mu\text{m}^2$  is the median value.

The constant fluence threshold is consistent with a desorption and ionization process in which the laser pulse simply creates ions at the surface of the DPA particles. Measured



**Figure 42.** The number of DPA particles analyzed in each of the size bins studied is shown. The smallest size bin actually contains all diameters observed below 0.45  $\mu\text{m}$  and the largest size bin contains all diameters above 0.6  $\mu\text{m}$ . Differences in the total numbers analyzed primarily reflect the size distribution of particles produced and the size-dependent instrument sampling and focusing efficiency (the differences do not indicate different fluence thresholds).

values for the optical and physical properties of DPA necessary to perform a quantitative absorption calculation are not available, but DPA is known to absorb strongly at 266 nm<sup>101</sup> (consistent with the low fluence threshold observed). As a result, it is not unreasonable to believe that much of the energy incident upon a DPA particle is absorbed at its surface. If true, ions are likely to be produced at the surface and the DPA fluence threshold should be largely independent of the particle size, just as observed. In the contrasting case of spores, only ~60% of the directly incident energy is absorbed. The energy that is absorbed is thus likely to be spread more uniformly throughout a spore than a DPA particle (ignoring the heterogeneous spore structure for the moment). In this type of situation (i.e. semi-uniform energy absorption), a stronger size dependence should be expected for the fluence threshold.

#### **5.4 Direct Observation of Variability Reduction with Flattop**

The purpose of the experiments described in chapter 4 was to determine whether the “Gaussian” DI laser profile, which had always been used up to that point, was truly a significant source of mass spectral variability or not. Every method available to produce a flattened profile has some shortcomings (e.g. inefficiency, size, alignment sensitivity, cost, etc.), so there was little desire to modify the original profile until it was shown in chapter 4 that the laser profile’s fluence distribution did indeed contribute to the spectral variability and that a reduction in the variability would almost certainly result if the laser profile were flattened. Once this conclusion was reached, the profile was “flattopped” (as described above). Once the flattop profile was available, it became possible to directly

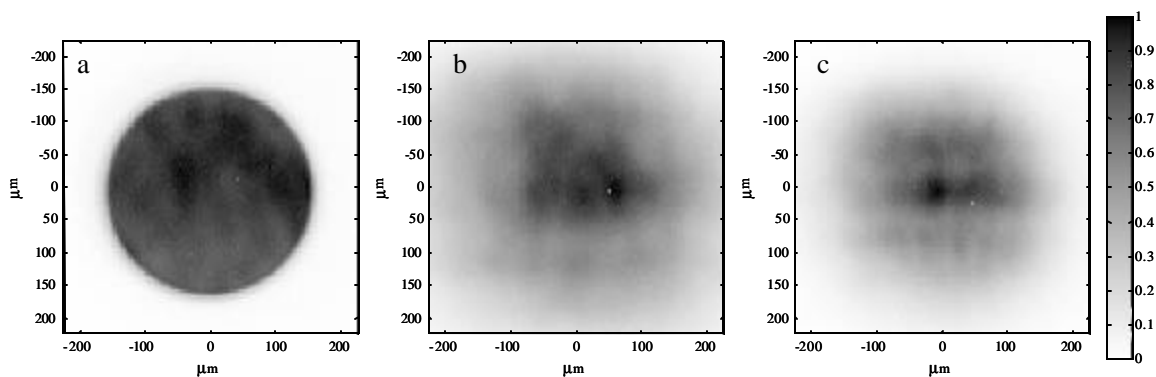


measure and quantify the reduction in spectral variability produced relative to a Gaussian profile. This is the subject of this section.

Data can easily be collected with a Gaussian profile and then compared to data collected with a flattop profile, but unless the properties of both profiles are carefully matched, the comparison will have limited value. In the experiment described here, the properties of the flattop profile were precisely measured at two different average pulse energies. “Matching” Gaussian (or at least approximately Gaussian) profiles were then carefully created. The Gaussian profiles contained the same amount of area at or above the fluence threshold as the matching flattop profiles and also contained the same amount of energy in this area. To be clear, it is not the total pulse energy that was the same, but only the energy contained in areas of the beam where the fluence met or exceeded the chosen threshold value. In this way it was ensured that the “effective” average fluence was the same in both profiles. The match actually depends upon the properties of the particles to be analyzed; *B.at.* spores were used here because their DI fluence threshold was known (and because of their general importance as a surrogate for anthrax).

As was shown above, different spores have slightly different thresholds so it was not entirely clear which exact threshold value to choose for the comparisons. Ultimately, a value slightly below the mean was chosen ( $0.95 \text{ nJ}/\mu\text{m}^2$ ) since this single value seemed best at predicting the hit rate behavior in the fixed threshold model (see Figure 33).

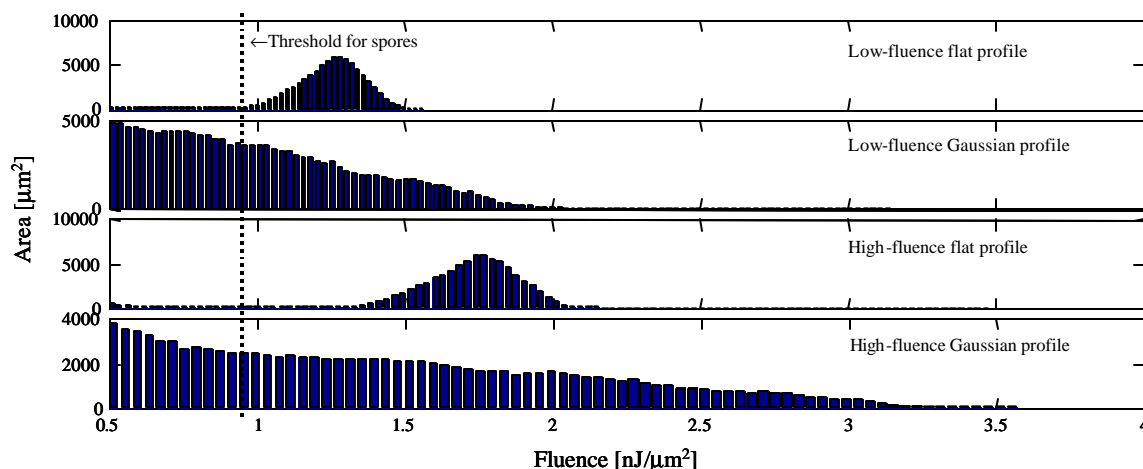
Data was collected with flattop profiles using average pulse energies of  $106 \mu\text{J}$  and  $147 \mu\text{J}$ . Twenty-one new images of the flat profile at the target plane (Figure 43a) were collected and the area in each image that had a fluence greater than or equal to  $0.95 \text{ nJ}/\mu\text{m}^2$  was calculated. The average of these areas was then found and converted to an



**Figure 43.** Sample images show a) the “flat” profile (which looks the same at any energy), b) the low-fluence “Gaussian” profile and c) the high-fluence “Gaussian” profile. These are single images, but the calculations used to determine if the profiles “matched” were based on sets of 21 images each. The maximum fluence in each image is scaled to one to enhance contrast. The true fluence distributions are shown in Figure 44.

“effective diameter” for convenience. (The effective diameter of an area or region is defined as the diameter of a perfect circle with the same area.) The effective diameters of the flattop profiles were 314  $\mu\text{m}$  and 320  $\mu\text{m}$ , respectively. The average energies contained within the above threshold regions were 97  $\mu\text{J}$  and 136  $\mu\text{J}$  (the remainder of the energy was spread over lower fluence regions). Other energies could have been used, but these produced “desirable” fluence distributions as shown in Figure 44. Notice that the two distributions are almost entirely above the  $0.95 \text{ nJ}/\mu\text{m}^2$  fluence threshold and are also fairly distinct from one another.

The optical scheme used to obtain matching “Gaussian” pulses was largely identical to the one shown in Figure 28, except that the aperture was removed and the original lens was replaced and repositioned to obtain the desired laser spot sizes at the target plane. The laser was operated at less than its maximum energy setting to produce a more Gaussian-like profile (this was, after all, the normal mode of operation in the past), but the waveplate and polarizer were still used to fine-tune the average pulse energy. To match the low-fluence flat profile, a laser spot with a diameter of  $\sim 330 \mu\text{m}$  (FWHM) and an average total pulse energy of 197  $\mu\text{J}$  was produced. This resulted in an above



**Figure 44.** Fluence distributions for the four profiles used to collect data. Each distribution is based on 21 images. Care should be taken when comparing the vertical scales because the histogram bins widths are not the same. (This was necessary to avoid aliasing effects.) The areas and the energies above the threshold are equal for the two low-fluence profiles and are also equal for the two high-fluence profiles.

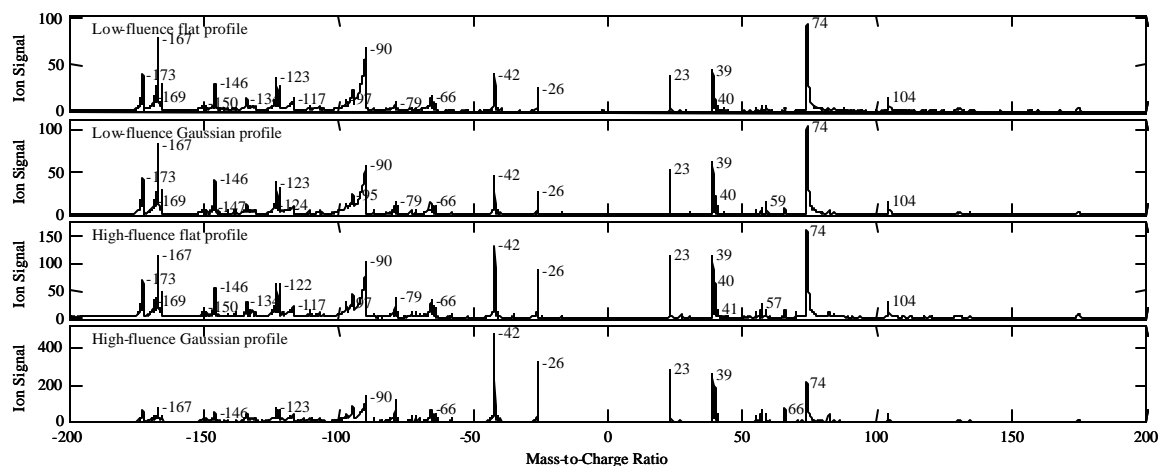
threshold region with an average effective diameter of  $312 \mu\text{m}$  containing an average of  $97 \mu\text{J}$  of energy (Figure 43b). To match the high-fluence flat profile, a  $\sim 240 \mu\text{m}$  (FWHM) spot was created with a total average energy of  $187 \mu\text{J}$ . This produced an effective diameter of  $318 \mu\text{m}$  and a contained energy of  $138 \mu\text{J}$  (Figure 43c). The fact that the total energy in this high-fluence Gaussian was less than the total energy in the low-fluence Gaussian is not surprising. Less energy was wasted in large low-fluence wings. The fluence distributions for the flattop and Gaussian profiles in both the low and high fluences cases are shown in Figure 44.

Approximately 1200 spectra were collected using the flat profile at each of the two average pulse energies. Significantly more spectra were collected with the Gaussian profiles because the pulse-to-pulse energy variations at the reduced laser setting were larger. Although these energy variations were present in the past, the purpose of this experiment was not to observe spectral differences caused by shot-to-shot pulse energy variations, but rather the spectral differences truly caused by the laser profile at constant pulse energy. Spectra produced by unusually large or small energies were discarded;

exactly 1000 mass spectra from each of the four sets of data were kept. The standard deviation of the laser pulse energies recorded for the reduced data sets were  $\sim 1.2\%$  for the flattop profiles and  $\sim 5\%$  for the Gaussian profiles. Compared with the range of fluences present in these pulses, the remaining energy fluctuations should not be significant. If the full data sets had been retained, the Gaussian profiles would only appear worse when compared with the flat profiles.

Average spectra from the four reduced sets of data are shown in Figure 45. The fact that the spectra from the low-fluence Gaussian and flattop profiles look very similar is encouraging. This is evidence that well matched profiles have been generated, in spite of the fact that the profiles look very different (Figure 43). The fact that the high-fluence spectra look so different must be expected since the high-fluence Gaussian profile produces a much larger range of fluences than the high-fluence flattop profile (Figure 44 bottom).

There are many potential ways to quantify the variability of complex mass spectra with a single parameter, but none of them are perfect. The distance metric that was



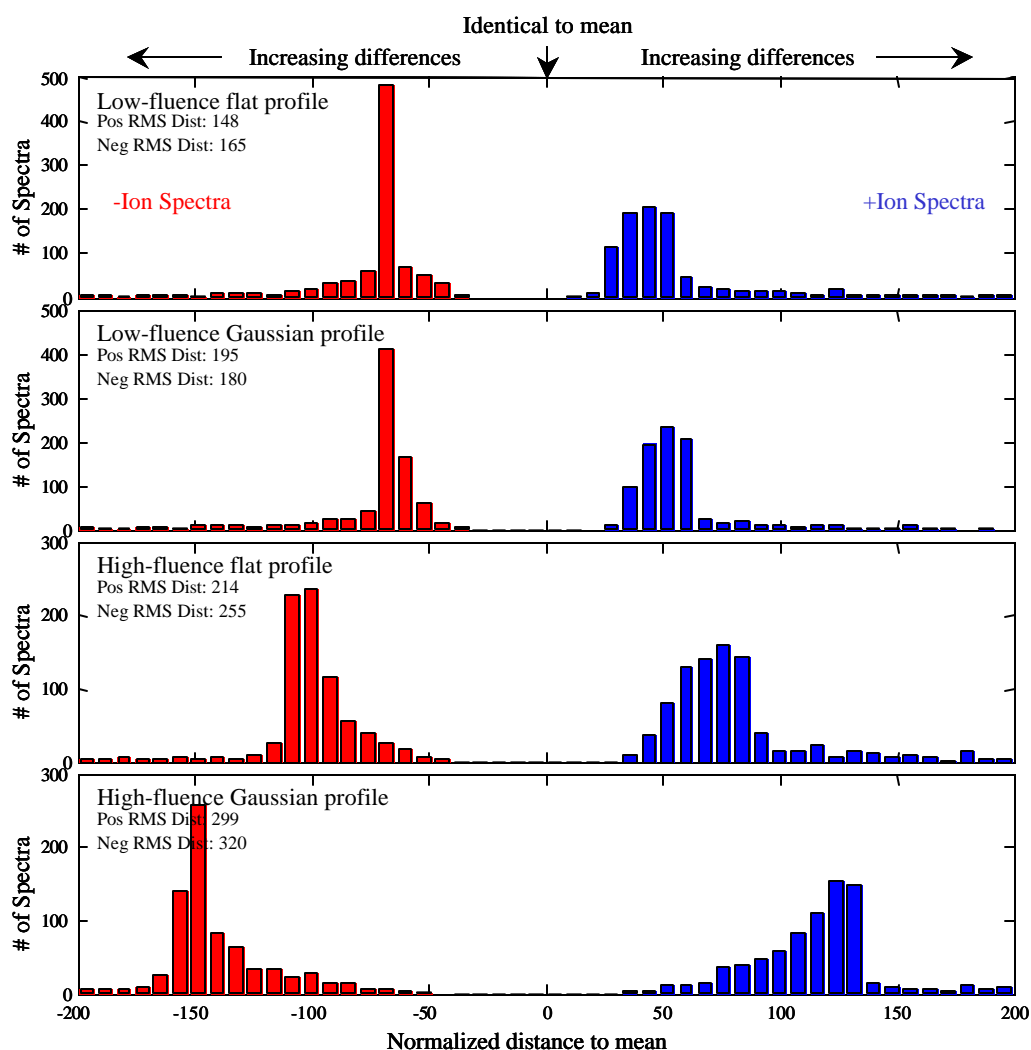
**Figure 45.** Average *B.at.* ( $\frac{1}{4}\times TY$ ) spore spectra from the four reduced sets of data used to compare the amount variability associated with the modified profiles and with the original “Gaussian” profiles.

developed in chapter 3 is used here. Each half spectrum is represented by a vector. The elements of each vector represent the areas of peaks near integer mass-to-charge ratios. The vectors at this point are identical to those used to cluster or identify particles. In the next step, however, the data is “normalized” as described in section 3.5.2. For each set of data, the mean vector is found and then each individual vector is divided, element-by-element, by the square root of the matching element of the mean vector. Many of the elements simply represent “chemical noise” or mass-to-charge ratios where no true peaks are present. Only the ten positive and ten negative vector elements with the largest mean values (based on an average of all the spectra in all data sets) are retained here. Based on these new reduced 10-element vectors, the distance from each individual vector to the appropriate mean vector is calculated (this is done separately for the positive and negative polarities). Histograms of these normalized distances are shown in Figure 46 (distances between negative ion vectors are assigned negative values for the purposes of plotting). The rms normalized distance from an individual vector to the mean vector is also calculated and labeled for each polarity of each set (this is the true metric of variability).

The flattop profiles clearly produce less variability than their matching Gaussian profiles as is observed by eye and quantified by the rms distance. It is interesting to note, however, that the variability reduction is less pronounced for the low fluence profiles than for the high fluence profiles. Figure 44 and Figure 45 provide part of the explanation. Although potentially interesting changes in peak ratios occur at lower laser energies ( $\sim 70\text{-}150\ \mu\text{J}$ ) and hence lower fluences, the biggest absolute changes in the peak heights (or more importantly areas) occur at higher energies ( $\sim 150\text{-}300\ \mu\text{J}$ ) and higher

fluences. In a certain sense, it is thus easier to make big reductions in the variability at high fluence than at low fluence. This glosses over a few subtleties involved in the proper application and interpretation of the variability metric, but at this point it is far more interesting to discuss the impact that the variability reduction will have.

The ~20% average reduction of the rms distances is more significant than it may first seem. Each mean vector defines a point in 10-dimensional space. The individual vectors



**Figure 46.** The normalized distances between the individual spectra and the mean spectrum for each data set are shown (based on the ten largest positive peaks and ten largest negative peaks). The RMS distances to the means are labeled for each set. In each case, the flattop data shows less variability than the Gaussian data. It should be pointed out that the RMS distances are affected heavily by the tails of the distributions, which extend well beyond the horizontal limits of these plots.

define points that are spread about the mean point. In the ideal case, for which the normalization scheme was rigorously developed, the individual points will be spread across a hypersphere centered at the mean point (ignoring correlations between peaks for the moment). Clearly if the hyperspheres from different types of particles overlap, it will be difficult if not impossible to differentiate the represented particle types. The number of hyperspheres that can be fit within a specific hypervolume of 10-dimensional space without overlapping is proportional to the hypervolumes of the hyperspheres, not their diameters or radii. The hypervolume of a hypersphere is proportional to its radius (i.e. the rms-distances calculated above) raised to the tenth power. A 20% reduction in the radius corresponds to a 90% reduction in hypervolume and thus may increase the number of different particles that can be discriminated by up to a factor of 10. This is not a trivial concept to grasp, but the simple conclusion is that the reduction of observed variability will likely have its greatest impact when large numbers of different types of particles must be differentiated and identified.

Knowing that a threshold fluence range exists, it is worth briefly reviewing the clustering results and conclusions about the sources of variability from chapter 4. It was correctly argued that if the most significant source of variability were truly the laser profile, then spectra created at similar fluences would tend to cluster together. Average spectra from different clusters of data would appear to be generated at different fluences. In addition, the average laser energy used to create the spectra in a cluster would provide some indication of the fluence represented by the cluster. The data collected in chapter 4 appeared to behave in precisely this manner. If natural sources of variation caused random or at least distinct types of spectral features and were dominant, the observed

behavior would not have been expected. In light of the threshold measurements, it now seems possible that natural variations could cause spectral changes that mimic the fluence dependent changes. A spore that interacts with a fluence just slightly above its threshold could create a mass spectrum that appears to be created at a lower fluence than the spectrum from another spore, with a significantly lower threshold, even if the actual fluences are identical in the two cases. Natural particle variations may be more significant than initially realized.

## 5.5 Conclusions

It has been shown that different types of biological particles have significantly different fluence thresholds when using a 266nm, 6ns laser. Although this is certainly not surprising, it is important because it places constraints on a BAMS system. If the laser energy is adjusted to obtain ideal anthrax spectra, for example, the system may be left unable to efficiently detect toxins and viruses. If the power were increased to obtain ideal spectra from toxins and viruses, spectra would still be obtained from spores, but they might retain few (if any) of the peaks used to differentiate similar *Bacillus* species. The ideal solution to this problem would be to use another DI laser system that is less sensitive to the particle type and which produces ions with higher masses. (Such a system might utilize new wavelengths, pulse lengths or even multiple lasers.) An alternative solution is to install diagnostics upstream of the mass spectrometer to provide the system with preliminary information on each particle before the DI laser is fired. The size of each particle is already known. Separate instruments exist that can detect the fluorescence from single particles excited at 266nm; they could be integrated into a BAMS system. Fluorescence data coupled with size data might provide a very good indication of the



proper laser power to use. Regardless of the approach ultimately used, any controllable source of variability must always be minimized. The results presented in this chapter represent an important step forward in this direction. A flattop laser profile clearly produces less variability than comparable Gaussian profiles.

## Chapter 6. Pulse Length and Wavelength Effects on Mass Spectra

In the last chapter, an improved DI laser profile was shown to reduce the variability of bacterial spore mass spectra. Similar improvements are expected for other particle types, but it is clear that an improved profile will never remove all of the variability. Whatever its source may be, the remaining variability in the data makes it difficult to differentiate certain types of particles. The generation and observation of higher mass ions could significantly offset the deleterious effects of this variability. The ability to produce and detect intact molecules or even large ( $\geq 1$  kD) fragments of protein, RNA, lipids and other biological molecules is strongly desired. These molecules are far more likely to be indicative of particular types of organisms than small ions like sodium, potassium and calcium, which currently dominate many of the mass spectra. A different laser-based DI system may be able to meet this challenge. This chapter describes the initial results from several alternatives to the current DI system.

At 266 nm, the single photon energy ( $\sim 4.7$  eV) is insufficient to ionize the vast majority of molecules that are likely to be encountered. Two or more photons must be absorbed, and in order for this to occur, the flux of photons must be significant. Consequently, the pulse duration (and thereby the intensity) is varied in section 6.1. In section 6.2, the pulse length is held fixed at  $\sim 6$  ns, but the wavelength is lengthened to 355 nm. There is little theoretical reason to expect that this should be more generally effective at producing large ions (since the photon energy is lower), but 355 nm is more typical of wavelengths used in MALDI and is worth exploring. In section 6.3 a wavelength of  $3.05 \mu\text{m}$  is used since past LLNL mass spectrometry experiments with

bulk bacterial spores samples and infrared DI laser pulses showed some promise. Ions with masses of many kilodaltons are not expected<sup>127</sup>, but there is reason to believe that ions with masses greater than those currently observed may be obtainable.

The current 266 nm Nd:YAG laser is not, however, the only factor limiting observation of large mass ions. The current mass spectrometer was designed primarily to study the chemical composition of common environmental aerosols of which biological particles constitute only a small fraction. Limited effort was expended toward ensuring that the spectrometer efficiently collected ions with large masses. The performance of the spectrometer is modeled in section 6.4 and its performance is shown to be less than ideal for large ions.

### **6.1 266 nm Wavelength, ~130 fs, ~6 ns and 40ns Pulse Length Experiments**

As already mentioned, two or more photons must generally be absorbed at 266 nm in order to produce ionization in common organic molecules. The time scale on which this absorption occurs is crucial. Consequently the DI laser pulse length is an important variable to consider. Numerous reports indicate that femtosecond lasers produce less molecular fragmentation than nanosecond lasers<sup>128-132</sup>. At the intensities produced by nanosecond lasers, many of the excited electronic states generated by the absorption of a single photon are able to relax before an additional photon is absorbed. In many cases, the energy from the excited electronic state is transferred to excited vibrational states. Molecules that become highly excited vibrationally are likely to dissociate. At the intensities generated by femtosecond pulses, on the other hand, two photons can be absorbed in a very short period of time thus preventing relaxation of the intermediate

excited electronic states. Photoionization and production of a molecular ion are consequently much more likely with femtosecond pulses. In spite of this, it is still interesting to see what effect a longer pulse length will have on the spectra for both scientific and practical reasons. Future versions of the BAMS system will require high repetition rate DI lasers. On-demand Q-switching at a high repetition rate is often achieved with an acoustooptical Q-switch, but this tends to result in longer pulse lengths than those currently used (~6 ns).

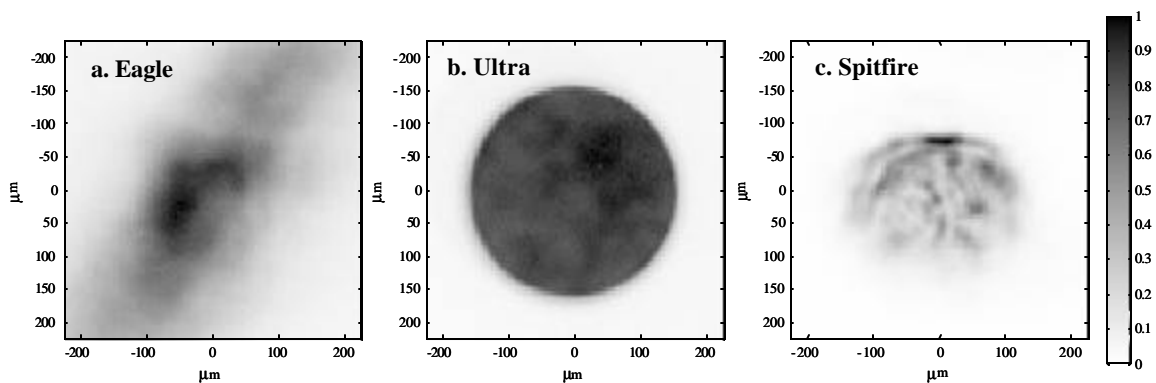
Three separate DI lasers were used to produce 266 nm laser pulses with three different durations. The first laser, a Quantronix Eagle, produced 40 ns pulses. The second laser, the “standard” Big Sky Ultra, produced ~6 ns pulses. The third laser, a Positive Light Spitfire, produced pulse widths on the order of 130 fs. A few of the basic properties of these laser systems will be described here first before discussing the beam profiles, fluence distributions and ultimately the mass spectra produced.

The Eagle was used to produce pulses with a duration of 40 ns (FWHM) and a maximum energy of ~0.5 mJ. This quadrupled Nd:YAG laser is pumped continuously with an arc lamp, which simplifies the required triggering system slightly (since there is no flash lamp to trigger). Unfortunately, the continuous pumping also requires the laser to be fired at a high average rate (~3 kHz for optimal performance). This high repetition rate is well suited for future BAMS systems, but it greatly exceeds the ability of the current ATOFMS system to track particles. The standard triggering system was consequently modified to include blank shots to ensure an acceptable firing rate. An external pulse generator produced a train of “artificial” trigger signals at 1440 Hz (a pulse every ~700  $\mu$ s). This pulse train was then combined with the real trigger signals resulting from

tracked particles. Once one of these trigger signals (real or artificial) actually fired the laser, an electronic gate prevented any of the subsequent signals from triggering the laser for  $\sim 350 \mu\text{s}$ . The fastest rate at which the laser could fire was thus  $\sim 2.9 \text{ kHz}$ , which is close to the optimal rate. Since  $350 \mu\text{s}$  of the  $700 \mu\text{s}$  between most laser shots was dead time, only  $\sim 50\%$  of the particles that were tracked were actually fired upon. This cut the rate of data collection at least in half, but this was not a significant problem for laboratory experiments.

The same Ultra and flattopping optics used in chapter 5 were used again here to produce  $\sim 6 \text{ ns}$  pulses. The Ultra's general properties were described previously (chapters 2 and 5) and will not be repeated. The same triggering scheme used in chapter 5 was also employed here.

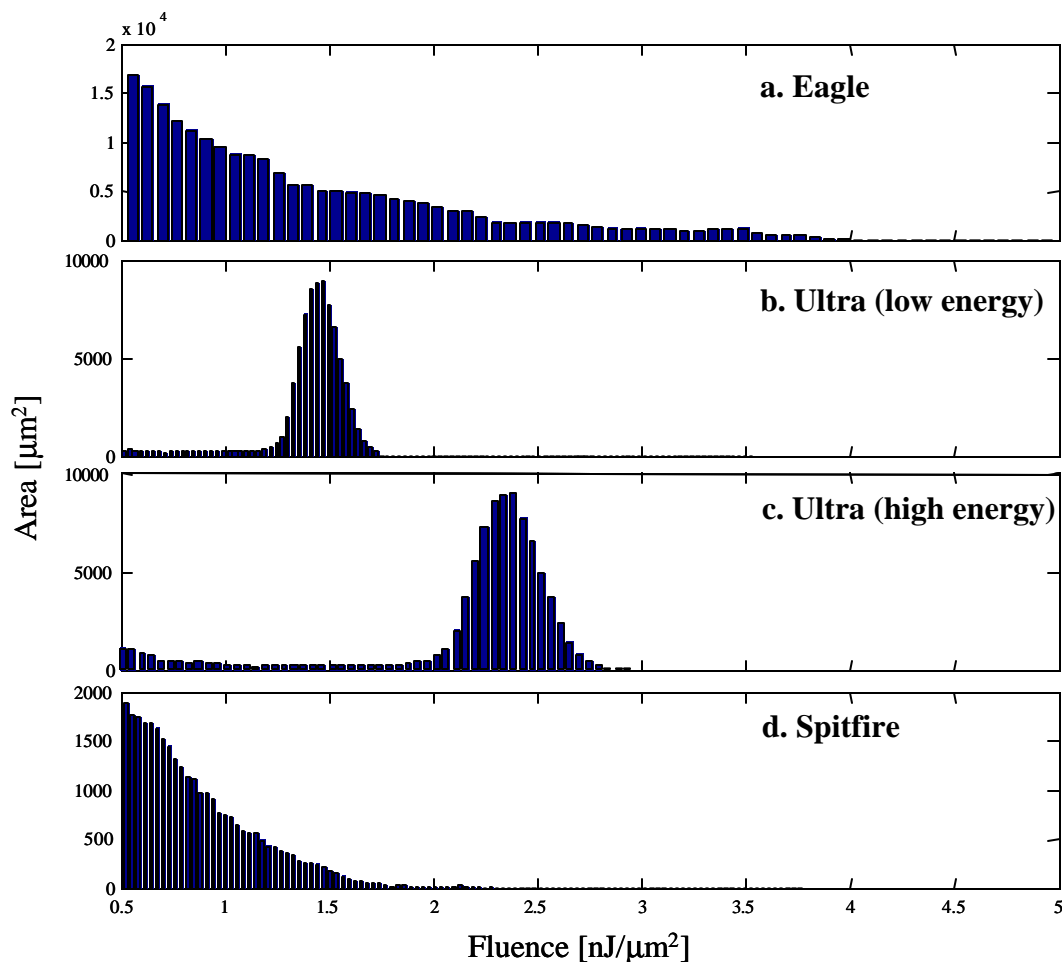
The third and final "laser" is a mode-locked, frequency-tripled, Ti:sapphire system used to produce  $266 \text{ nm}$ ,  $\sim 130 \text{ fs}$  pulses. The Spitfire mentioned earlier is actually just one component of this system (specifically, the amplifier). The Spitfire is pumped by a Positive Light Merlin and seeded by a Spectra-Physics Tsunami, which is in turn pumped by a Spectra-Physics Millennia. The Tsunami is a mode-locked Ti:sapphire oscillator. It produces a quasi-CW train of  $800 \text{ nm}$ ,  $\sim 100 \text{ fs}$  pulses at a repetition rate of  $82 \text{ MHz}$ . The pulses are used to seed the chirped-pulse regenerative amplifier contained in the Spitfire. The Spitfire is capable of producing  $800 \text{ nm}$ ,  $1 \text{ mJ}$  pulses at  $1 \text{ kHz}$ . The Spitfire pulses are converted to  $266 \text{ nm}$  with a KD\*P based frequency tripler from Spectra Physics. This tripler is noteworthy in that it uses a phase plate to obtain proper phase matching and high conversion efficiency.



**Figure 47.** Profiles of 266 nm pulses with durations of (a) 40 ns, (b)  $\sim 6$  ns, and (c)  $\sim 130$  fs. It is clear that profiles are less than ideal for the new lasers. The left and right image actually represent averages of multiple shots since the profiler was operated as though the lasers were operating in a CW rather than pulsed mode.

The specifications for the Spitfire claim a pulse length of 130 fs (at 800 nm) and this was quickly confirmed using a Swamp Optics GRENOUILLE. The 266 nm pulse width was not measured directly, since the appropriate equipment was not readily available, but it is expected to be similar to the 800 nm pulse width. In theory, tripling could shorten the initial 800 nm pulses, but dispersion and other effects tend to offset this in practice. At any rate, the 266 nm pulse length will be referred to as  $\sim 130$  fs to distinguish it from other pulse lengths in the present discussion. The whole system was triggered using the same scheme described for the Eagle above, with the exception that the various times were scaled to produce a repetition rate of 500 Hz in the absence of tracked particles.

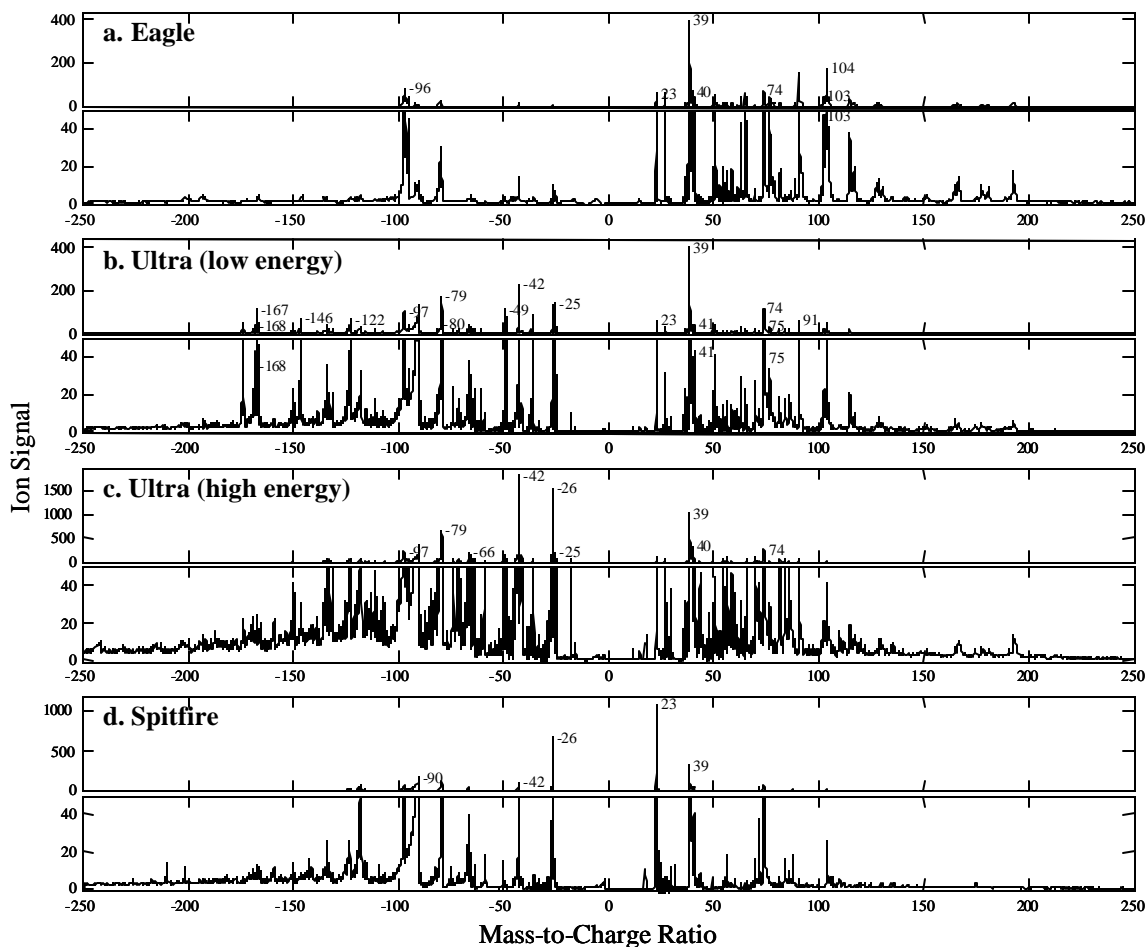
An image of the Eagle laser profile at the target plane is shown in Figure 47a. It is elliptical in shape with an intense central region and extensive wings. Since the primary purpose of these experiments was to look for gross differences between spectra no effort was made to improve the profile. The effective diameter of the pulses at the target plane was 240  $\mu\text{m}$  (FWHM). The mean pulse energy at the target plane was 360  $\mu\text{J}$  with a standard deviation of  $\sim 10\%$ . The flattened Ultra profile is shown in Figure 47b. As before the diameter is approximately 320  $\mu\text{m}$  (FWHM) at the target plane. Two separate pulse



**Figure 48.** Fluence distributions for the a) 40 ns pulses, b) ~6 ns, 123  $\mu\text{J}$  pulses, c) ~6 ns, 199  $\mu\text{J}$  pulses, and d) ~130 fs pulses.

energies (123  $\mu\text{J}$  and 199  $\mu\text{J}$ ) were used in the experiments for reference purposes. The profile of the tripled Spitfire pulses is shown in Figure 47c. The pulses had a reasonably good profile when they first left the Spitfire's regenerative amplifier, but it was degraded significantly by the gratings in the Spitfire's pulse compressor. The structure in the profile was then further accentuated by the nonlinear processes required for frequency conversion in the frequency tripler. The final ultrashort pulses at 266 nm had a mean energy of 38  $\mu\text{J}$  and a standard deviation of 13  $\mu\text{J}$  at the target plane.

The distribution of fluences in the Eagle profile is shown in Figure 48a. Fluence distributions for the two Ultra energies are shown in Figure 48 (b and c). The fluence



**Figure 49.** Average *B.at.* spore spectra. a) Average spectrum obtained using 40 ns pulses (lower spectrum is shows an expanded vertical range). b) Average spectrum obtained using  $\sim 6$  ns, 123  $\mu\text{J}$  pulses. c) Average spectrum obtained using  $\sim 6$  ns, 199  $\mu\text{J}$  pulses. d) Average spectrum obtained using  $\sim 130$  fs pulses.

distribution in the Spitfire's profile is shown in Figure 48d. Only a small fraction of the area in the Spitfire profile is above the spore fluence threshold determined for 6 ns pulses, but the intensities produced by the ultrashort pulses are orders of magnitude higher than for the 6 and 40 ns pulses. It is interesting to note that the Spitfire's fluence distribution is qualitatively similar in shape to that produced by the Eagle, even though the profiles have very different appearances.

Approximately 250 *B.at.* spore spectra were collected using the Eagle and averaged to produce the spectrum shown in Figure 49a. Spectra were collected with the Ultra on the same day using the same preparation of spores. Two sets of data (containing  $\sim 250$  spectra



each) were collected using the two different pulse energies mentioned earlier (123  $\mu\text{J}$  and 199  $\mu\text{J}$ ). Their averages are shown in Figure 49 (b and c). One thousand spectra were collected from *B.at.* spores using the Spitfire and averaged to produce the spectrum shown in Figure 49d (other applications, not discussed here, required more spectra to be collected than usual).

Although peak heights and ratios vary significantly with the pulse length, it is not difficult to see that most peaks produced using either the 130 fs or 40 ns pulses are also produced, to some extent, using the standard 6 ns pulses. This was not entirely expected, but a potential explanation can be provided. As was mentioned in chapter 5, ion formation in UV MALDI is thought to result from a two-stage process. Primary ions are generated more or less directly by the laser and must as a result depend on the particular properties of the laser (i.e. pulse duration, wavelength, etc.). In the plume, however, the production of secondary ions is largely governed by thermodynamics. These secondary reactions can thus conceal differences in the formation of the primary ions. It is also important to keep in mind that spores contain only a finite variety of molecules. It should not be particularly surprising, therefore, that different ionization schemes may produce many of the same ions, just in different amounts (particularly when the same wavelength is used). If separate pulses were used for desorption and ionization, it might be possible to exert much greater influence over the spectra produced<sup>133-137</sup>.

Perhaps the most striking feature of the spectra is that the 40 ns pulses appear to produce relatively few negative ions in comparison to the number of positive ions. This cannot yet be quantitatively explained, but it will be investigated more thoroughly in future experiments. Another interesting property of the 40 ns spectra is that many of the

peaks observed in the raw time of flight data are narrower than the laser pulse width. It is not difficult, in fact, to find fairly substantial peaks in individual spectra that have a width of 20 ns (FWHM) or less. The static fields employed in the mass spectrometer do not have the ability to “time focus” ions thus the laser pulses seemingly must produce free ions over a briefer window than the pulse duration. This may be consistent with the tentative hypothesis that ions are produced primarily in the spore core but cannot be observed until enough energy has been absorbed to rupture the outer spore layers. It could also be related to some other type of threshold behavior.

No large peaks appear in the mass spectra outside of the mass range shown in Figure 49. A triplet of very small peaks, with spacings of 14 Daltons, can be observed in the 6 ns data centered at approximately  $m/q = -710$ . These may be consistent with a lipid (the peak spacings in this case would correspond to the loss of one or two  $\text{CH}_2$  groups). The largest peak is about 0.6 units tall in the spectrum shown in Figure 49b. If it were not for the fact that these three peaks are observed in data from other experiments, it might be tempting to ignore them. There is perhaps an even smaller peak at approximately  $m/q = -1245$  in the 130 fs spectra, but more data is needed to confirm its existence. Whether the peak is real or not, all of these “high mass” peaks are far too small (in terms of signal) to be of value for identification with the current instruments.

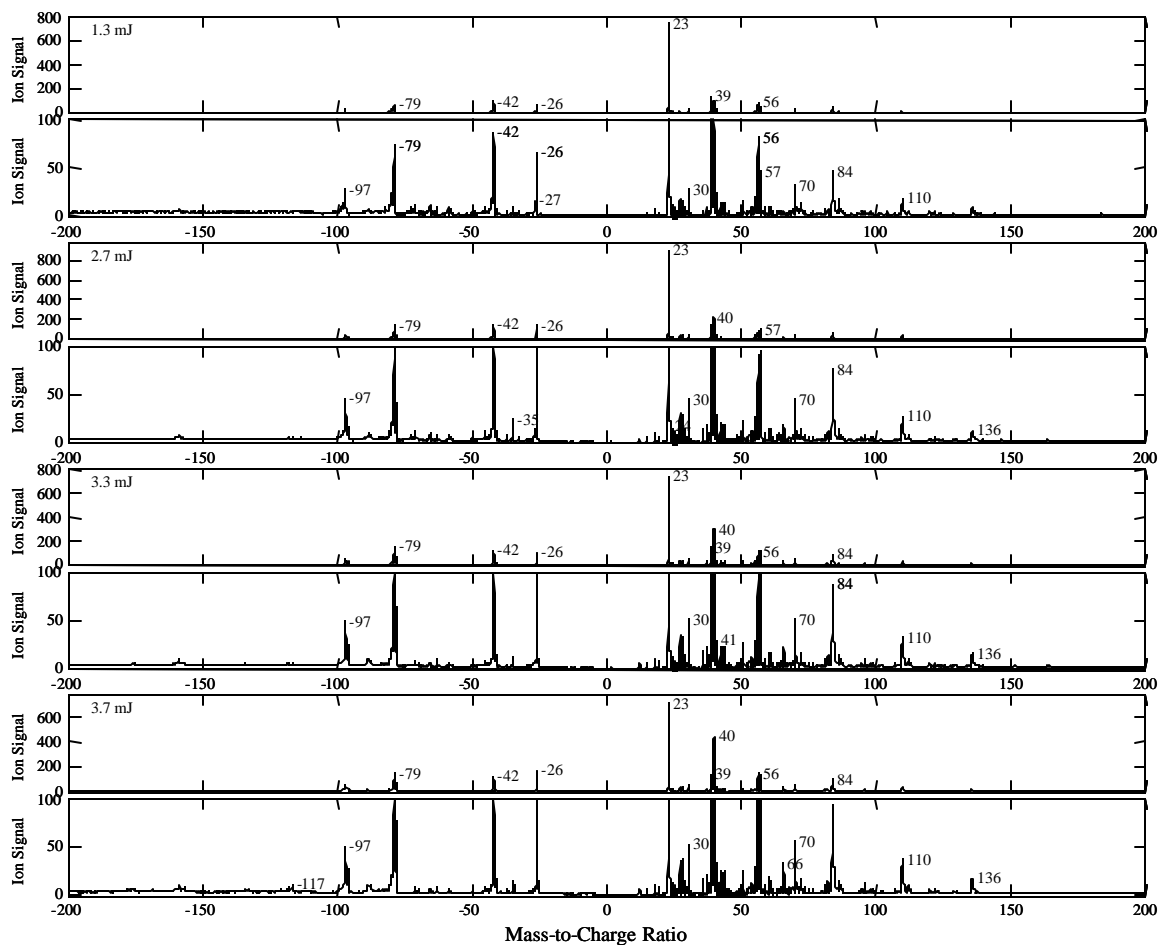
The use of 266 nm lasers with different pulse lengths did not result in the observation of significant new ion peaks at masses greatly in excess of those seen previously (with the standard 266 nm, 6 ns laser). This does not necessarily mean, however, that high mass ions were not created. As will be shown in section 6.4, such ions must be generated in large numbers to produce an obvious peak in a spectrum. Whether new high mass ions

were generated or not, the 130 fs and 40 ns lasers are able to operate at a far higher rate than the 6 ns laser and may be of some value for this reason.

## 6.2 355 nm Wavelength, ~6 ns Pulse Length Experiments

Common mass spectrometry wavelengths include 308 nm (XeCl), 337 nm (N<sub>2</sub>), 351 nm (XeF) and 355 nm (3×Nd:YAG) among others. Gieray et al. used a XeCl excimer laser for laser ablation mass spectrometry of single *B.at.* spores<sup>75</sup>. Several of the single particle MALDI experiments utilized N<sub>2</sub> lasers<sup>85, 86</sup>. A tripled Nd:YAG (355 nm) laser was used here to explore the wavelength dependence of spore mass spectra. The particular laser utilized is virtually identical to the standard 266nm laser (used in chapter 4). Both are Ultras from Big Sky Laser Technologies; the only difference is the harmonics package. 355nm photons are not absorbed strongly by DPA or any of the amino acids, so significant differences are expected between the spectra produced at 355 nm and the earlier spectra produced at 266 nm.

Figure 50 shows average spectra collected at 355 nm from *B.at.* spores using 4 different average pulse energies: 1.3, 2.7, 3.3 and 3.7 mJ. Compared to the energy induced spectral changes at 266 nm, these spectra appear much more consistent. At fluences far above threshold, the 266 nm spectra were observed to become more consistent, but the fluences produced at the lowest energies here cannot be far above the 355 nm threshold. The hit rate at the lowest energy in the figure is less than 2% and grows to almost 14% for the highest energy. Unfortunately, equipment was lacking to take images of the laser profile at the time these spectra were collected so it was not possible to properly measure the size or shape of the unmodified laser beam (which is



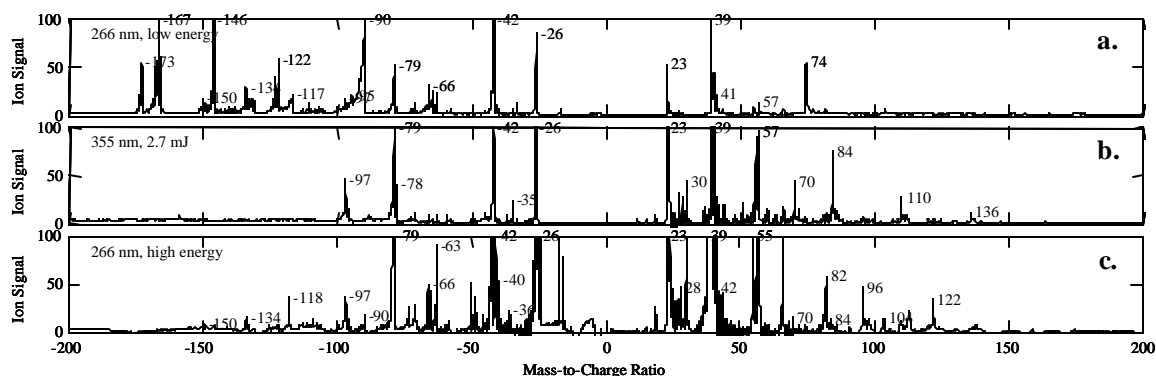
**Figure 50.** Average mass spectra from *B. at.* spores collected using a 355 nm laser at four different average pulse energies. For each pulse energy two spectra are shown: one full vertical scale, and one expanded scale.

necessary to determine the true fluence distributions produced). Nonetheless, the profile is known to have been roughly Gaussian in form. The diameter of the pulses at the target plane was  $\sim 500 \mu\text{m}$  (within a factor of 2).

Given the uncertainties in the true shape of the profile, it is not possible to accurately determine the 355 nm fluence threshold. Nonetheless, it is possible to determine a lower limit for its value. Based on the hit rate data, most of the area in the profile of lowest-energy, 355 nm pulse must have a fluence below the true threshold. Even if the 1.3 mJ of energy contained in the low energy pulse were spread uniformly over a 1 mm diameter, the mean fluence would still exceed  $1.6 \text{ nJ}/\mu\text{m}^2$ . The true fluence threshold must be

higher. This increased threshold (relative to the value of  $\sim 1 \text{ nJ}/\mu\text{m}^2$  at 266 nm) is at least qualitatively consistent with absorption data that has been obtained for *B. subtilis* spores suspended in water<sup>121</sup>. The absorption coefficients at 265 nm, 300 nm and 400 nm are  $6540 \text{ cm}^{-1}$ ,  $6200 \text{ cm}^{-1}$ , and  $5560 \text{ cm}^{-1}$  respectively. No unusual behavior is expected between 300 nm and 400 nm so the absorption coefficient at 355 nm is expected to be roughly  $5900 \text{ cm}^{-1}$ . More energy is absorbed at 266 nm (for a given fluence) and so the fluence threshold at 266 nm is expected to be lower.

Of greater relevance to the BAMS system is the pattern of ions formed. No significant peaks were observed outside of the mass range shown in Figure 50. Virtually all of the peaks produced at 355 nm were also produced at 266 nm, although some of the peaks are significantly smaller. Figure 51 shows two average spectra collected at 266 nm on either side of the second 355 nm spectrum from Figure 50. The peaks attributed to DPA, arginine (-173), glutamic acid (-146) and aspartic acid (-134), which are present in the low energy 266 nm spectrum, are gone at 355 nm. In reality, this is not surprising since DPA and the aromatic amino acids do not absorb at 355 nm. Although the primary amino



**Figure 51.** The top (a) and bottom (c) average spectra were collected using 266 nm pulses with relatively low and high energies respectively. The middle average spectrum (b) was collected using 355 nm pulses. It appears more similar to the high energy 266 nm spectrum, but the fluences produced could not have been far above the 355 nm fluence threshold. Virtually every peak produced at 355 nm can be identified in the 266 nm spectra. The spectra have been vertically cropped to make the small peaks more clear.

acids observed at 266 nm were not actually aromatic (and therefore did not absorb strongly at 266 nm) other members of the BAMS group have shown that the presence of aromatic amino acids and DPA can enhance the production of non-aromatic amino acid ions at 266 nm which probably explains their presence in the 266 nm spore spectra<sup>138</sup>. It is also interesting to note that two 266 nm photons provide enough energy to ionize the majority of amino acids<sup>125</sup>, while two 355 nm photons do not.

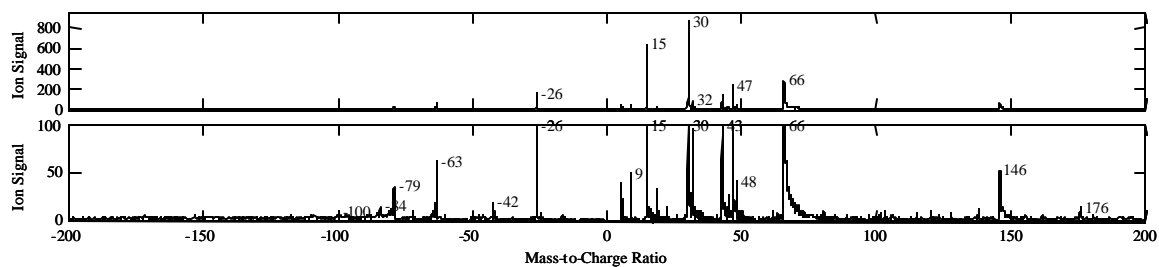
It is encouraging that the differences between the 355 nm spectra and the 266 nm spectra can be qualitatively explained. As rather expected, however, the 355 nm spectra do not appear to be as useful as the 266 nm spectra for differentiating particles since important marker peaks are absent and no significant new peaks have appeared. In applications where the use of reagents is not a disadvantage, however, the 355 nm laser may still be of some use. Preliminary experiments by other members of the BAMS group indicate that adding 2,5-Dihydroxybenzoic acid (DHB), a common MALDI matrix, to a solution can greatly enhance the sensitivity of the instrument to certain analyte molecules at 355 nm. Perhaps this may hold some promise for mass spectrometry on single cells as well.

### **6.3 3.05 $\mu$ m Wavelength, ~4 ns Pulse Length Experiments**

As an alternative to the ultraviolet wavelengths used hitherto, it is interesting to consider infrared wavelengths. Whereas UV photons tend to produce excited electronic states, IR photons produce excited vibrational states. (The energy of a 3050 nm photon is only ~0.4 eV.) Early work performed at LLNL by some of the founding members of the BAMS group showed that mass spectra could be obtained from collections of whole

bacterial spores deposited on a metal substrate without the addition of a chemical matrix<sup>139</sup>. An N<sub>2</sub> laser was used to collect some spectra at 337 nm, but superior results were obtained using infrared wavelengths (3.05 to 3.80 μm) from a custom-built optical parametric oscillator (OPO). At 3.34 μm it was possible to generate an ion signal near  $m/q=19,050$ , along with numerous smaller peaks, and clear differences between *Bacillus* species were observed. Those measurements, however, required many spores; the sensitivity limit was  $\sim 2 \times 10^4$  spores, but good quality spectra required  $\sim 10^7$  spores. Since the single particle instruments used here demonstrate single spore sensitivity at UV wavelengths, new experiments were performed with IR wavelengths.

The experiments described here employed an Opotek IR OPO 2732. This is an infrared optical parametric oscillator pumped by a Q-switched, Nd:YAG laser. It uses a KTiOAsO<sub>4</sub> (KTA) crystal to produce  $\sim 4$  ns pulses of light continuously tunable from 1.55 to 1.75 μm (signal) and 2.7 to 3.2 μm (idler). The signal is discarded and work is performed with the idler ( $\sim 5$  mJ/pulse max). The pump is a Brilliant from Quantel. Unlike the Ultra, the Brilliant's pulse energy is adjusted by changing the delay between the flashlamp and Q-switch. This results in a more consistent profile at the fundamental wavelength of 1064 nm. Unfortunately, the laser requires itself to fire at an average rate of  $\sim 10$  Hz (presumably to maintain a stable output). This is not directly compatible with the ATOFMS triggering system, which simply fires the laser whenever an aerosol particle is tracked. To remedy this situation, the triggering system was modified to generate artificial trigger signals when particles arrived at an insufficient rate and gate out trigger signals that occurred too rapidly after a preceding trigger (the scheme was similar, but not identical, to that used for the Eagle described in section 6.1). In general this



**Figure 52.** An average of 83 individual spectra collected from an aerosolized solution of *B.at.* spores at 3050 nm. The bottom spectrum is simply an enlarged version of the spectrum at top. The IR laser did not efficiently produce mass spectra so only a small set of data was collected.

worked quite well, but the rate of data acquisition was still reduced ( $\sim 2\times$ ) because of the necessary dead time following each laser shot.

The profile of the IR pulses was not ideal, but since pulse energy could not be sacrificed (as explained shortly) no attempt was made to modify it. A Spiricon Pyrocam beam profiler confirmed that the profile could be made approximately Gaussian, with a diameter of  $\sim 300\ \mu\text{m}$  (FWHM), when properly focused at the target plane. The beam was highly divergent upon emission from the OPO and at some wavelengths absorption by the atmosphere, presumably due to water, was noticeable. The maximum pulse energy at the target plane was about 2.5 mJ at 3050 nm, which is a wavelength close to an absorption peak in water and presumably other OH containing molecules. Figure 52 shows an average spectrum collected under these conditions from an aerosolized solution of *B.at.* spores. Far more significant than the actual spectrum is the fact that more than 28,000 particles had to be tracked in order to obtain the 83 spectra represented by the average. This hit rate of less than 0.3% was the best that could be obtained after significant care was taken to align the IR laser. In contrast, the hit rate for spores at 266 nm can approach 30% at high pulse energies.

The reason for the low hit rate is not fully understood. It is possible that the fluences produced by the IR laser pulses were simply below the DI threshold at this wavelength,



except for a few hot spots in unusual pulses. It also is possible that the true DI threshold for intact spores at 3050 nm is much higher than any of the fluences produced and that the resulting spectra are from damaged or atypical particles. The early IR-LDI work at LLNL shows that spectra from spores on a substrate can be produced using 1.5 mJ, 3050 nm pulses focused to a spot size of  $0.9 \text{ mm}^2$  (area at or above half of maximum intensity)<sup>139</sup>. This represents an average fluence of  $\sim 1.7 \text{ nJ}/\mu\text{m}^2$ . This is of the same order of magnitude as the fluences typically used in IR MALDI<sup>140</sup>. In the current IR single spore experiment, a  $\sim 2.5 \text{ mJ}$  pulse is focused to  $\sim 300 \mu\text{m}$  producing a fluence of  $\sim 35 \text{ nJ}/\mu\text{m}^2$ . This is obviously much higher than  $1.7 \text{ nJ}/\mu\text{m}^2$ , but because an isolated single spore is significantly smaller than the IR laser wavelength, the absorption may be affected and care must be taken in comparing the two values.

The energy absorbed by a spore depends on its complex index of refraction. The approximate complex index of refraction for *B.at.* spores at a wavelength of  $\sim 3 \mu\text{m}$  is  $n=1.45+0.06i$ <sup>123</sup>. A Mie calculation based on this value indicates that the absorption cross section of a  $1 \mu\text{m}$  diameter spore is  $0.13 \mu\text{m}^2$  ( $\sim 1/6$  of its geometrical cross section) at a wavelength of 3050 nm. At a fluence of  $35 \text{ nJ}/\mu\text{m}^2$  the spore should absorb  $\sim 4.6 \text{ nJ}$  of energy (assuming that the index of refraction stays fixed throughout the laser pulse). This is more than enough energy to desorb molecules. At the 266 nm fluence threshold, a  $1 \mu\text{m}$  spore is only estimated to absorb  $0.47 \text{ nJ}$  (chapter 5). The relative lack of spectra (i.e. low hit rate) at 3050 nm does not actually indicate that molecules were not desorbed; it merely indicates that free ions were not formed. It is possible that large numbers of free neutral molecules were produced (they simply were not and could not be detected). There

is some evidence to support this from initial experiments performed using two DI laser pulses (an IR pulse for desorption followed rapidly by a UV pulse for ionization).

In trying to better understand the low observed hit rate, it is useful to briefly consider current theories of IR MALDI. It is not unreasonable to expect that some of the molecules in spores may act as a MALDI-like matrix for other molecules. Some of the DI processes involved here in single spores may thus be similar to processes occurring in standard IR MALDI. In IR MALDI, the desorption/ablation process is strongly tied to the properties of the matrix and the laser. Spallation, various forms of photoablation and thermal processes have all been proposed as possible desorption mechanisms<sup>140</sup>. In spores, thermal desorption processes are perhaps the most likely candidate. Even without precise knowledge of the material properties of spores, it seems doubtful that all of the other processes can be relevant. Spallation, for example, occurs when the laser deposits energy in a surface layer faster than thermal conduction allows it to be removed. Since lower layers remain basically unheated, significant stresses may develop. A spore is so small, however, that energy absorption is likely to be relatively uniform. The small size also means that even if temperature gradients develop, they will likely diffuse away very rapidly. Nonetheless, the energy cannot escape from the spore altogether so it seems probable that some desorption must occur.

The mechanisms that form primary ions in IR MALDI are even less well understood than the desorption/ablation processes. The photon energy is so low (~0.4 eV at 3050nm) that direct photoionization is highly unlikely. It has been suggested that spallation, cavitation, and explosive phase transitions may cause sufficiently energetic events to produce ions<sup>140</sup>. It is possible that these mechanisms do not operate effectively in spores

(hence explaining the low hit rate). It is beyond the scope of this thesis, however, to quantitatively evaluate the roles that these processes may play. Once primary ions are formed in IR MALDI, they interact with other molecules in the plume to produce secondary ions. Because of the small size of single isolated spores, the plume density is likely to drop rapidly. Whereas a MALDI plume from a large spot may undergo an almost one-dimensional expansion, the spore plume is free to expand in all three dimensions. The production of secondary ions is almost certainly limited, therefore, in comparison with experiments that employ bulk samples (e.g. the past IR experiments at LLNL). This may partially explain the lack of high mass peaks seen in the few mass spectra successfully collected here (Figure 52) compared to the past LLNL IR results. On a more practical note, however, it is very important to consider the efficiency of the current mass spectrometer as a function of ion mass. The mass spectrometer used for the past experiments is believed to have been more sensitive (at least at high mass) and was given a far larger sample to analyze.

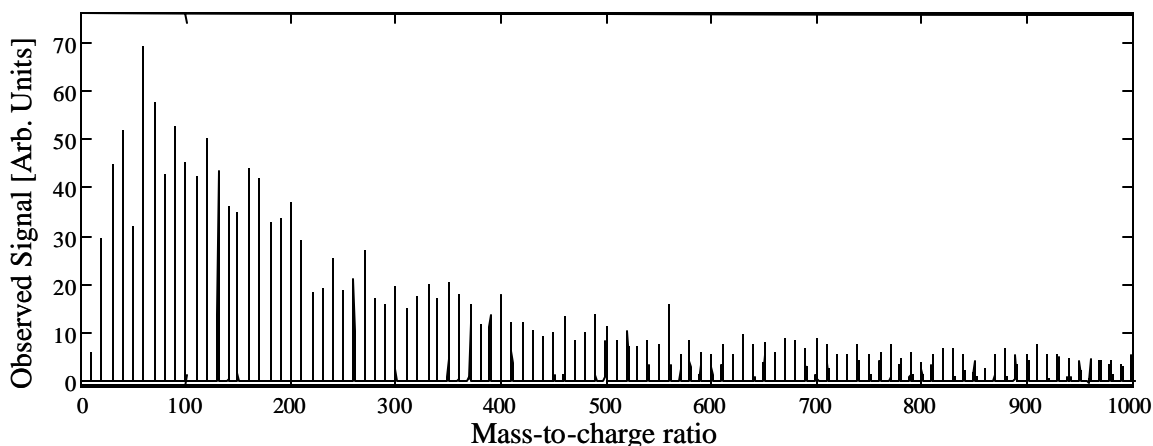
#### **6.4 ATOFMS Performance for High Mass Ions**

The instrument used for the experiments described here is basically a commercial aerosol time-of-flight mass spectrometer originally designed for the chemical analysis of organic and inorganic aerosol particles. As such, the instrument was designed primarily to detect relatively small ions and not the larger ions that are of greater interest here. Since large ions have relatively long flight times, a large ion with a given initial radial velocity (i.e. a velocity perpendicular to the axis of the spectrometer) will have a larger radial displacement at the MCP detector than a small ion with the same initial radial

velocity. The instrument can clearly only detect an ion if that ion actually hits one of the detectors. Large ions are more likely to miss the detector and thus remain unobserved.

In chapter 3, a simple (but lengthy) equation for the time of flight of an ion in the mass spectrometer was derived. It is a relatively straightforward extension of this calculation to predict the radial position (i.e. the transverse displacement) of an ion in any plane perpendicular to the mass spectrometer axis. This is done here to determine whether an ion (with a specific mass, charge and initial velocity) will hit the active area of the detector or even escape the ion source region. The efficiency of the mass spectrometer at high masses is in turn determined by tracking many ions with a range of initial properties.

For the present calculations, it is assumed that the radial velocities of the ions created from an aerosol particle result purely from the kinetic energy imparted by the DI event and the initial downward velocity of the aerosol particle (~300 m/s). Although the particle is broken apart, the center of mass of the desorbed material (and any remnant bulk material) must continue traveling downward with the same velocity. The electric fields inside the mass spectrometer are treated as purely axial and do not change the initial radial velocities of the ions. Significant work has been done to determine the velocities of ionized and neutral analyte and matrix molecules in MALDI experiments<sup>141</sup><sup>142</sup><sup>143</sup><sup>144</sup>. The velocities depend on the choice of analyte and matrix, but values of 500-1000 m/s appear typical. The angular distribution of ion velocities caused by the DI event may not be spherically symmetric<sup>145, 146</sup>, but since the true distribution cannot be measured without significant modification of the instrument, a spherically symmetric distribution (relative to the particle's center-of-mass) will be used here as a simple approximation. The x, y, and z components of each ion's velocities are all described by



**Figure 53.** An artificial spectrum shows the reduced efficiency of the mass spectrometer at high mass. Each “peak” results from 100 ions initially located at the center of the ion source region. At high mass, however, the ions of a particular mass-to-charge ratio are dispersed and no longer form a single peak. When superimposed with chemical noise and shifted by calibration jitter, the individual peaks may be very hard to identify and properly associate with one another.

normal distributions with means of zero and standard deviations of 500 m/s. The particular values are chosen by a normal random number generator (the “randn” function in MATLAB).

To produce an artificial spectrum, the flight times for 100 ions (randomly assigned initial velocities conforming to the distributions described above) are calculated at every tenth integer mass-to-charge ratio out to a maximum of 1000 Daltons. A histogram with 2 ns time bins is made from the time of flight data (the MCP response is assumed to be instant with no temporal width). As discussed in chapter 3, the MCP detectors produce a broad range of responses to single ions. The distribution is roughly Gaussian with a standard deviation equal to ~40% of the mean. For plotting purposes, the mean response of the MCP was simply assumed to be one in arbitrary units. (The specific response to a given ion or ions is determined using again the “randn” MATLAB function.) A calibration equation of standard form (Eq. 9) was fit to the mass-to-charge ratio and flight time data and then used to produce a proper mass scale for the spectrum. An example is shown in Figure 53.

If the mass spectrometer were ideal, all of the ions of a particular mass would arrive at the detector at exactly the same time regardless of their initial velocities or even positions (this would constitute perfect energy and space focusing, respectively). Each peak in the artificial mass spectrum would consequently have an average height of 100 units. If the detectors were perfect, all of the peaks would be exactly 100 units tall. Imperfect detectors inevitably cause some fluctuations. For masses near 50 Daltons, the model indicates that the mass spectrometer should perform fairly well (Figure 53). At high masses, however, the few ions of a given mass that actually make it to the detector no longer form a single peak. When superimposed with random ion signals (i.e. chemical noise) and shifted randomly by calibration jitter, the individual peaks will be very hard (if not impossible) to identify and properly associate with one another.

This shortcoming of the mass spectrometer is particularly unfortunate because there is no guarantee that a single particle will even produce 100 ions at a single high mass-to-charge ratio. With optimal conditions in a MALDI experiment, less than 1 in 1000 of the analyte molecules are likely to be ionized<sup>124</sup>. In the present case of single spores, the ionization efficiency appears to be much worse. A typical spore contains on the order of  $4 \times 10^8$  DPA molecules, but the corresponding peak at  $m/q=167$  is thought to represent no more than a few hundred ions at best. For very large molecules, such as DNA, there may be only one or a few copies in the entire aerosol particle.

The software that operates the current mass spectrometer imposes an additional limitation that is not directly related to the design of the mass spectrometer. The system is configured to collect data for 60  $\mu$ s after the laser is fired which means that ions with

masses greater than ~1500 will never be observed (the necessary MCP data simply isn't collected). This cannot currently be changed.

It is clear that a new mass spectrometer is needed. The new spectrometer must transport high mass ions to the detectors much more efficiently, focus the ions better (in time), have reduced calibration jitter and of course acquire data over an adequate range of flight times. Such a mass spectrometer is currently under construction by the BAMS group.

## 6.5 Conclusions

Ion peak heights and ratios vary significantly with the wavelength and pulse length of the laser used for desorption and ionization. It is conceivable that some of the new spectral features encountered with the "non-standard" lasers might enable better differentiation of certain particle types, but no significant new high mass peaks were observed. The cause for this is not certain. It is possible that the new lasers simply did not produce such ions. It is also possible that some high mass ions were generated but simply not detected. The current mass spectrometer has been shown to perform poorly at high mass. Unless large ions are produced in large numbers, the ions are not likely to produce clear peaks in the mass spectra obtained with the current system. Fortunately, a new mass spectrometer is being built and even more advanced systems are planned. As soon as one of these systems becomes available, the experiments described here can be repeated with the new system. Regardless of the mass spectrometer used, it is important to better understand the chemical reaction dynamics within the desorption plume. These reactions are critical to the success of standard MALDI, but their role here is unclear and may vary

significantly with the laser parameters. The use of a second laser pulse to probe the plume is likely to provide very informative data. The use of a second laser pulse may also allow far greater control over the ions produced.



## Chapter 7. Performance Predictions

### 7.1 Introduction

The goal of the ongoing work described in the previous chapters is ultimately to produce mass spectra that enable accurate differentiation and identification of individual bioaerosol particles. There can be no question that this is essential to the BAMS system; better identification of individual particles will lead to better overall system performance. Nonetheless, the ability to accurately analyze individual particles is not a guarantee that a useful bioaerosol detector can be achieved. Two important system parameters that must be quantified are the probability of false alarm and the probability of detection. It is easy to produce a detector that has a low false alarm rate or a high detection probability, but it is very difficult to obtain both in a rapid (or even a slow) detector. The particular requirements for sensitivity, specificity and speed will be determined by the details of each specific deployment or application scenario encountered. It is thus critical that the performance of the BAMS system be modeled quantitatively and that the relationships between these three parameters and their dependence on the performance of the individual components of the BAMS system be understood fully. Such a model is the subject of the current chapter.

The comprehensive model described here requires as input a number of aerosol particle properties, performance parameters for individual BAMS system components and at least a partial set of user requirements concerning the desired detection probability, detection time and false alarm rate. The aerosol particle properties are first combined with the instrument component parameters to determine the “efficiency” of the total

BAMS system for each particle type in the local aerosol. This efficiency links the concentration of each particle type to the number of such particles measured per minute. The single particle identification and misidentification rates (which are known or measurable) are then used to calculate the average number of particles that are correctly and incorrectly identified as agent. A statistical model finally determines the actual detection limit and expected false alarm rate. A statistical model is necessary because there is, for example, always a possibility that no agent particles will be drawn into the instrument and analyzed in a given sampling period even though agent particles are present in the aerosol. Similarly, there is always a finite probability that one or more harmless particles will be misidentified as agent in any given period of time. The statistical model quantifies these probabilities.

Fortunately, an alarm need not be sounded on the basis of the identification of a single particle as agent. In an environment that is very clean, the identification of only “a few” particles as a specific type of agent in a short period of time might be required to sound an alarm. In an environment filled with particles that are similar to an agent of interest, but harmless (e.g. *B.t.* when searching for anthrax), a greater number of agent identifications would be required. Setting the “alarm threshold” properly is critical. The threshold is defined here as the number of particles identified as agent per sampling period required to sound an alarm. If this threshold is set too high, the instrument loses sensitivity. If the threshold is set too low, too many false alarms will result. One of the outputs from the model (the ROC curve discussed later) explicitly shows this tradeoff.

The requirements to sound an alarm will generally be different and largely independent for each detectable agent type. The fact that a few particles may be identified

as plague in a hypothetical scenario does not necessarily change the requirement to sound an alarm for anthrax any more than the identification of a few particles of a harmless background material. The instrument must ultimately perform many simultaneous but independent calculations to determine whether there is sufficient reason to sound an alarm for each of the individual agent particle types. These calculations are identical mathematically, so only the calculations necessary for one particular “agent of interest” are derived here. One of the few scenarios in which this simultaneous but independent approach would be inappropriate is if a synchronized release of multiple agent types was expected. The model can be expanded to include this type of scenario, but this is not done here.

Before embarking on more detailed descriptions, it is worthwhile to briefly summarize the basic steps in the model. First, the interaction of a specified aerosol with a specified instrument is modeled to determine the average number of particles of each aerosol particle type analyzed per sample period. Next, the single particle identification and misidentification rates are applied to determine the average number of these particles either correctly or incorrectly identified as agent. A statistical equation is then utilized to determine the appropriate alarm threshold for a particular agent of interest. Finally, a second equation is applied to determine the minimum concentration of that agent necessary to sound an alarm with a certain desired probability. The ultimate output of the model can take many forms, but two particular types of graphs are produced here. The first of these will be referred to as a performance plot. The second is the well-known Receiver Operating Characteristic (ROC) curve. Both are described in detail further below.

The necessary equations for the statistical model are derived in sections 7.2 and 7.3, before considering details of specific systems, because the equations are generally applicable to many types of instruments. The required aerosol properties and instrument parameters are described in section 7.4. Numerical values are assigned and a hypothetical aerosol and instrument are modeled in section 7.5. Various modifications to the basic system and the resulting performance improvements are described in section 7.6. Tradeoffs between detection probability, detection time and false alarm rate are illustrated in section 7.7.

## 7.2 Probability of False Alarm

False alarms result from the misidentification of background particles as agent particles. Mass spectral fluctuations are unavoidable so there is a small, but non-zero chance that a harmless particle will produce a mass spectrum that is similar enough to the spectrum from an agent to cause a misidentification. Assume that  $m$  background particle types are present in the air. (One “type” might be sea salt, another *B.t.*, another a particular species of pollen, etc.) For each of these types there is a probability,  $P_{a,i}$  ( $i=1..m$ ), that an individual particle will be misidentified as the agent of interest when it is analyzed. (The reduction of spectral variability described in chapter 5 reduces this probability, but a finite chance for misidentification will always remain.) Different particle types have different densities, sizes, and other properties that cause them to be transported and analyzed with different efficiencies in the BAMS instrument. All of the various factors influencing these efficiencies will, for now, be summarized with a single “efficiency” parameter,  $h_i$ , for each particle type. If the concentration of particles of type

$i$  is  $C_i$ , then  $h_i C_i$  such particles will be analyzed by the system per minute on average. If a sampling period lasts for  $T$  minutes, the average number of particles misidentified as agent will be  $N_i = P_{a,i} h_i C_i T$ . Generally, the probability,  $p_{N_i}(k_i)$ , to observe exactly  $k_i$  misidentifications of particle type  $i$ , given an average of  $N_i$  can be approximated by a Poisson distribution

$$p_{N_i}(k_i) = e^{-N_i} \frac{N_i^{k_i}}{k_i!}.$$

**Eq. 20**

It is possible, however, to derive a general equation for the false alarm rate without taking advantage of the specific functional form of Eq. 20. This general equation, derived in the following, is actually useful since some situations that involve very small aerosol releases or instruments that do not sample at a uniform rate may not be described well by Eq. 20.

Based upon the scenario in which the instrument is being used, the probability of false alarm will be required to be less than or equal to  $P_{fa}$  (a specific number specified by the user). If the aerosol contained only a single particle type (particle type 1, for example), it would be possible to simply sum over Eq. 20 (or whatever the true distribution might be) to find a value of  $k_1$  such that the probability to obtain  $k_1$  or more misidentifications would be less than or equal to  $P_{fa}$ .

$$P_{fa} \leq \sum_{i=k_1}^{\infty} p_{N_1}(i) = 1 - \sum_{i=0}^{k_1-1} p_{N_1}(i).$$

**Eq. 21**

(Note that the second part of Eq. 21 relies upon the proper normalization of  $p_{N_1}$ .) The number,  $k_1$ , that just satisfies Eq. 21 would define the threshold to sound an alarm. There

will generally be more than a single background particle type, however, so the calculation becomes more complicated.

The probability to obtain exactly  $k_1$  misidentifications of type 1 particles,  $k_2$  misidentifications of type 2 particles,... and  $k_m$  misidentifications of type  $m$  particles, all in one sampling period is

$$\prod_{i=1}^m p_{N_i}(k_i).$$

**Eq. 22**

The particular values of the  $k_i$  do not actually matter; all that is important is their sum,  $k$ . A value of  $k$  (not  $k_i$ ) must be found such that the probability to obtain  $k$  or more total misidentifications (of background particles as a particular type of agent particle) is less than or equal to the desired  $P_{fa}$ . This requires a summation over all the different combinations of the  $k_i$  that add up to  $k$  or more. The smallest value of  $k$  that satisfies the following inequality must be found. This value defines the proper alarm threshold.

$$P_{fa} \geq \sum_{k_1+\dots+k_m \geq k} \prod_{i=1}^m p_{N_i}(k_i) = 1 - \sum_{k_1+\dots+k_m < k} \prod_{i=1}^m p_{N_i}(k_i).$$

**Eq. 23**

Eq. 23 is valid independent of the details of the probabilities  $p_{N_i}(k_i)$  (so long as they are independent), but may be rather difficult to use in practice. Fortunately, this equation simplifies greatly if a Poisson distribution accurately predicts the number of misidentifications for each of the individual particle types (as assumed in Eq. 20). In this case, Eq. 23 reduces (as shown in appendix A.8) to

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} e^{-N} \frac{N^j}{j!} \quad \text{where} \quad N = \sum_{i=1}^m N_i .$$

**Eq. 24**

It is possible, in theory, to obtain an arbitrarily low false alarm rate; the instrument can simply require that larger and larger numbers of particles be identified as agent before sounding an alarm. As the alarm threshold (i.e. the required  $k$  value) is increased, however, it becomes less and less likely that a given concentration of agent (once added to the background) will actually produce enough agent identifications to trigger the alarm. It is thus important to also determine the minimum concentration of agent that is likely to produce  $k$  or more agent identifications in one sampling period.

### 7.3 Probability of Detection

Assume that the agent concentration is  $C_a$ , the efficiency parameter is  $h_a$ , and the probability that an agent particle is properly identified is  $P_{a,a}$ . (Reduction of spectral variability causes  $P_{a,a}$  to become larger, approaching 100% in the ideal case.) The average number of agent particles properly identified per sampling period  $T$  is thus  $N_a = P_{a,a} h_a C_a T$ . Once again, the probability,  $p_{N_a}(k_a)$ , to obtain a specific number,  $k_a$ , of correct agent identifications given an average of  $N_a$  is generally described by a Poisson distribution.

$$p_{N_a}(k_a) = e^{-N_a} \frac{N_a^{k_a}}{k_a!} .$$

**Eq. 25**

The probability to properly identify  $k_a$  or more agent particles (whether the exact distribution is Poissonian or not) is just the sum

$$\sum_{i=k_a}^{\infty} p_{N_a}(i) = 1 - \sum_{i=0}^{k_a-1} p_{N_a}(i).$$

**Eq. 26**

If this equation was set equal to the desired probability of detection,  $P_d$ , and solved for the agent concentration, the sensitivity of the instrument would actually be underestimated. The alarm is triggered whenever  $k$  or more total particles (whether truly agent or not) are identified as agents. Eq. 26 does not include the contribution from misidentified background particles. This contribution can be relevant, as the following example demonstrates. Consider a hypothetical situation in which the background aerosol contains sufficient concentrations of easily misidentified particles to produce an average of 25 “agent” identifications (even though no agent is present). Application of Eq. 24 indicates that the alarm threshold must be set to  $k=50$  to ensure that  $P_{fa}=10^{-5}$ . If the misidentification of background particles was ignored, it would appear that ~50 agent particles would have to be identified as agent on average to sound the alarm reliably. In reality, only ~25 agent particles must be identified as agent, on average, because the background supplies the other ~25 agent identifications necessary to reach the alarm threshold. In this hypothetical case, the true sensitivity of the instrument would be roughly twice that predicted ignoring the background.

To find the true sensitivity of an instrument in general, the minimum value of  $N_a$  (or really  $C_a$ ) that satisfies the following inequality must be found.

$$P_d \leq \left( 1 - \sum_{j=0}^{k-1} p_{N_a}(j) \right) + \sum_{j=0}^{k-1} \left( p_{N_a}(j) \times \left( 1 - \sum_{k_1+\dots+k_m < k-j} \prod_{i=1}^m p_{N_i}(k_i) \right) \right).$$

**Eq. 27**



The first term on the right is identical to Eq. 26. It is the probability that  $k$  or more agent particles are properly identified. The second term is the sum (over  $j$ ) of the probability that  $j$  agents are properly identified ( $j < k$ ) multiplied by the probability that more than  $k-j$  background particles are misidentified (so that the total number of particles identified as agent is always  $\geq k$ ). This can be rewritten to closely resemble Eq. 23, but this is not done here. As in the case of Eq. 23, Eq. 27 simplifies greatly when the various distributions involved are all Poissonian. The result is

$$P_d \leq 1 - \sum_{j=0}^{k-1} e^{-N} \frac{N^j}{j!} \quad \text{where} \quad N = N_a + \sum_{i=1..m} N_i .$$

**Eq. 28**

Given a certain background of various types of particles and a certain desired probability of false alarm, Eq. 24 can be used to determine the necessary alarm threshold. Once the alarm threshold is determined, Eq. 28 can be used to find the concentration of agent needed to meet or exceed that threshold with a probability equal to the desired probability of detection.

## 7.4 Aerosol Properties and Instrument Parameters

The basic theoretical framework now exists to predict how an instrument will perform in virtually any aerosol environment. Numerical values are needed for the various aerosol properties and instrument parameters to determine if a hypothetical BAMS system, with reasonably obtainable characteristics, can perform well enough to be a practical detector or not. The exact performance of a real BAMS system is sensitive information and cannot be revealed here. Order of magnitude approximations and estimates for the various required parameters are used here to predict the performance of a hypothetical system. A

few aerosol properties occur explicitly in the above equations and these will be discussed first. All of the instrument parameters and a number of additional aerosol properties must be combined to obtain values for the efficiency parameters,  $h_a$  and  $h_i$ . This will be discussed subsequently.

The concentration of each aerosol particle type must be specified as well as the single particle identification and misidentification probabilities. For modeling purposes, the concentrations can simply be chosen at will. In the field, the actual particle concentrations of each type should be measured. (With proper calibration, the BAMS system will be able to perform these measurements directly and no additional hardware will be required.) The precise probability that particles of a given type are misidentified as a particular agent ( $P_{a,i}$ ) and the precise probability that agent particles are properly identified as agent ( $P_{a,a}$ ) can only be obtained from laboratory experiments with those particle types. (Estimates of these probabilities may be obtained by extrapolation of laboratory measurements on similar or related particle types however.) The BAMS group has already demonstrated the ability to analyze several thousand spectra from certain particle types without misidentifying any of them as agents (or more precisely as agent surrogates). This is encouraging, but it does not prove that particles will never be misidentified (i.e.,  $P_{a,i}=0$ ). A reasonable upper limit for a value of  $P_{a,i}$  in this type of statistics-limited experiment can be determined as follows. The probability,  $P_0$ , to obtain no misidentifications of a certain particle type in  $N$  trials is simply  $P_0=(1-P_{a,i})^N$ . This can be rearranged to obtain

$$P_{a,i} = \left(1 - P_0^{\frac{1}{N}}\right) \approx \frac{1}{N} \ln\left(\frac{1}{P_0}\right).$$

**Eq. 29**

The approximation is valid in the limit of large  $N$ . If  $P_0$  is set to 5%, an upper limit for  $P_{a,i}$  is simply  $3/N$ . This makes a convenient rule of thumb. If 3000 *B.t.* spore spectra, for example, were analyzed without a single misidentification (as a particular agent or agent surrogate, e.g. *B.at.*) a reasonable estimate for  $P_{a,i}$  would simply be  $3/3000=10^{-3}$ . If the true value of  $P_{a,i}$  were any higher, it would be extremely likely ( $\geq 95\%$ ) to observe at least one particle identified as agent (or agent surrogate).

Specific values for the efficiency parameters could be determined relatively easily for well-defined aerosols in carefully controlled laboratory experiments, but the values will change whenever the aerosol or instrument is modified. It is simply not practical to reproduce every aerosol and instrument combination in the lab that might be encountered in actual field use or deployment. Specific properties of single aerosol particles and specific instrument parameters must be measured or estimated and then combined mathematically to produce values for the efficiency parameters. As already mentioned, the efficiency parameters ultimately relate the concentration of each type of particle in the local aerosol surrounding the BAMS system to the average number of particles of each type analyzed by the system per minute. For the purpose of organization, the efficiency parameters will be determined by factoring in the properties of each component of the BAMS system sequentially, more or less as encountered by the aerosol particles entering the system.

To begin, particle types with different aerodynamic diameters will be focused differently by the inlet nozzle. Particles with diameters that are poorly focused are not

likely to be fully analyzed. At least one parameter (and in practice several) are needed to properly quantify the size dependent focusing. In the present case, the probability that a particle is tracked given that it has been sampled and the probability that a particle can be hit by the DI laser given that it has been tracked are needed to assess the particle focusing efficiency (rate limitations are intentionally excluded at this point). Fortunately, relatively simple experiments can be devised to measure these parameters directly. After the inlet, the next major component that must be considered is the tracking system. In the current calculations, the tracking system is treated as having a certain maximum operating rate just like the DI laser and the data acquisition system. The probability that a given particle can be tracked (assuming that it has been focused well enough to cross the necessary tracking laser beams) can be derived easily. Since a similar equation for the rate limitation also applies to the DI laser (which is the next major instrument component after the tracking system) and the DI laser system is easier to understand, the equation for the rate limitation is developed considering the DI laser explicitly in the next paragraph.

Consider a DI laser with a maximum firing rate of  $R$  pulses per minute that is given  $N$  randomly incoming particles per minute to fire at. If the laser has just fired at a particle, the laser must wait  $1/R$  minutes until it is able to fire again. Once it is able to fire, the system must wait  $1/N$  minutes (on average) for the next particle to arrive. The total time between laser shots is thus  $1/N + 1/R$  minutes and the laser is expected to fire at  $NR/(N+R)$  particles per minute on average. The probability that an individual particle is fired at is consequently  $R/(N+R)$ . If the DI laser has a maximum operating rate of 1 kHz, for example, and particles arrive randomly at an average rate of 1 kHz, only 50% of the particles can be fired upon. Note also that not every particle that is fired upon is hit

because of imperfect focusing. As was mentioned earlier, the basic rate equation is applied to the tracking system, the DI laser system and the DAQ and data analysis system. Each system has a different maximum rate (in units of tracks per minute, shots per minute and spectra analyzed per minute, respectively), but these are easily determined from manufacturer specifications or experiments.

Additional aerosol particle properties must still be factored in to the performances of the system components just mentioned to determine the overall efficiency parameters. Very small particles, or particles with low indices of refraction may not scatter sufficient light to be detected by the tracking system. If so, the particles will not be analyzed. Similarly, it was shown in chapter 5 that different particle types have different fluence thresholds. For a given laser fluence distribution, not all particles that interact with the laser beam may produce spectra. At low laser energy, for example, it is very unlikely that BSA or MS2 spectra will be produced, given the results from chapter 5, regardless of the concentration of BSA and MS2 particles. Fortunately, all of these properties can be quantified.

In the next section, numerical values are assigned to the parameters described in order to obtain predictions for the performance of the basic hypothetical system described hitherto. Modifications and additions to this hypothetical BAMS system are discussed later (in 7.6) together with the additional parameters introduced by those modifications.

## **7.5 Performance Predictions**

In the following, a generic background aerosol is considered that consists of 100 parts non-biological particles and 1 part biological particles. For the particular output that is

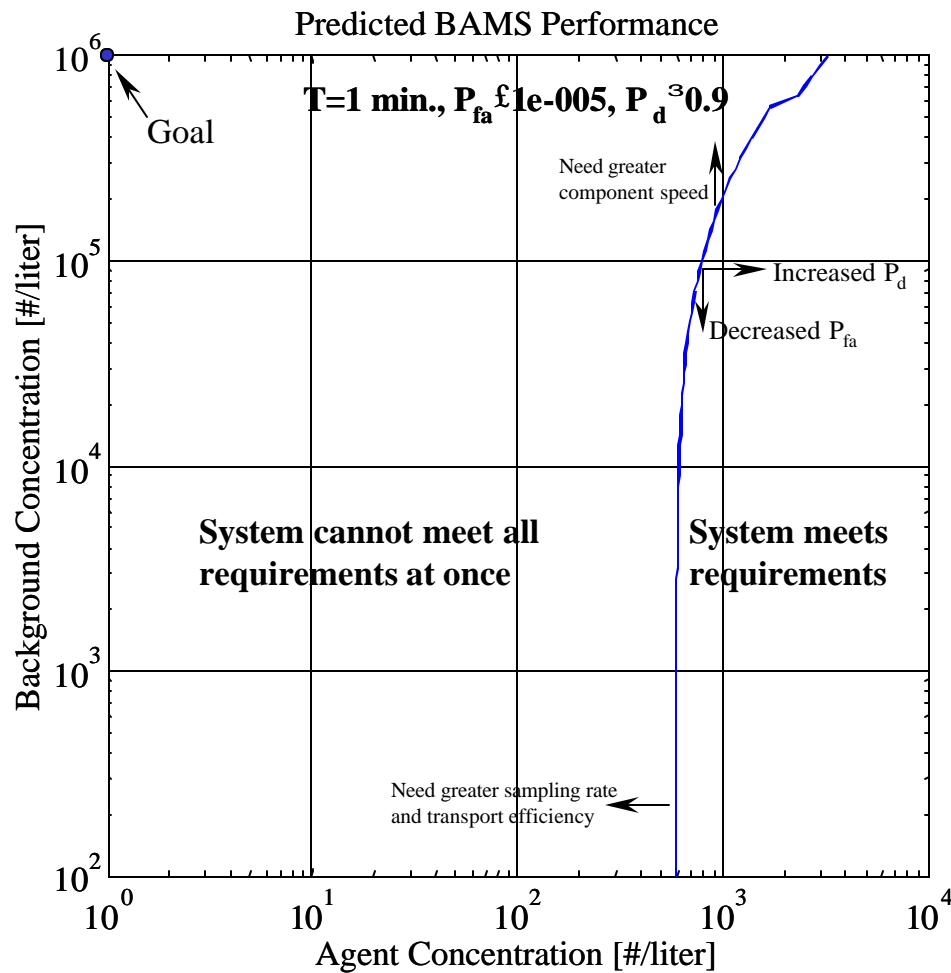
shown shortly, the model itself determines and varies the total concentration keeping the ratio constant. The non-biological particles (which constitute most of the aerosol) are assumed to produce mass spectra that are very distinct from agent particle mass spectra and are unlikely to be confused as agent. The biological particles (a small fraction of the total aerosol) are assumed to produce spectra that are more similar to the agent spectra and are thus more likely to be misidentified as agent. It is assumed that the misidentification rate for non-biological particles,  $P_{a,i}$ , is  $10^{-5}$  and the misidentification rate for biological particles is  $10^{-3}$ . The probability that an agent particle is correctly identified,  $P_{a,a}$ , is assumed to be 90%.

The rate at which air and entrained particles are sampled into the hypothetical BAMS system is  $\sim 1$  l/min. It is assumed that 10% of the particles drawn into the instrument can be tracked and that 10% of the particles that are tracked can be hit by the DI laser producing spectra. These rates account for imperfect particle focusing and will be treated here as independent of particle size and composition for simplicity. It is assumed that all particles scatter enough light to be tracked and that all particles that are hit by the DI laser produce ions. The hypothetical tracking system has an effective repetition rate of 100 kHz (i.e. can track a maximum of  $10^5$  particles per second) while the DI laser has a maximum firing rate of 1 kHz. The DAQ and data analysis system can acquire, process and identify spectra at 100 Hz. For the present calculations, it is assumed that some type of detector is used near the ion source region to determine if a particle has produced ions so that the DAQ system is only triggered if it is certain that a spectrum will be acquired.

A 90% or greater probability of detection is desired with a false alarm rate less than or equal to  $10^{-5}$  and a sampling period of 1 minute. Although it may not be required

mathematically, the BAMS system will be forced to always require at least three agent identifications before sounding an alarm. This is simply an extra safeguard to help ensure that the false alarm rate remains low in practice.

A computer program was written to perform all of the necessary model calculations and produce various types of output (the code is included in A.7). Upon inserting all of the numerical values just listed into the program, the “performance plot” shown in Figure 54 is produced. The horizontal axis shows the concentration of agent, while the vertical axis shows the total background concentration (not including the agent). In this example,



**Figure 54.** The predicted performance of the hypothetical BAMS system. The system can operate anywhere under the curved line with the desired  $P_{fa}$ , and  $P_d$  given a 1 minute sampling time. Outside of the curve, it is generally possible to meet some but not all of the requirements. An ideal detector should be able to operate in the upper left hand region of the plot (one particular goal has been indicated).

the ratio of the concentrations of the two background particle types remains fixed (non-biological:biological=100:1) as the total background concentration is scaled. The curved line on the plot shows the boundary between the region where the BAMS instrument can meet or exceed the required probability of detection and false alarm rate, and the region where it cannot (which are labeled in the figure). If one picks a point on the boundary and then moves right (further into the region where the system can operate), the probability of detection will increase since a greater and greater amount of agent is present in a given concentration of background particles. Similarly, if one picks a point on the boundary and then moves downward, the probability of false alarm could be decreased without affecting the probability of detection because the number of background particles is reduced while the amount of agent is held fixed.

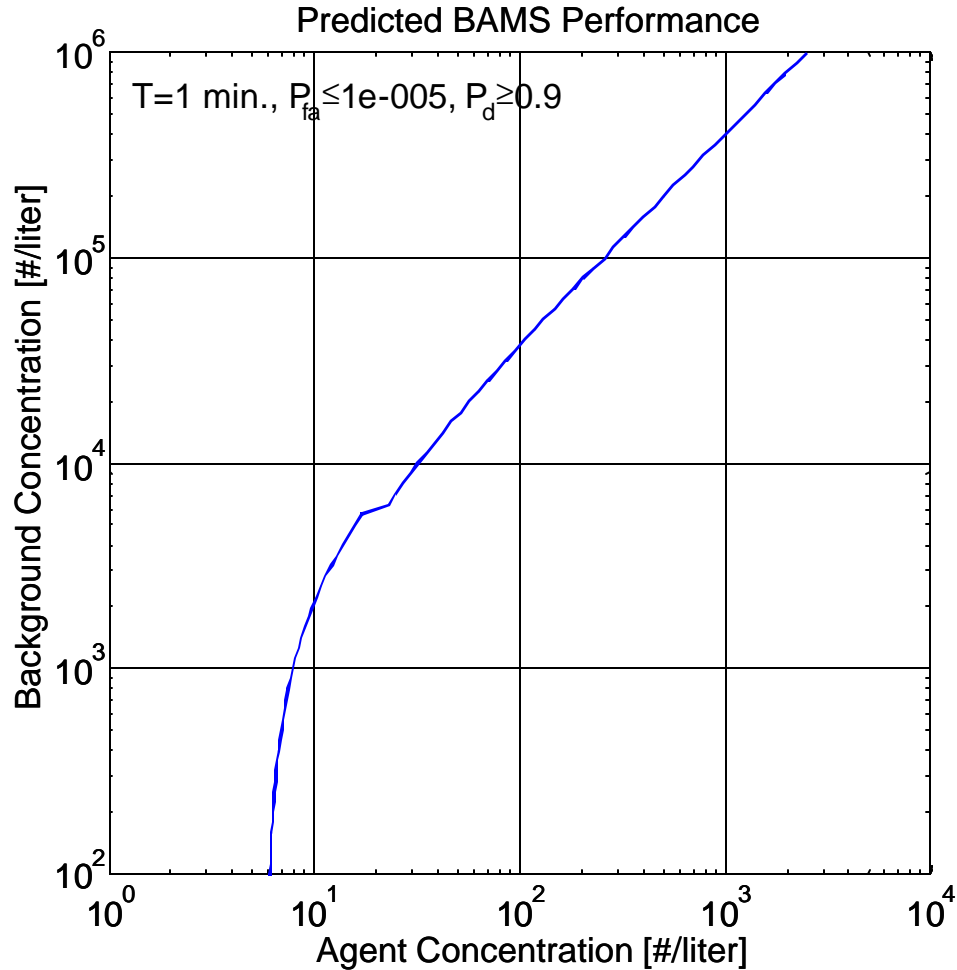
Clearly, it is desirable to push the region where the instrument meets the false alarm rate and detection probability goals towards the upper left corner of Figure 54. The BAMS system should ultimately be able to detect small agent concentrations in large concentrations of background materials. (Relevant concentrations of agent and background particles are described below.) There are several ways to advance the system towards this goal (in addition to simply sampling for a longer period of time). Generally separate techniques are needed to push the curve upward (towards higher background) and to push the curve leftward (towards lower agent concentration). Examples of such improvements are described in the next section.



## 7.6 Preconcentrators and Prescreeners

A real BAMS system with the configuration described hitherto would rarely be used in practice because its sensitivity would be relatively low. The ultimate sensitivity of the instrument is limited by the amount of air that can be sampled and by the fraction of the particles drawn into the system that are successfully tracked and analyzed. If three particles must be identified as agent in one minute to sound an alarm, at least 300 agent particles must be drawn into the system (given the assumed tracking and hitting efficiencies of 10% each). Since the system draws in 1 liter of air per minute, there must be at least 300 agent particles per liter of air in the environment to have a significant chance of triggering the alarm. To meet the 90% required probability of detection (with a 90% correct identification rate for single agent particles) there must be a concentration of almost 600 agent particles per liter of air, even if no background particles are present. Consequently, this is the agent concentration at which the curve in Fig. 54 intercepts the x-axis.

The performance of such a system can be improved significantly by adding a preconcentrator to concentrate the aerosol before it is drawn into the instrument and a prescreener to screen out particles that are obviously not agents before they are fully analyzed by the mass spectrometer. A hypothetical virtual impactor can be added to the modeled instrument that draws in 100 l/min of air and concentrates the entrained aerosol particles into a 1 l/min flow sampled directly by the BAMS system. (The virtual impactor concentrates the particles by exploiting their inertia.) The addition of the virtual impactor produces the results shown in Figure 55.

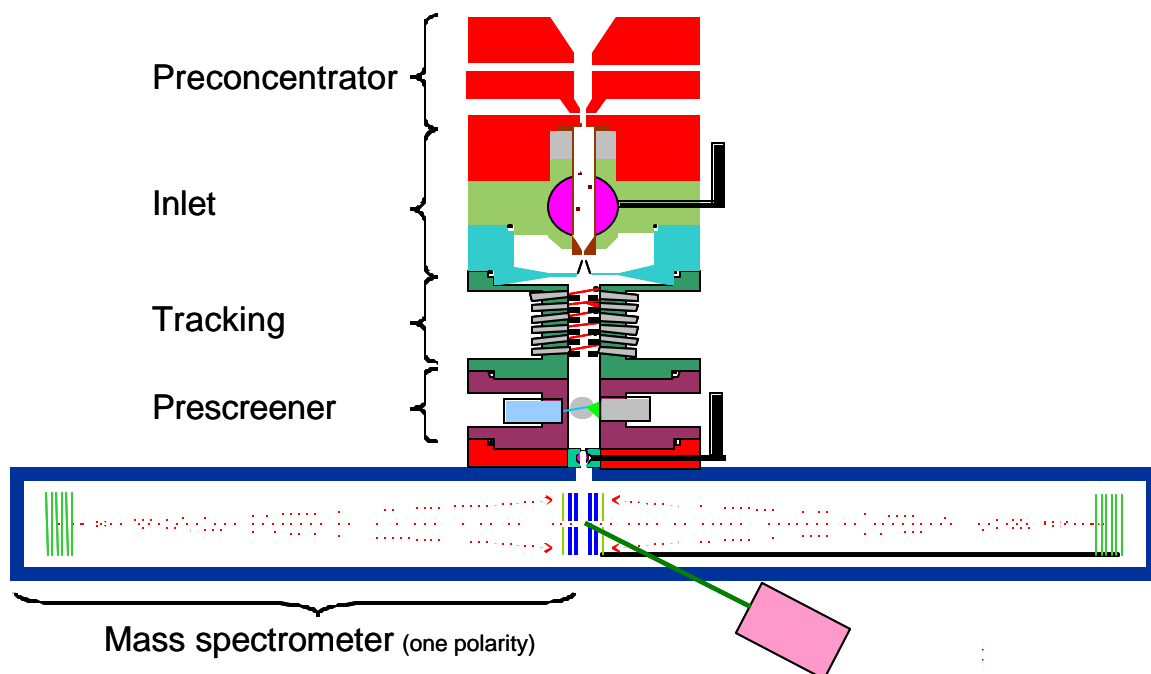


**Figure 55.** Performance of the BAMS instrument with the addition of a virtual impactor. The sensitivity of the resulting detector is greatly improved compared to the instrument without a VI (see Figure 54).

The minimum detectable concentration of agent is clearly improved. It is also apparent, however, that as the concentration of background particles increases, the minimum detectable concentration of agent increases as well (i.e. the sensitivity decreases with increasing background). This is because of the finite speeds of the DI laser and DAQ systems. Even if enough particles are sampled to fire the DI laser at 1 kHz (its maximum rate), the hit rate is assumed to be 10% so only 100 particles per second will produce ions on average (all particles that are hit produce ions in the current example). Since the times at which the ions are produced is basically random, the DAQ and data analysis system (with a maximum rate of 100 Hz) will only be able to acquire data from

and analyze 50% of these events on average. This means that no more than 3,000 particles can be fully analyzed per minute on average. Since at least 3 agent particles must be identified in one sampling period to sound an alarm, the aerosol cannot contain more than  $3000/3=10^3$  times as many background particles as agent particles if even a relatively low probability of detection is required. For a 90% detection probability the ratio between background and agent (with the properties described above) cannot exceed ~400. This limit for the ratio is equally valid for the instrument without the virtual impactor, but in that instrument much of the performance curve (Figure 54) was dominated by the limited number of agent particles that could be sampled.

With or without a virtual impactor, it is important to notice that a great deal of time is wasted analyzing particles that are very distinct from the agent whenever a significant concentration of background particles is present. The performance would improve significantly if an additional prescreening technique could be integrated to identify clearly distinct (non-agent like) particles before they reached the mass spectrometer so that the instrument could simply let them pass through the ion source region without firing the DI laser or trying to analyze them further. Laser shots and processing time could be reserved for those particles that are at least somewhat similar to the agent particles being sought. Several existing biodetectors already use UV laser induced fluorescence (UV-LIF) to identify biological particles (chapter 1). It is easy to imagine placing a fluorescence prescreener between the tracking system and the mass spectrometer system (Figure 56). The prescreener would attempt to probe each of the tracked particles with a low energy, pulsed UV laser. If sufficient fluorescence was detected (or perhaps, if the proper spectral pattern of fluorescence was recognized), the



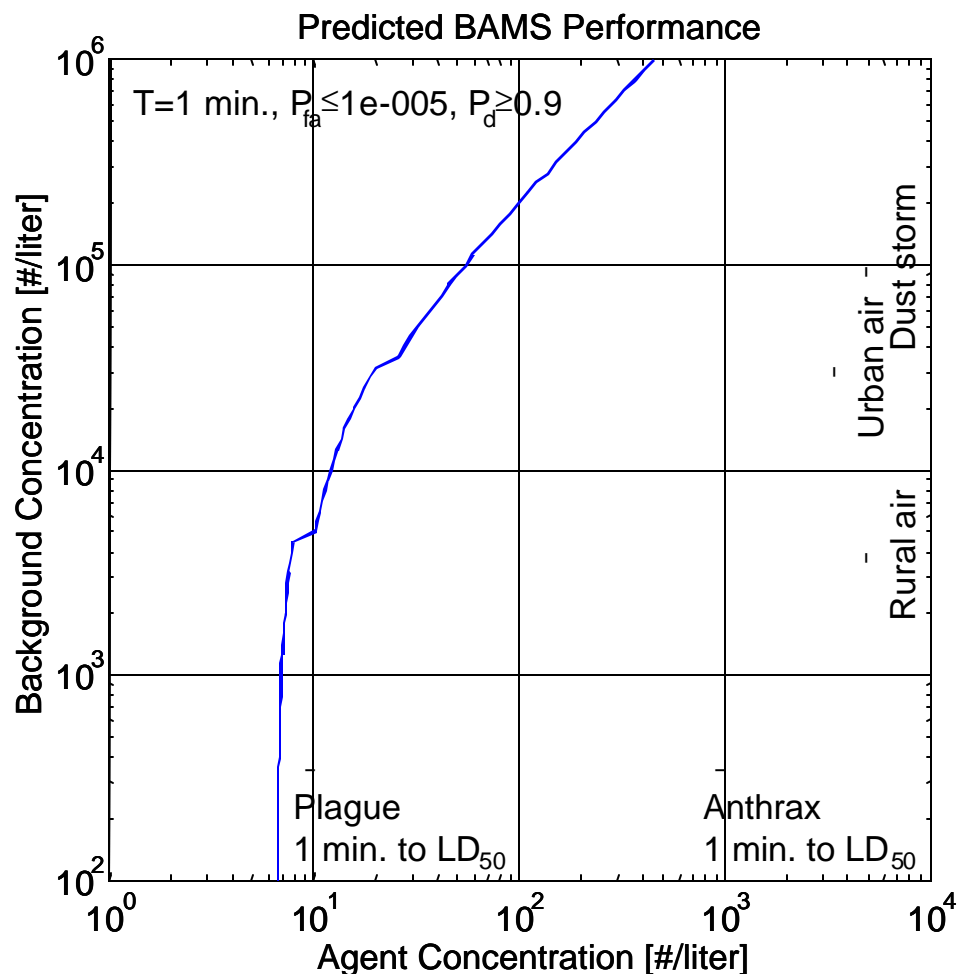
**Figure 56.** A hypothetical BAMS system with a virtual impactor (concentrator) and a UV-LIF prescreener.

prescreener would inform the DI laser that the particle was “interesting” and should be fully analyzed.

For the calculations that are performed shortly, it is assumed that 10% of non-biological particles and 90% of biological and agent particles fluoresce sufficiently to be identified as interesting and that the ratio of non-biological to biological particles is 100:1, as before. The imperfect particle focusing must also be taken into account. Since the UV-LIF prescreener’s laser is closer to the tracking system than the DI laser, and of comparable diameter, the probability that a tracked particle will interact with a given pulse from the fluorescence laser should be higher. Assume that the fluorescence laser hits 40% of the tracked particles that it fires at. If a particle is hit by the fluorescence laser, assume that the DI laser has a 25% chance of hitting it (25% of 40% is 10%, which was the original hit rate of the DI laser). Clearly the DI laser should not be fired at particles that were missed by the fluorescence laser. The performance of the BAMS

system with a virtual impactor and a UV-LIF prescreener (assumed to have a maximum rate of 10 kHz) is shown in Figure 57.

The performance curve in Figure 57 has effectively shifted upward (relative to the curve in Figure 54 or Figure 55). A given concentration of agent can now be reliably detected in a higher concentration of background aerosols. The minimum detectable concentration of agent, however, has remained basically unchanged compared to the system with only a prescreener (Figure 55). A few examples help to clarify the



**Figure 57.** Performance of a BAMS system utilizing a virtual impactor and a fluorescence based prescreening system. A few relevant agent concentrations are included. Typical background concentrations are also shown for a few different environments (a broad range of concentrations is possible in each environment however). For the simple model here, the particle size has largely been ignored, but the background concentrations represent particles in the “respirable” size range (roughly 1-10  $\mu\text{m}$ )

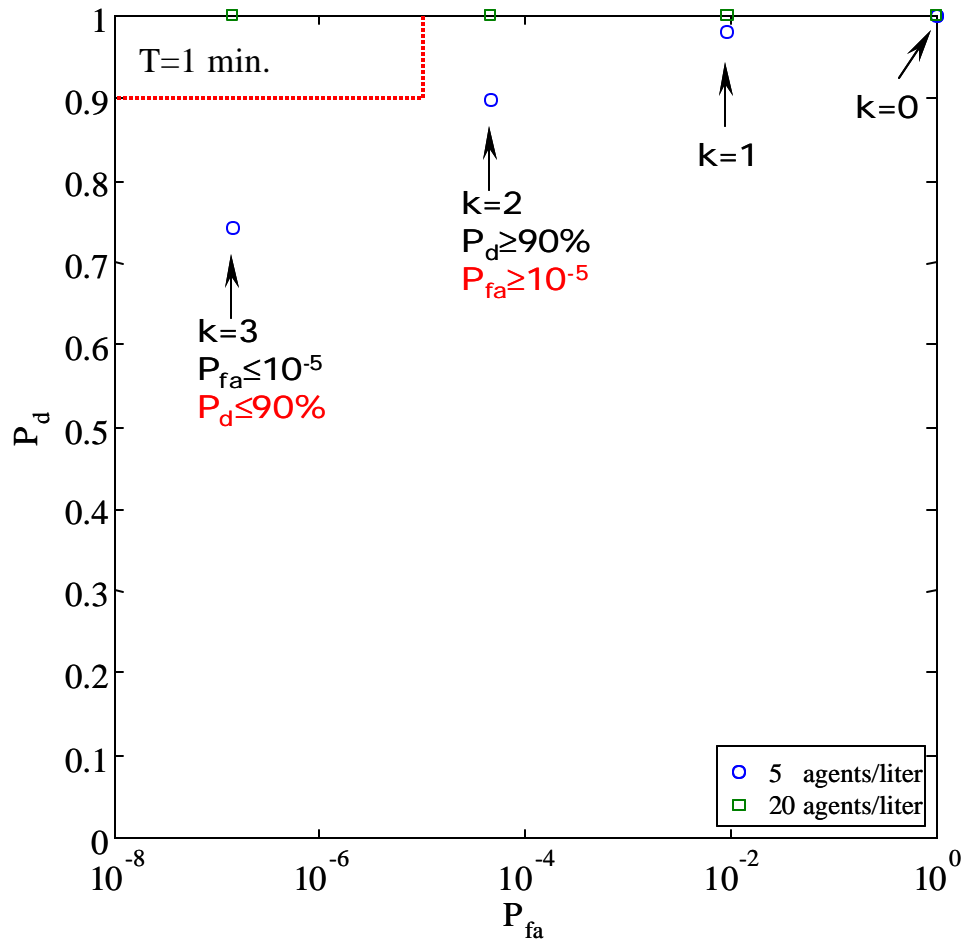
significance of the horizontal and vertical ranges shown in Figure 57. Very small amounts of certain agents can be lethal. The LD<sub>50</sub> for anthrax is on the order of 10,000 spores<sup>40</sup>. A normal person inhales approximately 10 liters of air per minute. A one-minute exposure to 1000 anthrax spores per liter thus results in a potentially lethal dose. An infectious dose of plague may only contain 100 cells<sup>41</sup>. It is important, as a result, to be able to detect 10 plague cells per liter in a minute or less. The vertical scale on the plot is determined by the background aerosol. In a rural environment, the total concentration of particles between 1 and 10 μm is on the order of a few thousand per liter. This concentration is also comparable to that found in a typical building. In an urban environment, the concentration is on the order of a few tens of thousands per liter. In a dust storm, the concentration could reach 10<sup>5</sup> particles per liter or more.

In just one minute, the hypothetical BAMS system can detect an agent concentration of less than 10 particles/liter in a background of more than 10<sup>3</sup> particles/liter (some of which are relatively similar to the agent) with a probability of 90% and a false alarm rate of 10<sup>-5</sup>. This is a powerful ability. Particular note should be taken of the false alarm rate; it represents less than one false alarm every two months if the machine made a separate measurement every minute, 24 hours a day, 7 days a week. The exact false alarm rates of existing detectors with similar detection times are not publicly available (and are in fact unknown by the author), but there seems to be a general consensus that they are significantly worse than 10<sup>-5</sup>. Similarly, the sensitivities of existing detectors are not publicly available, but are believed to be comparable or inferior to the sensitivity of the BAMS system.

The level of performance expected for the BAMS system with a virtual impactor and a UV-LIF prescreener is clearly useful, but various improvements could still be made. If more sensitivity was desired, more particles could be sampled and the particles that were sampled could be analyzed more efficiently. Based on published literature<sup>26</sup>, there is good reason to believe that the efficiency of the nozzle could be improved so that the incoming particle beam would be focused more tightly and that significantly more than 1% of the sampled particles would be tracked and hit by the DI laser. If operation in more polluted environments was needed, better prescreeners and faster lasers and electronics could be used. The detection of 1 agent particle per liter in a background of  $10^6$  particles per liter may be achievable with such improvements.

### **7.7 Tradeoffs between $P_d$ , $P_{fa}$ and $T$**

The curve in Figure 57 separates the region on the right where the modeled BAMS system can meet all of the current performance requirements ( $P_{fa} \leq 10^{-5}$ ,  $P_d \geq 90\%$ ,  $T=1$  min) from the region on the left in which it cannot. That does not mean, however, that the system is useless to the left of the curve. Consider a point on the plot at an agent concentration of 5 particles per liter and a background concentration of 1000 particles per liter, which is slightly to the left of the blue curve. Figure 58 shows a Receiver Operating Characteristic (ROC) curve for this scenario. ROC curves are another form of output that can be produced by the performance model using the same equations derived above. (ROC curves were originally used in the 1940s to quantify the discrimination of radio signals in the presence of noise as a function of detection threshold. Recently, ROC curves have been used more generally to show the tradeoff between detection probability



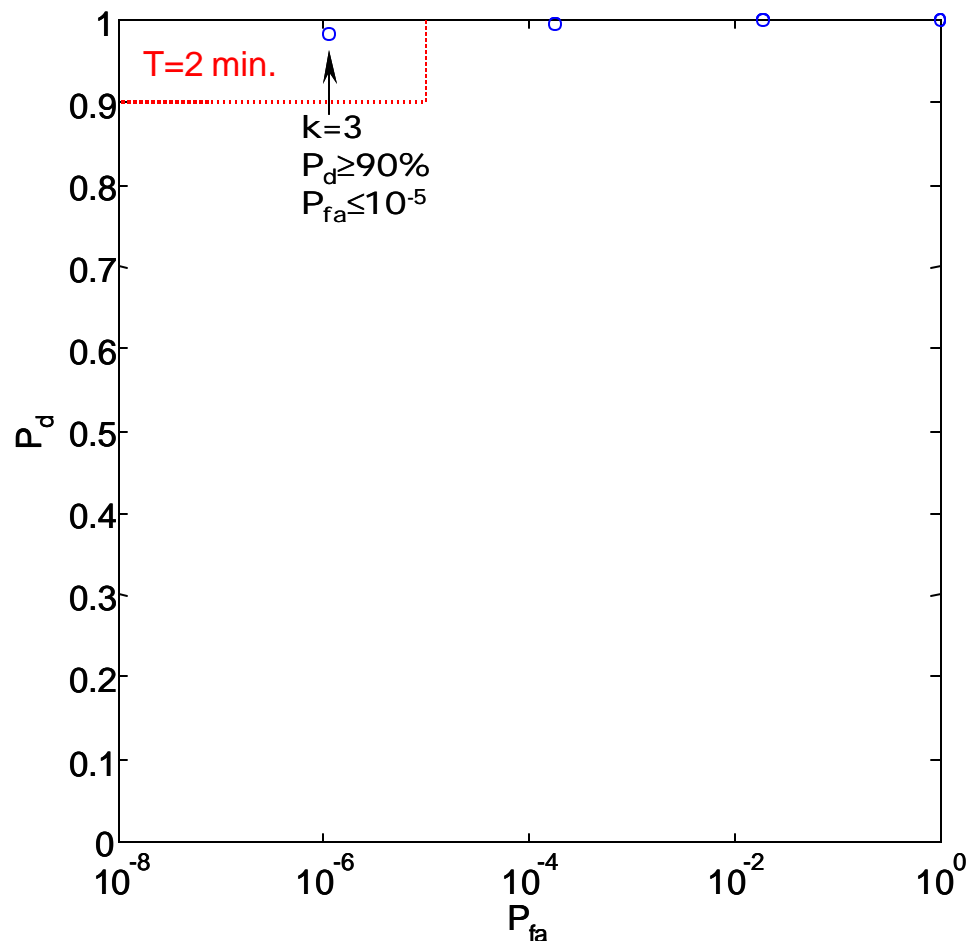
**Figure 58.** ROC curve for 1 minute detection of 5 agent particles/liter in a background of 1000 particles/liter. If the alarm threshold is set to  $k=2$ , the probability of detection meets the requirement, but the probability of false alarm is a bit too high (although it's still less than  $10^{-4}$ ). If the threshold is set to  $k=3$ , the probability of false alarm is less than  $10^{-6}$ , but the probability of detection is a bit too low. Both alternatives may be perfectly acceptable in certain scenarios. A concentration of 20 agent particles/liter is easily detected while meeting all performance requirements.

and false alarm.) The horizontal axis shows the probability of false alarm (log scale in this case). The vertical axis shows the probability of detection (linear scale). Clearly a good detector should be able to operate in the upper left hand corner of the plot. For some instruments, the ROC curve may truly be a continuous curve. The BAMS system analyzes individual discrete particles, however, so the ROC curve consists of discrete points corresponding to different alarm thresholds.

For the 5 agent particle/liter scenario under consideration, none of the points on the appropriate ROC curve (○) fall in the box in the upper left hand corner of the plot. This



indicates that the performance requirements cannot all be met simultaneously (as expected, since the chosen scenario corresponds to a point left of the curve in Figure 57). If the alarm threshold is set at  $k=2$ , the probability of detection is satisfactory, but the probability of false alarm is not. It is, however, still quite low and may be perfectly acceptable in some situations (e.g. the protection of a military base). If the “high” false alarm rate cannot be tolerated, the alarm threshold can simply be increased to  $k=3$ . The false alarm rate is now less than  $10^{-6}$  and the probability of detection is still greater than 70%. Again, this constitutes a very useful state of operation.



**Figure 59.** ROC curve for detection of 5 agent particles/liter in a background of 1000 particles/liter in 2 minutes. Since the concentration of agent is so low, there is little need to detect it in 1 minute or less. If the sampling period is increased from 1 minute to 2 minutes, the agent can easily be detected with the required probability of false alarm and probability of detection.

If both the probability of detection and probability of false alarm requirements have to be met simultaneously, the sampling time can simply be lengthened. Figure 59 shows a ROC curve for the same scenario shown in Figure 58, except that the sampling time has been increased to 2 minutes. An alarm threshold of  $k=3$  now meets both of the probability requirements. In the case of anthrax, it would take far longer than 2 minutes to inhale an  $LD_{50}$  at 5 spores/liter so the increased sampling time should not be cause for great concern.

These ROC curves apply only for one particular scenario; nonetheless they illustrate the importance and value of the BAMS performance model. Given a particular background aerosol, the tradeoffs between sensitivity, specificity and speed can easily be determined using the model. With the aid of a computer, these calculations take only seconds to perform. In addition to simply changing the alarm threshold, the model could be used to dynamically modify the operation of a real system in the field. In some circumstances, for example, it might be beneficial to turn off the virtual impactor or to turn off the UV-LIF prescreener. Variations of the mathematical model developed here will be used to tailor the performance of the BAMS system to end user preferences and to optimize the system for deployment in virtually any environment.

## Outlook

The current “proof-of-concept” BAMS instruments can already analyze individual biological aerosol particles and differentiate at least some closely related species (e.g. *B.at.* vs. *B.t.*). The flattop 266 nm DI laser profile, implemented and quantified here, reduces the variability of the raw data produced by the systems. This is already beneficial, but it is likely to become more and more important as the library of spectral types (for particle identification) continues to grow. The flattened profile has also facilitated the measurement of scientifically interesting particle properties, such as the DI fluence threshold, and will be used for future experiments.

Even without further improvements in the DI system (and subsequent additional improvements in particle identification rates), the potential value of the BAMS system for many applications should not be underestimated. Even if it proves difficult, for example, to consistently differentiate a *B. anthracis* spore from a close relative like *B. cereus*, neither of these types of particles should suddenly appear in the middle of a city, or in the air conditioning system of a building. A BAMS system would know if they did. In many, if not most, scenarios the differentiation of very similar particle types may not be particularly important. For every real bioterrorist event there will be hundreds if not thousands of hoaxes and most of these will be very crude. Even the current BAMS systems have no problems differentiating common white powders (e.g. baking powder, Equal, Gold Bond, etc.) from bacterial spores and other worrisome biological particles. There is no question that a BAMS system will be far more selective than current trigger systems already in use like BAWS.

Once a better mass spectrometer becomes available for experiments, some of the alternative DI laser systems introduced in chapter 6 will be reevaluated. The short pulse laser system and the IR OPO may both prove useful. The absorption of DPA, amino acids and many other molecules are known to be much stronger at wavelengths below 266 nm so these wavelengths will be explored as well. VUV wavelengths may be investigated since single photon ionization has been shown elsewhere to produce less fragmentation of large mass ions.

Another alternative is to employ more than a single DI laser pulse. Some interesting results have already been obtained from the combination of an IR pulse for desorption followed by a weak 266 nm pulse for ionization. In applications where the use of a reagent is not problematic, “on-the-fly” MALDI is another DI method that is quite likely to produce significant improvements.

Other components of the BAMS instrumentation are also rapidly improving. The BAMS group is designing and building a whole new generation of BAMS instruments. A new particle inlet will be built to better focus particles. New tracking stages and prescreeners have been built and are being implemented. A new mass spectrometer with improved high mass efficiency has been built and is in testing. Even the data analysis will be made faster and more efficient. The end result of all these efforts will be a far more efficient and sensitive instrument. Using a model like that described in chapter 7, the performance of the system can be quantitatively predicted and even tailored for specific scenarios. In fact, portions of the model itself will likely be integrated with the instrument software to dynamically define the conditions necessary to sound an alarm.

In the not too distant future, a fieldable version of the BAMS instrument may be used to detect aerosols released from airplanes or from envelopes. The system could be deployed outdoors, in the middle of a city, or indoors, monitoring the air conditioning system of a large building. One day, a BAMS system may even be able to analyze human effluents directly and screen them for the presence of pathogens or disease markers. A person without any outward symptoms of a respiratory infection might cough into an instrument and instantly obtain a diagnosis. BAMS detectors built for this application could be placed at borders, airports and hospitals. In the more distant future, the ability to rapidly analyze single cells by bioaerosol mass spectrometry may even enable ultra-sensitive screening for cancer.

## Appendix:

All of the code shown in this appendix was written for use with MATLAB (version 6.5 from The Mathworks, Inc.). With the exception of the material in section A.7, the following fragments of code do not constitute independently useful programs. They are merely excerpts from a much larger volume of code that have been included to clarify the comments and descriptions given in the body of the thesis (primarily chapters 3 and 7). All text following a “%” character represents author comments. In some cases the comments continue to the next line even though the next line is not marked with a separate “%” character. This is a result of the limited column width available on a printed page. (Many of the original single lines of code in fact have been wrapped and now occupy multiple lines.) This is unfortunate but the wrapping is easily spotted.

The final section of the appendix (A.8) contains a simple mathematical derivation relevant to the material in chapter 7 and the performance model in section A.7.

## A.1 Smoothing Function

This algorithm convolves the raw spectral data from a single polarity with a narrow Gaussian function to reduce random point-to-point fluctuations. “Polarity” is a structured array with multiple fields. The “Data” field (i.e. Polarity.Data) holds the raw mass spectral data for one polarity.

---

```
function Polarity = BoxcarAverage(Polarity);

if Polarity.BoxcarAverage == 0;           %THEN DO GAUSSIAN SMOOTHING
    GausW=5;                               %THE GAUSSIAN ARRAY WIDTH - NEEDS TO BE ODD NUMBER
    GausH=ceil(GausW/2);                   %THE LARGE HALF (3)
    GausL=GausW-GausH;                     %THE SMALL HALF (2)
    GausCar=exp(-(1:GausW)-GausH).^2/1^2); %THE GAUSSIAN SMOOTHING FUNCTION
    GausCar=GausCar/sum(GausCar);          %NORMALIZE

    TempData=[ones(GausH,1)*Polarity.Data(1); Polarity.Data;
ones(GausH,1)*Polarity.Data(end)];
    GausData=zeros(length(Polarity.Data),length(GausW));
    for I=1:GausW;
        GausData(:,I)=TempData(I:(end-GausW+I),1);
    end
    Polarity.Data=GausData*GausCar;        %THE FINAL SMOOTHED DATA
end

return
```

---

## A.2 Baseline Subtraction Function

This simple algorithm subtracts the nearly constant baseline from the raw mass spectral data. AMSData is an array holding the raw data from one half spectrum (AMSData is equivalent to Polarity.Data in A.1).

---

```
function AMSData = SubtractBaseline(AMSData);

Y1=sort(AMSData(1001:5000));
Y1=mean(Y1(1001:2000)); %the average value of the lowest 25%-50% of the data between 1001
and 5000
X1=3000; %the "time" associated with the new baseline value

Y2=sort(AMSData((end-4000):end));
Y2=mean(Y2(1001:2000)); %the average value of the lowest 25%-50% of the last 4000 pts
X2=length(AMSData)-2000; %the "time" associated with the new baseline value

AMSData=AMSData-( (Y2-Y1)/(X2-X1)*((1:length(AMSData))-X1)+Y1 )'; %subtract interpolated
baseline from data
AMSData(1:1000)=zeros(1000,1); %the first data points don't hold data

return
```

---



### A.3 Autocalibration Function

The following functions refine an average starting calibration for individual mass spectra. MassFit is basically the “calibration quality function” referred to in chapter 3. By minimizing the value of MassFit, improved calibration parameters are generally obtained for the data held in Polarity.

Polarity is the same structured variable as in A.1. HalfCalibration is a structured variable that contains the initial calibration information for the mass spectral data held in Polarity.

---

```
function HalfCalibration = AutocalibratePolarity(Polarity, HalfCalibration);

if Polarity.AutoCal == -1;
    return
end

OldEqn=HalfCalibration.Equation; %[a,b] where  $m^{1/2}=at+b$ 
OldRevEqn=HalfCalibration.ReverseEquation; %[c,d] where  $t=cm^{1/2}+d$    $c=1/a$    $d=-b/a$ 

CnstMass=170; %The upper mass at which to place a constraint on the motion
CnstLim =0.4; %The maximum amount by which the mass can be moved in one direction

Mass15Time=round(polyval(OldRevEqn,sqrt(15))); %Min. time passed to fit thing (also
serves as the lower constraint point)
MassCnstTime=round(polyval(OldRevEqn,sqrt(CnstMass))); %Constraint at high mass
MaxTime=round(polyval(OldRevEqn,sqrt(250))); %Max time passed to the fit thing (~m/z=250)

FitTime=(Mass15Time:MaxTime)'; %All the times we might want to consider
FitData=Polarity.Data(FitTime); %All the data we might want to consider
GoodPts=find( FitData>=Polarity.Threshold ); %Ignore small pts

if length(GoodPts)==0;
    disp('Signal level too small to auto calibrate - calibration unchanged');
    return;
end

FitTime=FitTime(GoodPts);
FitData=FitData(GoodPts).^0.5; %take the square root to “normalize” peak heights

%the problem with fminsearch is that it just finds the local minimum. Unfortunately the
correct
%calibration may not be the closest local minimum, therefore I'll run the thing three
times with
%slightly different initial conditions to try to find the true minimum we're looking for
[NewEqn1,Val1]=fminsearch(@MassFit,OldEqn,[],FitTime,FitData,Mass15Time,MassCnstTime,Cnst
Mass,CnstLim); %[a,b] where  $m^{1/2}=at+b$ 
[NewEqn2,Val2]=fminsearch(@MassFit,[ (sqrt(CnstMass+CnstLim)-OldEqn(2))/MassCnstTime
,OldEqn(2)],[],FitTime,FitData,Mass15Time,MassCnstTime,CnstMass,CnstLim); %[a,b] where
 $m^{1/2}=at+b$ 
[NewEqn3,Val3]=fminsearch(@MassFit,[ (sqrt(CnstMass-CnstLim)-OldEqn(2))/MassCnstTime
,OldEqn(2)],[],FitTime,FitData,Mass15Time,MassCnstTime,CnstMass,CnstLim); %[a,b] where
 $m^{1/2}=at+b$ 
```

```
[minVal,minInd]=min([Val1,Val2,Val3]);
NewEqn=[NewEqn1; NewEqn2; NewEqn3];
NewEqn=NewEqn(minInd,:); %so this should be the correct calibration
```

```
HalfCalibration.Equation = NewEqn;
HalfCalibration.ReverseEquation = [1/NewEqn(1), -NewEqn(2)/NewEqn(1)];
HalfCalibration.Mass = [50, 100];
HalfCalibration.ChannelNumber = polyval(HalfCalibration.ReverseEquation,
sqrt(HalfCalibration.Mass));
HalfCalibration.ZeroPoint = polyval(HalfCalibration.ReverseEquation, 0);
```

```
return
```

```
function MassMissMatch =
MassFit(Param,FitTime,FitData,Mass15Time,MassCnstTime,CnstMass,CnstLim);
    Mass=(Param(1)*FitTime+Param(2)).^2; %the masses for each data element
    Weight=exp(-(Mass-round(Mass)).^2/(0.15)^2); %Gaussians with 0.3 1/e full widths
    SigInRange=FitData'*Weight; %stuff falling within +/-0.15 of integer mass

    Mass15Shift=(((Param(1)*Mass15Time+Param(2))^2-15)/0.4)^10+1; %blows up if small mass
    peaks move too much
    MassCnstShift=(((Param(1)*MassCnstTime+Param(2))^2-CnstMass)/(1.2*CnstLim))^10+1;
    %blows up if the big mass peaks move too much

    MassMissMatch=MassCnstShift*Mass15Shift/SigInRange;
return
```

## A.4 Peak Finding and Vector Formation Function

The following function identifies all of the peaks in a half spectrum and then produces a vector representation of that spectrum based on those peaks. The actual peak finder was written by David Fergenson. `Polarity.AreaBar` and `Polarity.HeightBar` are the “vector” representations of the spectra referred to frequently in the text and were implemented in their current form by the author. Only `Polarity.AreaBar` is typically used.

---

```
function Polarity = FindAMSPeaks(Polarity, HalfCalibration);

%---find where each peak starts and stops
HighPointIndex = find(Polarity.Data>=Polarity.Threshold);
if isempty(HighPointIndex); %if none of the points are above threshold then stop
    Polarity.Mass = [];
    Polarity.Area = [];
    Polarity.Height = [];
    Polarity.AreaBar = zeros(1, Polarity.MaxMass);
    Polarity.HeightBar = zeros(1, Polarity.MaxMass);
else %if there is at least one peak
    if length(HighPointIndex) == 1; %if there is only one peak
        Start = 1;
        Stop = 1;
    else %if there are multiple peaks
        Edges = find((HighPointIndex(2:length(HighPointIndex)) -
HighPointIndex(1:length(HighPointIndex) -1)) > 1);
        if isempty(Edges);
            Start = 1;
            Stop = length(HighPointIndex);
        else
            Start = [1; Edges+1];
            Stop = [Edges; length(HighPointIndex)];
        end
    end
end

%---figure out peak areas, heights, locations
for J = 1:length(Start);
    Polarity.Area(J) =
sum(Polarity.Data(HighPointIndex(Start(J)):HighPointIndex(Stop(J))));
    Polarity.Height(J) =
max(Polarity.Data(HighPointIndex(Start(J)):HighPointIndex(Stop(J))));
    TempMass = ChannelNumber2Mass(HighPointIndex(Start(J)):HighPointIndex(Stop(J)),
HalfCalibration);
    Polarity.Mass(J) =
mean(TempMass(find(Polarity.Data(HighPointIndex(Start(J)):HighPointIndex(Stop(J))) ==
Polarity.Height(J))));
end

%---Basically treat each peak like a vector and rotates it as necessary. Don't rotate
the vector at all unless the peak is more
%than 0.3 units from an integer mass. If a peak has a half integer mass exactly, the
two bins on
%either side get equal contributions from the peak
Polarity.AreaBar = zeros(1, Polarity.MaxMass);
Polarity.HeightBar = zeros(1, Polarity.MaxMass);
```

```

PkMass=Polarity.Mass; %the mass of the peaks
LoMass=floor(PkMass);
HiMass=LoMass+1;
RotAngle=pi/2*(PkMass-LoMass-0.3)./0.4; %~angle to rotate the peak vector
RotAngle(find(RotAngle<0))=0; %make negative rotations equal to zero
RotAngle(find(RotAngle>pi/2))=pi/2; %make big rotations=pi/2
LoWeight=cos(RotAngle);
HiWeight=sin(RotAngle);

for J=1:length(PkMass); %step through all the peaks identified
    if LoMass(J) >=1 & LoMass(J) <= Polarity.MaxMass;
        Polarity.AreaBar(LoMass(J)) =Polarity.AreaBar(LoMass(J)) +Polarity.Area(J)
*LoWeight(J);

Polarity.HeightBar(LoMass(J))=Polarity.HeightBar(LoMass(J))+Polarity.Height(J)*LoWeight(J)
);
        end
        if HiMass(J) >=1 & HiMass(J) <= Polarity.MaxMass;
            Polarity.AreaBar(HiMass(J)) =Polarity.AreaBar(HiMass(J)) +Polarity.Area(J)
*HiWeight(J);

Polarity.HeightBar(HiMass(J))=Polarity.HeightBar(HiMass(J))+Polarity.Height(J)*HiWeight(J)
);
        end
    end
end
return;

```

---

## A.5 New Baseline and Vector Formation Function

This function subtracts the baseline offset from the data more accurately and then calculates new vectors (based effectively on the peak area) without actually identifying the peaks. It assumes that all the relevant data is already loaded in the BLOCK variable.

---

```
function MakeNewBaselineAndVectors;

global BLOCK
[BN,ok] = listdlg('ListString',{BLOCK.Name},'SelectionMode','single','Name','Select block
to fix','ListSize',[250,300]); %BN is the block number
if ~ok
    return
end

NumHunks=10; %number of hunks to divide each polarity into
MaxMass=400; %the max mass-to-charge ratio included in the area bar
w=0.1; %half width of rising/falling edge of data weight array thing
CentMass=1:MaxMass; %note that the minimum mass is always one
LtBotMass=(CentMass-0.5-w).^0.5; %the square root of the mass of the bottom left edge of
the weight array
LtTopMass=(CentMass-0.5+w).^0.5;
RtTopMass=(CentMass+0.5-w).^0.5;
RtBotMass=(CentMass+0.5+w).^0.5;

WaitHandle = waitbar(0,['Fixing ',BLOCK(BN).Name,'...']);
NumSpec=length(BLOCK(BN).Spectrum);
for j = 1:NumSpec

%-----Positive ions-----

LtBotTime=round(BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(1).*LtBotMass+
BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(2));

LtTopTime=round(BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(1).*LtTopMass+
BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(2));

RtTopTime=round(BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(1).*RtTopMass+
BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(2));

RtBotTime=round(BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(1).*RtBotMass+
BLOCK(BN).Spectrum(j).Calibration.Positive.ReverseEquation(2));
    MinTime=LtBotTime(1);
    MaxTime=MinTime+NumHunks*ceil( (RtBotTime(end)-MinTime+1)/NumHunks )-1; %this makes
sure the number of channels is evenly divisible by NumHunks...

    Data=reshape( double(BLOCK(BN).Spectrum(j).Pos.Data(MinTime:MaxTime)) ,[],NumHunks);
%reshape the data into an array with NumHunks columns
    HunkSize=size(Data,1);
    Center=sort(Data); %sort the data in each hunk
    Center=mean( Center( round(0.3*HunkSize):round(0.5*HunkSize),:) ); %the value of the
baseline at the center of each hunk of data (the range 0.3-0.5 can be changed)
    LtEdge=( [Center(1),Center(1:(end-1))] +Center)/2; %the value of the baseline at the
left edge of each hunk of data
    RtEdge=(Center+[Center(2:end),Center(end)])/2; %the value of the baseline at the
right edge of each hunk of data
end
end
```

```

%linear interpolation from center to center
CentLt=round(HunkSize/2); %the index of the last point on the left half of the hunk
(the one closest to center)
RtSize=HunkSize-CentLt; %the size of the right half of the hunk
BaseLine=0*Data; %an empty array the same size as Data
BaseLine(1:CentLt,:)=((1:CentLt)'./CentLt)*(Center-LtEdge)+ones(CentLt,1)*LtEdge;
BaseLine((CentLt+1):end,:)=((1:RtSize)'./RtSize)*(RtEdge-
Center)+ones(RtSize,1)*Center;

Data=reshape( Data-BaseLine ,[],1); %shape the data back into a column array
% BLOCK(BN).Spectrum(j).Pos.Data(MinTime:MaxTime)=Data; %I don't really like this
much... seems like we should fix the whole baseline or else leave the data alone

bin_sums=zeros(size(CentMass));
for i = 1:length(bin_sums)
    TmpData=Data( (LtBotTime(i):RtBotTime(i)) -MinTime+1)'; %this is all the data we
need for mass i
    LtPts=( LtBotTime(i) : LtTopTime(i)) -LtBotTime(i)+1;
    CtPts=((LtTopTime(i)+1):(RtTopTime(i)-1))-LtBotTime(i)+1;
    RtPts=( RtTopTime(i) : RtBotTime(i)) -LtBotTime(i)+1;

    LtSum=sum( TmpData(LtPts).*((1:length(LtPts))-1)./(length(LtPts)-1) ); %weight
from 0-1
    CtSum=sum( TmpData(CtPts) ); %weights are all 1
    RtSum=sum( TmpData(RtPts).*(1-((1:length(RtPts))-1)./(length(RtPts)-1)) );
%weight from 1-0

    bin_sums(i) = LtSum+CtSum+RtSum;
end;
BLOCK(BN).Spectrum(j).Pos.AreaBar = bin_sums.*(bin_sums>0); %the ".*(bin_sums>0)"
sets negative values to zero...
BLOCK(BN).Spectrum(j).Pos.MaxMass=MaxMass;

%----Negative ions----

LtBotTime=round(BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(1).*LtBotMass+
BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(2));

LtTopTime=round(BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(1).*LtTopMass+
BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(2));

RtTopTime=round(BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(1).*RtTopMass+
BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(2));

RtBotTime=round(BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(1).*RtBotMass+
BLOCK(BN).Spectrum(j).Calibration.Negative.ReverseEquation(2));
MinTime=LtBotTime(1);
MaxTime=MinTime+NumHunks*ceil( (RtBotTime(end)-MinTime+1)/NumHunks )-1; %this makes
sure the number of channels is evenly divisible by NumHunks...

Data=reshape( double(BLOCK(BN).Spectrum(j).Neg.Data(MinTime:MaxTime)) ,[],NumHunks);
%reshape the data into an array with NumHunks columns
HunkSize=size(Data,1);
Center=sort(Data); %sort the data in each hunk
Center=mean( Center( round(0.3*HunkSize):round(0.5*HunkSize),: ) ); %the value of the
baseline at the center of each hunk of data (the range 0.3-0.5 can be changed)
LtEdge=(Center(1),Center(1:(end-1))]+Center)/2; %the value of the baseline at the
left edge of each hunk of data
RtEdge=(Center+[Center(2:end),Center(end)])/2; %the value of the baseline at the
right edge of each hunk of data

%linear interpolation from center to center
CentLt=round(HunkSize/2); %the index of the last point on the left half of the hunk
(the one closest to center)
RtSize=HunkSize-CentLt; %the size of the right half of the hunk
BaseLine=0*Data; %an empty array the same size as Data
BaseLine(1:CentLt,:)=((1:CentLt)'./CentLt)*(Center-LtEdge)+ones(CentLt,1)*LtEdge;
BaseLine((CentLt+1):end,:)=((1:RtSize)'./RtSize)*(RtEdge-
Center)+ones(RtSize,1)*Center;

Data=reshape( Data-BaseLine ,[],1); %shape the data back into a column array

```

```

%      BLOCK(BN).Spectrum(j).Neg.Data(MinTime:MaxTime)=Data; %I don't really like this
much... seems like we should fix the whole baseline or else leave the data alone

    bin_sums=zeros(size(CentMass));
    for i = 1:length(bin_sums)
        TmpData=Data( (LtBotTime(i):RtBotTime(i)) -MinTime+1)'; %this is all the data we
need for mass i
        LtPts=( LtBotTime(i) : LtTopTime(i)) -LtBotTime(i)+1;
        CtPts=((LtTopTime(i)+1):(RtTopTime(i)-1))-LtBotTime(i)+1;
        RtPts=( RtTopTime(i) : RtBotTime(i)) -LtBotTime(i)+1;

        LtSum=sum( TmpData(LtPts).*((1:length(LtPts))-1)./(length(LtPts)-1) ); %weight
from 0-1
        CtSum=sum( TmpData(CtPts) );
        RtSum=sum( TmpData(RtPts).*(1-((1:length(RtPts))-1)./(length(RtPts)-1)) );
%weight from 1-0

        bin_sums(i) = LtSum+CtSum+RtSum;
    end;
    BLOCK(BN).Spectrum(j).Neg.AreaBar = bin_sums.*(bin_sums>0); %the ".*(bin_sums>0)"
sets negative values to zero...
    BLOCK(BN).Spectrum(j).Neg.MaxMass=MaxMass;

    waitbar(j/NumSpec,WaitHandle);
end
close(WaitHandle)

BLOCK(BN).Name=[BLOCK(BN).Name, '-NewAreaBase'];

```

---

## A.6 The BART Clustering Function

The following code clusters similar spectra together based on the angle between the vectors (which were defined by A.4 or A.5). ART2aWorkData holds the vectors from each particle to be clustered. The WeightMatrix variable holds the neurons that represent each cluster.

---

```
function [ART2aClusterData, WeightMatrix, Iteration] = BartCore(ART2aWorkData,
VigilanceFactor, LearningRate, MaxIterations);

VigAng=acos(VigilanceFactor)*180/pi;

load c:\matlabprograms\Others\randomstate.mat; %tells the rand function where to start
rand('state', RandomState); %this starts in the same place every time

WaitHandle = waitbar(0, 'BART Calculation In Progress... (2/3)');

for Iteration = 1:MaxIterations;
    clear ART2aClusterData; %Dump details of the last round (keep weight vectors)

    SortOrder = randperm(size(ART2aWorkData.SpectrumNumber, 2)); %p = randperm(n) returns
a random permutation of the integers 1:n.
    ART2aWorkData.PosData = ART2aWorkData.PosData(SortOrder, :); %reorder all the
positive spectra
    ART2aWorkData.NegData = ART2aWorkData.NegData(SortOrder, :); %put the negative in the
same order
    ART2aWorkData.SpectrumNumber = ART2aWorkData.SpectrumNumber(SortOrder); %keep the
spectrum numbers associated with the right spectra

    for I = 1:size(ART2aWorkData.PosData, 1); %step through each spectrum
        if ~exist('WeightMatrix'); %for the first spectrum of the first iteration, the
weightmatrix isn't defined
            WeightMatrix.Pos = ART2aWorkData.PosData(I, :); %so the very first cluster is
just the first spectra
            WeightMatrix.Neg = ART2aWorkData.NegData(I, :);
            ART2aClusterData{1} = ART2aWorkData.SpectrumNumber(I);
        else
            PosAng=acos(ART2aWorkData.PosData(I, :) * WeightMatrix.Pos')*180/pi; %angle
between current spectrum and current weights (=>a row with a col for each weight)
            NegAng=acos(ART2aWorkData.NegData(I, :) * WeightMatrix.Neg')*180/pi;

            SimScore=max([PosAng;NegAng]); %simply find which angle is bigger and use
that as the similarity score
            SimScore(find(PosAng > VigAng | NegAng > VigAng))=-1; %SimScore must be >=0
so the negative values will be used to exclude the bad points (with Ang>VigAng)
            SimScore=1./(SimScore+0.01); %We can now find the max of this, if it's
negative that means no match. (0.01 just prevents div by zero in case of perfect match)

            [Max, MaxIndex] = max(SimScore); %find the max SimScore and where that value
is in the row
            if Max > 0; %So if there is a match do the following (if either angle is
>VigAng then Max will be <0)
                if exist('ART2aClusterData');
                    if MaxIndex > length(ART2aClusterData); %this condition is possible
because ART2aClusterData is erased every iteration while WeightMatrix remains
```





```

        if NumI >= NumJ
            NumPosPts=size(ART2aWorkData.PosData,2);
            NumNegPts=size(ART2aWorkData.NegData,2);
            PosSpec=zeros(NumJ,NumPosPts);
            NegSpec=zeros(NumJ,NumNegPts);
            for K=1:NumJ

SpecIndex=find(ART2aWorkData.SpectrumNumber==ART2aClusterData{J}(K));
                PosSpec(K,:)=ART2aWorkData.PosData(SpecIndex, :); %load the
specs from WorkData into PosSpec and NegSpec
                NegSpec(K,:)=ART2aWorkData.NegData(SpecIndex, :);
            end
            PosSimilarities = PosSpec * WeightMatrix.Pos(I,:); % => column
vector with similarity scores
            NegSimilarities = NegSpec * WeightMatrix.Neg(I,:);
            if min(PosSimilarities) >= VigilanceFactor & min(NegSimilarities)
>= VigilanceFactor; %if all the spectra are close enough to the WeightVector
                %Time to join J to I and delete J
                WeightMatrix.Pos(I, :) = ( WeightMatrix.Pos(I, :)*NumI +
sum(PosSpec,1) )/(NumI+NumJ); %The weighted average of the big cluster weight vector and
the average spectrum from the little cluster
                WeightMatrix.Pos(I, :) = WeightMatrix.Pos(I,
:)./norm(WeightMatrix.Pos(I, :)); %renormalize the thing
                WeightMatrix.Neg(I, :) = ( WeightMatrix.Neg(I, :)*NumI +
sum(NegSpec,1) )/(NumI+NumJ); %The weighted average of the big cluster weight vector and
the average spectrum from the little cluster
                WeightMatrix.Neg(I, :) = WeightMatrix.Neg(I,
:)./norm(WeightMatrix.Neg(I, :)); %renormalize the thing

ART2aClusterData{I}=[ART2aClusterData{I},ART2aClusterData{J}];

                WeightMatrix.Pos(J, :)=[];
                WeightMatrix.Neg(J, :)=[];
                ART2aClusterData(J)=[];
                disp('Joined similar clusters');
            end
        else %if NumJ>NumI
            NumPosPts=size(ART2aWorkData.PosData,2);
            NumNegPts=size(ART2aWorkData.NegData,2);
            PosSpec=zeros(NumI,NumPosPts);
            NegSpec=zeros(NumI,NumNegPts);
            for K=1:NumI

SpecIndex=find(ART2aWorkData.SpectrumNumber==ART2aClusterData{I}(K));
                PosSpec(K,:)=ART2aWorkData.PosData(SpecIndex, :); %Add each
new spectrum to a new row
                NegSpec(K,:)=ART2aWorkData.NegData(SpecIndex, :);
            end
            PosSimilarities = PosSpec * WeightMatrix.Pos(J,:); %get column
vector with similarity scores
            NegSimilarities = NegSpec * WeightMatrix.Neg(J,:);
            if min(PosSimilarities) >= VigilanceFactor & min(NegSimilarities)
>= VigilanceFactor; %if all the spectra are close enough to the WeightVector
                %Time to join J to I and delete J
                WeightMatrix.Pos(I, :) = ( WeightMatrix.Pos(J, :)*NumJ +
sum(PosSpec,1) )/(NumI+NumJ); %The weighted average of the big cluster weight vector and
the average spectrum from the little cluster
                WeightMatrix.Pos(I, :) = WeightMatrix.Pos(I,
:)./norm(WeightMatrix.Pos(I, :)); %renormalize the thing
                WeightMatrix.Neg(I, :) = ( WeightMatrix.Neg(J, :)*NumJ +
sum(NegSpec,1) )/(NumI+NumJ); %The weighted average of the big cluster weight vector and
the average spectrum from the little cluster
                WeightMatrix.Neg(I, :) = WeightMatrix.Neg(I,
:)./norm(WeightMatrix.Neg(I, :)); %renormalize the thing

ART2aClusterData{I}=[ART2aClusterData{I},ART2aClusterData{J}];

                WeightMatrix.Pos(J, :)=[];
                WeightMatrix.Neg(J, :)=[];

```

```

                ART2aClusterData(J)=[];
                disp('Joined similar clusters');
            end
        end
    end %of matching weight vector if statement
end %of J for loop
I=I+1;
end %of while loop
end %of length>=2 if statement

for I=1:length(ART2aClusterData); %Sort the spectrum numbers so that they're
consistent from iteration to iteration before we compare below
    ART2aClusterData{I}=sort(ART2aClusterData{I});
end

if exist('BackupOfClusterData'); % All of this just asks if ART2aclusterData and
BackupOfClusterData are identical
    if length(BackupOfClusterData) == length(ART2aClusterData); %if you got the same
number of clusters you did last time
        for I = 1:length(BackupOfClusterData);
            if length(BackupOfClusterData{I}) == length(ART2aClusterData{I}); %if the
number of spectra in the clusters are the same
                if BackupOfClusterData{I} ~= ART2aClusterData{I};
                    clear BackupOfClusterData;
                    break
                end
            else
                clear BackupOfClusterData;
                break
            end
        end
    else
        clear BackupOfClusterData;
    end
end

%at this point, if the Backup wasn't identical, it was cleared
if exist('BackupOfClusterData'); %so if it was identical
    break; %stop iterating
end
BackupOfClusterData = ART2aClusterData;
WaitBar((Iteration/MaxIterations), WaitHandle);
end

disp(['Reached Iteration ',num2str(Iteration),' out of a maximum of
',num2str(MaxIterations)]); %It seems pretty rare that this is ever less than the max

close(WaitHandle);

return;

```

---

## A.7 Performance Model

The following is a fully functional model of a hypothetical BAMS instrument (described in chapter 7). By design, the version presented here does not contain or utilize actual performance parameters of a real BAMS system (or its components) and thus does not accurately predict the sensitivity of a real system to an actual agent aerosol.

---

```
function BAMSThesisModel
%This is basically just a wrapper where the aerosol and instrument parameters are
defined. It calls other
%functions to do all the real work and then plots the results.

%-----Probability of detection and false alarm
ReqPfa=1e-4;                               %The required probability of false
alarm                                       alarm
ReqPd=0.90;                                %The required probability of
detection

%-----Aerosol parameters
Aerosol.Name='Generic Aerosol';
Aerosol.Type ={'Agent' 'Bio' 'Non-Bio'};    %Types of background particles
Aerosol.AD =1;                             %Aerodynamic diameter represented by
each row of each array (the values chosen are independent of the aerosol to be studied)
Aerosol.ADDst=[1 1 1];                     %Size Distributions for each particle
type (sum=1) [fraction of particles at size]
Aerosol.Conc =[1e1 1e1 1e3];               %Total concentration of each particle
type [particles/liter]
Aerosol.SctMn=[1.0 1.0 1.0];               %Mean scattered signal from lum
particle [arb units]
Aerosol.Pps =[.90 .90 .10];                %Prob that particle passes
prescreener
Aerosol.Pion =[1.0 1.0 1.0];               %Prob that enough ions are produced
when hit by DI laser to generate a spectrum
Aerosol.Pagnt=[0.9 1e-3 1e-5];             %Prob that each type is identified
(or misidentified) as agent (independent of size)

%-----Instrument parameters
Instrument.Name='BAMS';                     %BAMS system described in thesis
Instrument.VIon=1;                           %if =1 then virtual impactor used, if
=0 then not used
Instrument.VIeff=100;                        %Size dependent virtual impactor
efficiency [concentration factor]
Instrument.SctThresh=0;                      %Threshold to be detected by tracking
stage [arb units]
Instrument.PSon=1;                           %if =1 then prescreening used, if =0
then not used
Instrument.PSRep=1e4*60;                     %Rep rate of prescreener
(shots/minute)
Instrument.DIRep=1e3*60;                     %Rep rate of DI laser (shots/minute)
Instrument.DAQRep=1e2*60;                   %Rep rate of DAQ system
(spectra/minute)
Instrument.Pfocus=0.25;                     %size dependent probability that
particle is focused well enough to hit first tracking laser
Instrument.Ptrack=0.40;                      %Size dependent probability that
particle is tracked (assuming that it hit first laser)
Instrument.Ppresc=0.40;                      %Size dependent probability that
particle is hit by prescreening laser(s)
```

```

Instrument.PhitDI=0.25; %Size dependent probability that
particle is hit by DI laser(s)
Instrument.SampRate=1; %Sampling Rate [liters/minute]
Instrument.SampTime=1; %Sampling Time [minutes]
Instrument.MinMaxAD=[0.7 10]; %Min and Max Aero Diam of particles
allowed to be identified as agent
TrackRep=1e5*60; %Rep rate of tracking system
[tracks/min]
Instrument.NumParts=0:5e5:1e8; %number of particles arriving per
minute
Instrument.GudTrack=TrackRep./(Instrument.NumParts+TrackRep); %fraction properly tracked
(will trig and theoretically intercept laser)
Instrument.BadTrack=Instrument.NumParts*0; %fraction tracked improperly (will
trig, but not intercept laser)

```

```

%-----Calculate ROC data
disp('Calculating ROC data');
[ROCPfa,ROCPd]=CalcROCDData(Aerosol,Instrument);

```

```

%-----Make a ROC plot
figure('Name',['ROC Curve - ',Aerosol.Name,' ',Instrument.Name],'Position',[100 100 700
700])
axes('FontSize',16)
semilogx([ROCPfa,1],[ROCPd,1],'o','LineWidth',2); % (1,1) always present
ylim([0,1]);
axis square
xlabel('Pfa')
ylabel('Pd')
title(['Predicted BAMS ROC curve']);
text(min(ROCPfa),0.9,{[Instrument.Name,' ',Aerosol.Name]; ['Non-bio:
',num2str(Aerosol.Conc(3)),' Particles/Liter']; ['Bio: ',num2str(Aerosol.Conc(2)),'
Particles/Liter']; ['Agent: ',num2str(Aerosol.Conc(1)),' Particles/Liter']; ['Time:
',num2str(Instrument.SampTime),' Minutes'] }); %,'FontName','FixedWidth');

```

```

%-----Calculate ACPLA data
disp('Calculating ACPLA data');
[ACPLA,TotPLA]=CalcACPLAData(Aerosol,Instrument,ReqPd,ReqPfa);

```

```

%-----Make an ACPLA plot
ZeroLine=10.^(0:0.5:7);

figure('Name',['ACPLA Plot - ',Aerosol.Name,' ',Instrument.Name],'Position',[200 200 700
700])
axes('FontSize',16)
loglog(ACPLA,TotPLA,ZeroLine,ZeroLine,'LineWidth',2);
xlim([ZeroLine(1),ZeroLine(end)]);
ylim([ZeroLine(1),ZeroLine(end)]);
set(gca,'ytick',10.^(0:7),'xtick',10.^(0:7))
axis square
grid on
grid minor
xlabel('Agent Concentration [# /liter]')
ylabel('Total Concentration [# /liter]')
title(['Predicted BAMS Performance']);
text(min(ZeroLine)*2,max(ZeroLine)/4,{[Instrument.Name,' ',Aerosol.Name];
['Tsamp=',num2str(Instrument.SampTime),' Minutes']; ['Pfa<=',num2str(ReqPfa)];
['Pd>=',num2str(ReqPd)]});

```

---

```

function [Pfa,Pd]=CalcROCDData(Aerosol,Instrument)
%This produces the data for a ROC curve based on a specific aerosol scenario.

%Aerosol contains all the aerosol parameters
%Instrument contains all the instrument parameters

%Pfa Pfa(i) is the probability that the background will produce i or more
agent identifications

```

```

%Pd          Pd(i) is the probability that the agent and background combined will
produce i or more agent identifications

MinPfa=10^-6; %this roughly sets the range of the ROC curve (one point on ROC curve
must have this Pfa or less)

%-----Based just on the background, figure out how likely a false alarm is for different
alarm thresholds

BGCols=2:length(Aerosol.Conc); %The columns representing background particle types
(column 1 should always be the agent)

[Nbg,NbgStatus]=FindAvgNumAnalyzed(Aerosol,Instrument,BGCols); %Avg # of each type of
background particle analyzed per sample period.

if NbgStatus==0; disp('Warning: ROC curve may be inaccurate! Exceeded range of tracking
data. (CalcROCData 1)'); end;

[Alarmk,Pfa] = FindNumToAlarm(MinPfa,Aerosol.Pagnt(BGCols),Nbg); %Pfa holds the prob
that >=1, >=2, =>3... background particles are misidentified (Prob >=0 misID =1)

%-----Including both the agent and background, figure out how likely a detection is for
different alarm thresholds

AllCols=1:length(Aerosol.Conc); %The columns for every particle type (agent and
background)

[Nall,NallStatus]=FindAvgNumAnalyzed(Aerosol,Instrument,AllCols); %The average number of
each type of particle analyzed per sample period.

if NallStatus==0; disp('Warning: ROC curve may be inaccurate! Exceeded range of tracking
data. (CalcROCData 2)'); end;

k=length(Pfa); %The max k value represented by an element of Pfa

Pd = FindProbDetect(Nall,Aerosol.Pagnt,k); %Prob to detect vector (i.e. prob identify
>=1..k agent and background particles as agent)

```

---

```

function Pd = FindProbDetect(Ni,Pai,k)
%This is closely related to the stuff in FindMinAvgNumAgent.m and FindNumToAlarm.m

%Pd          probability of detection (i.e. sounding alarm)

%Ni          avg number of each type of particle analyzed (includes agent)
%Pai        probability that each type is ID'd as agent (includes agent)
%k          detection thresholds (i.e. sound alarm if >=k particles ID'd as agent)

N=sum(Pai.*Ni); %Avg # of particles (agent and background)
identified as agent

PoisProb=exp(-N); %Pois prob that exactly 0 particles ID'd as
agent
Pgreaterk(1)=1-PoisProb; %Prob that 1 or more particles ID'd as agent

for i=1:(k-1)
    PoisProb=PoisProb*N/i; %prob that exactly i particles ID'd as agents
    Pgreaterk(i+1)=Pgreaterk(i)-PoisProb; %Prob that i+1 or more particles ID'd as
agent
end

Pd=Pgreaterk; %Prob of detection is equivalent to prob >=k
particles ID'd as agent

```

---

```

function varargout = FindAvgNumAnalyzed(Aerosol,Instrument,Cols)
%varargout{1}=N           the average number of each type of particle analyzed by the
BAMS system in one sample period
%varargout{2}=status     everything okay => status=1, something bad happened =>
status=0

%Aerosol           structured variable that holds all the aerosol parameters
%Instrument        structured variable that holds all the instrument parameters
%Cols             vector of data columns to include in calculations (all particles or just
background particles)

%-----Make a few vectors for convenience
OneRow=ones(size(Cols));           %row of
ones                               %column
OneCol=ones(size(Aerosol.AD));
of ones

if nargout>1; varargout{2}=1; end; %if two
outputs are desired, the second arguement is the status and at this point it's okay (=1)

%-----Now figure out how many of each particle type get analyzed
N=(OneCol*Aerosol.Conc(Cols)).*Aerosol.ADDst(:,Cols); %Avg.
#/liter of each type and size of background particle [# /liter]

if Instrument.VIon
    N=N.*(Instrument.VIeff*OneRow); %If
virtual impactor is used, concentrate the aerosol...
end

N=N*Instrument.SampRate;           %Avg. #
of each type sucked into BAMS instrument in one minute [# /min]

N=N.*(Instrument.Pfocus*OneRow); %Avg. #
that make it to first tracking laser [# /min]

PassSct=((Aerosol.AD.^6)*Aerosol.SctMn(Cols))>=Instrument.SctThresh; %1 if
that size and type is detectable by tracking, =0 otherwise (need to add std dev stuff)
N=N.*PassSct;                     %Avg. #
that will scatter sufficiently to be detected by tracking system [# /min]

TotN=sum(sum(N));                  %Total #
of detectable particles delivered to first tracking laser [# /min]
if TotN>=max(Instrument.NumParts)
    disp('Warning: Rate of tracking exceeds tested limits!');
    if nargout>1; varargout{2}=0; end; %send
alarm that we're extrapolating past the end of the data... if the user wants to know
end

FracGudTrack=interp1(Instrument.NumParts,Instrument.GudTrack,TotN,'cubic'); %fraction
of the particles tracked properly
FracBadTrack=interp1(Instrument.NumParts,Instrument.BadTrack,TotN,'cubic'); %fraction
of the particles tracked improperly

N=N*(FracGudTrack+FracBadTrack); %Avg. #
tracked (both Gud and Bad will trigger lasers) based just on rate [# /min]

N=N.*(Instrument.Ptrack*OneRow); %Avg. #
tracked (both Gud and Bad) after geometrical divergence is accounted for [# /min]

if Instrument.PSon
    PfirePS=Instrument.PSRep/(sum(sum(N))+Instrument.PSRep); %Prob
that prescreening laser will fire at a given particle (rate constraint)

    N=N.*(Instrument.Ppresc*OneRow)*PfirePS*FracGudTrack/(FracGudTrack+FracBadTrack);
%Avg. # hit by prescreener (includes rate and geom. constraints) [# /min]

    N=N.*(OneCol*Aerosol.Pps(Cols)); %Avg. #
that pass prescreener [# /min]

```

```

end

PfireDI=Instrument.DIRep/(sum(sum(N))+Instrument.DIRep); %prob
that DI laser will fire at a given particle

if Instrument.PSon %If
prescreener was used, we don't have to worry about "BadTrack" particles
    N=N.*(Instrument.PhitDI*OneRow).*(OneCol*Aerosol.Pion(Cols))*PfireDI; %Avg. #
hit by DI laser that produce spectrum [# /min]
else
N=N.*(Instrument.Ppresc*OneRow).*(Instrument.PhitDI*OneRow).*(OneCol*Aerosol.Pion(Cols))*
PfireDI*FracGudTrack/(FracGudTrack+FracBadTrack); %Avg. # hit by DI laser that produce
spectra [# /min]
end

PfireDAQ=Instrument.DAQRep/(sum(sum(N))+Instrument.DAQRep); %Prob
that DAQ system is ready to acquire spectrum

N=N.*PfireDAQ; %Avg. #
fully analyzed by the system [# /min]

N=N.*((Aerosol.AD>=Instrument.MinMaxAD(1) & Aerosol.AD<=Instrument.MinMaxAD(2))*OneRow);
%Avg. # in the proper size range [# /min]

N=N*Instrument.SampTime; %Avg #
per sample period [#]

N=sum(N,1); %Sum
along the columns to get rid of size info which no longer matters

varargout{1}=N; %Return
the average number of each type of particle that gets analyzed per sample time

```

---

```

function [k,Pgreaterk] = FindNumToAlarm(Pfa,Pai,Ni)
% k          number of required agent IDs to ensure false alarm rate <=Pfa
% Pgreaterk(k) prob that >=k particles ID'd as agent

% Pfa          max desired probability of false alarm
% Pai          prob that particle type is midIDD as an agent. (1 x m)
% Ni          avg number of particles ID'd by mass spec per time period. (1 x m)

N=sum(Pai.*Ni); %Avg # of background particles misidentified

PoisProb=exp(-N); %Pois prob that exactly 0 particles ID'd as
agent
Pgreaterk(1)=1-PoisProb; %Prob that 1 or more particles ID'd as agent

i=1;
while Pgreaterk(end)>Pfa
    PoisProb=PoisProb*N/i; %prob that exactly i particles ID'd as agents
    Pgreaterk(i+1)=Pgreaterk(i)-PoisProb; %Prob that i+1 or more particles ID'd as
agent
    i=i+1;
end

k=max(3,length(Pgreaterk)); %So this defines our alarm threshold (i.e.
prob to obtain >=k agent ID's is <=Pfa) (must be at least 3)

```

---

```

function [ACPLA,TotPLA]=CalcACPLAData(Aerosol,Instrument,Pd,Pfa)
%This things calculates the data for ACPLA plots

%ACPLA          horizontal position of each point in ACPLA plot
%TotPLA         vertical position of each point in ACPLA plot

%Aerosol        all aerosol parameters
%Instrument      all instrument parameters
%Pd             minimum requested probability of detection
%Pfa           maximum requested probability of false alarm

```



```

%-----Figure out how many agent ID's needed and how much agent required to get it...
BGCols=2:length(Aerosol.Conc); %Columns representing
background particle types (column 1 is always agent)
Ci=Aerosol.Conc(BGCols); %The total concentration of
each background particle type [# /liter]
Cnrm=Ci./sum(Ci); %Nnormalized concentrations
(sum=1/liter)
Cmag=10.^(-1:0.05:7); %Concentration magnitude
(multiplied by Cnrm to determine actual concentrations)
NumConc=length(Cmag); %Number of concentrations to
try out

ACPLA=zeros(1,NumConc); %Will hold agent
concentration [# /liter]
TotPLA=zeros(1,NumConc); %Will hold total
concentration [# /liter]

WaitHandle = waitbar(0, 'Calculating performance data');
for i=1:NumConc
    Aerosol.Conc(BGCols)=Cnrm.*Cmag(i); %Total concentration of each
particle type

    [Ni,NiStatus]=FindAvgNumAnalyzed(Aerosol,Instrument,BGCols); %Avg # of each
background particle type analyzed per sample period

    if NiStatus==0 %if something has gone wrong
with FindAvgNumAnalyzed, then stop
        disp('Warning: Exceeded range of tracking data! (CalcACPLAData)')
        ACPLA(i:end)=[];
        TotPLA(i:end)=[];
        close(WaitHandle);
        return;
    end

    [k,Pgreaterk] = FindNumToAlarm(Pfa,Aerosol.Pagnt(BGCols),Ni); %k is the number
of agent ID's required to sound an alarm

    % disp([num2str(k),' agent IDs required for alarm when avg of ',num2str(sum(Ni)),'
bkgnd parts analyzed/min. Tot BkGnd Conc= ',num2str(sum(Aerosol.Conc(2:end)))]);

    [MinAvgNumAgentIDs,ProbError] = fzero(@FindMinAvgNumAgentIDs,[0 4*k],[],k,Pd); %min
avg num of particles (both agent and background) ID'd as agent to exceed k with prob of
Pd

    ACPLA(i)=FindMinConcAgent(Aerosol,Instrument,MinAvgNumAgentIDs); %The concentration
of agent required to obtain MinAvgNumAgentIDs (with help from background)

    if ACPLA(i)==-1 %if something has gone wrong with FindMindConcAgent, then stop
        ACPLA(i:end)=[];
        TotPLA(i:end)=[];
        close(WaitHandle);
        return;
    end

    TotPLA(i)=ACPLA(i)+sum(Aerosol.Conc(BGCols)); %Total concentration of
stuff in the air

    % disp([num2str(ACPLA(i)),' agents/liter required for alarm in tot. particle conc of
',num2str(TotPLA(i)),' particles/liter.']);

    WaitBar((i/NumConc), WaitHandle);
end
close(WaitHandle);

```

---

```

function ProbDif=FindMinAvgNumAgentIDs(N,k,Pd);
%N average number of agent IDs
%k number of agent IDs to sound alarm

```

```

%Pd          required prob that the actual number of agent IDs >=k
%ProbDif     difference between actual Pd and required Pd (we want this to be zero)

PoisProb(1)=exp(-N);          %Pois prob that exactly 0 particles ID'd
as agent
for i=1:k-1
    PoisProb(i+1)=PoisProb(i)*N/i;    %prob that exactly i particles ID'd as
agents
end
Pgreaterk=1-sum(PoisProb);      %prob that >=k ID'd as agents

ProbDif=Pd-Pgreaterk;          %The difference between the desired
probability of detection and the probability of detection for the given N and k

```

---

```

function ACPLA=FindMinConcAgent(Aerosol,Instrument,N)
%Aerosol      aerosol parameters
%Instrument    instrument parameters
%N            avg number of agent ID's required (from both agent and background)

%ACPLA        concentration of agent required to produce N (with help from background)
given aerosol and instrument

%-----make an estimate of the overall instrument efficiency and guess at required agent
concentration
AllCols=1:length(Aerosol.Conc);    %The columns for every particle
type (agent and background)

Aerosol.Conc(1)=1000;              %set the agent concentration to
1000 [#./liter]

Nall=FindAvgNumAnalyzed(Aerosol,Instrument,AllCols); %average number of each type
analyzed per minute

Nall=Nall.*Aerosol.Pagnt;          %average number ID'd as agent for
each type

Eff=Aerosol.Conc(1)/Nall(1);       %crude overall instrument
efficiency for agent

Nbg=sum(Nall(2:end));              %crude number of ID's expected
from background (bg)

ACPLAest=Eff*( N-Nbg );           %our guess at the required
concentration of agent

%-----now refine the estimate to get the real result
if DifNumAnal(0,Aerosol,Instrument,N)*DifNumAnal(4*ACPLAest,Aerosol,Instrument,N)>1
    ACPLA=-1;
    disp('Warning: The system is so saturated that it cant suck in enough agent... or
something like that! (FindMinConcAgent)');
    return
end

[ACPLA,NDif] = fzero(@DifNumAnal,[0 4*ACPLAest],[],Aerosol,Instrument,N);

Aerosol.Conc(1)=ACPLA;
[Nall,NallStatus]=FindAvgNumAnalyzed(Aerosol,Instrument,1);
if NallStatus==0
    ACPLA=-1;
    disp('Warning: Exceeded range of tracking data! (FindMinConcAgent)');
    return
end

% disp(['Estimated ACPLA: ',num2str(ACPLAest),' Actual ACPLA: ',num2str(ACPLA)])

```

```

if abs(NDif/N)>0.05
    disp('Warning: The derived ACPLA may be inaccurate! (FindMinConcAgent)')
end

```

---

```

function NDif=DifNumAnal(ACPLAguess,Aerosol,Instrument,N)
%ACPLAguess    estimate for the agent concentration
%Aerosol       aerosol parameters
%Instrument     instrument parameters
%N             avg number of agent ID's required

%NDif          difference between actual and desired avg # of agent particles analyzed

Aerosol.Conc(1)=ACPLAguess;

AllCols=1:length(Aerosol.Conc);           %The columns for every particle
type (agent and background)

Nall=FindAvgNumAnalyzed(Aerosol,Instrument,AllCols); %avg # of each type analyzed per
minute
Nall=Nall.*Aerosol.Pagnt;                 %avg # ID'd as agent

NDif=sum(Nall)-N;                         %difference between the number
actually ID'd and the number that should be analyzed ID'd

```

---

## A.8 Poisson Simplification of Eq. 23 for Two Particle Types

Eq. 23 describes the probability of false alarm based on an arbitrary number of particle types. In particular

$$P_{fa} \geq 1 - \sum_{k_1 + \dots + k_m < k} \prod_{i=1}^m p_{N_i}(k_i).$$

Consider the case for two particle types ( $i=1,2$ ).

$$P_{fa} \geq 1 - \sum_{k_1 + k_2 < k} p_{N_1}(k_1) p_{N_2}(k_2).$$

Write out the summation more explicitly.

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} \sum_{k_2=0}^j p_{N_1}(j-k_2) p_{N_2}(k_2).$$

Insert the definition of the Poisson distribution for both  $p_N$  terms.

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} \sum_{k_2=0}^j e^{-N_1} \frac{N_1^{j-k_2}}{(j-k_2)!} e^{-N_2} \frac{N_2^{k_2}}{k_2!}.$$

Rearrange.

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} e^{-(N_1+N_2)} \sum_{k_2=0}^j \frac{1}{(j-k_2)! k_2!} N_1^{j-k_2} N_2^{k_2}.$$

Multiply by 1 (i.e.  $j!/j!$ ).

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} \frac{e^{-(N_1+N_2)}}{j!} \sum_{k_2=0}^j \frac{j!}{(j-k_2)! k_2!} N_1^{j-k_2} N_2^{k_2}.$$

Evaluate the summation over  $k_2$ .

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} \frac{e^{-(N_1+N_2)}}{j!} (N_1 + N_2)^j.$$

Rearrange slightly.

$$P_{fa} \geq 1 - \sum_{j=0}^{k-1} e^{-(N_1+N_2)} \frac{(N_1+N_2)^j}{j!}.$$

This is the desired result (Eq. 24). It can easily be expanded to more than two particle types (although the initial summation will become more complex). This simplification ultimately works because the sum of two Poisson distributions is another Poisson distribution with a mean equal to the sum of the means.

## References

- (1) Perry, R. D.; Fetherston, J. D. *Clin. Microbiol. Rev.* **1997**, *10*, 35-&.
- (2) Inglesby, T. V.; Dennis, D. T.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Friedlander, A. M.; Hauer, J.; Koerner, J. F.; Layton, M.; McDade, J.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Schoch-Spana, M.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **2000**, *283*, 2281-2290.
- (3) Stolter, D. *The Coming of the Plague to Italy*; [www.brown.edu/Departments/Italian\\_Studies/dweb/plague/origins/spread.shtml](http://www.brown.edu/Departments/Italian_Studies/dweb/plague/origins/spread.shtml), 2002.
- (4) Macaulay, T. B. *The history of England from the accession of James II*; E. H. Butler & Co.: Philadelphia, 1856.
- (5) Henderson, D. A.; Inglesby, T. V.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Jahrling, P. B.; Hauer, J.; Layton, M.; McDade, J.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T.; Russell, P. K.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **1999**, *281*, 2127-2137.
- (6) Gani, R.; Leach, S. *Nature* **2001**, *414*, 748-751.
- (7) Fenner, F.; Henderson, D. A.; Arita, I.; Jezek, Z.; Ladnyi, I. D. *Smallpox and its eradication*; World Health Organization: Geneva, Switzerland, 1988.
- (8) Alibek, K. *Biohazard*; Arrow: London, 2000.
- (9) Ascenzi, P.; Visca, P.; Ippolito, G.; Spallarossa, A.; Bolognesi, M.; Montecucco, C. *FEBS Lett.* **2002**, *531*, 384-388.
- (10) Brossier, F.; Mock, M. *Toxicon* **2001**, *39*, 1747-1755.
- (11) Mock, M.; Mignot, T. *Cell Microbiol.* **2003**, *5*, 15-23.
- (12) Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Friedlander, A. M.; Hauer, J.; McDade, J.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **1999**, *281*, 1735-1745.
- (13) Inglesby, T. V.; O'Toole, T.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Friedlander, A. M.; Gerberding, J.; Hauer, J.; Hughes, J.; McDade, J.; Osterholm, M. T.; Parker, G.; Perl, T. M.; Russell, P. K.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **2002**, *287*, 2236-2252.
- (14) Dull, P. M.; Wilson, K. E.; Kournikakis, B.; Whitney, E. A. S.; Boulet, C. A.; Ho, J. Y. W.; Ogston, J.; Spence, M. R.; McKenzie, M. M.; Phelan, M. A.; Popovic, T.; Ashford, D. *Emerg. Infect. Dis* **2002**, *8*, 1044-1047.
- (15) Matsumoto, G. *Science* **2003**, *302*, 1492-+.
- (16) U.S. Congress Office of Technology Assessment *Proliferation of Weapons of Mass Destruction: Assessing the Risks*; US Government Printing Office: Washington, DC, 1993.
- (17) Kaufmann, A. F.; Meltzer, M. I.; Schmid, G. P. *Emerg. Infect. Dis* **1997**, *3*, 83-94.
- (18) Erbguth, F. J. *Movement Disorders* **2004**, *19*, S2-S6.
- (19) Arnon, S. S.; Schechter, R.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Hauer, J.; Layton, M.; Lillibridge, S.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Swerdlow, D. L.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **2001**, *285*, 1059-1070.

- (20) Atlas, R. M. *Annu. Rev. Microbiol.* **2002**, *56*, 167-185.
- (21) National Research Council. Subcommittee on Airborne Particles *Airborne Particles*; University Park Press: Baltimore, Maryland, 1979.
- (22) Cox, C. S.; Wathes, C. M., Eds. *Bioaerosols Handbook*; Lewis Publishers: Boca Raton, New York, 1995.
- (23) Spurny, K. R., Ed. *Analytical chemistry of aerosols*; Lewis Publishers: Boca Raton, 1999.
- (24) Willeke, K.; Baron, P. A., Eds. *Aerosol measurement: principles, techniques, and applications*; Van Nostrand Reinhold: New York, 1993.
- (25) Burge, H. A., Ed. *Bioaerosols*; Lewis Publishers: Boca Raton, 1995.
- (26) Johnston, M. V. *J. Mass Spectrom.* **2000**, *35*, 585-595.
- (27) Cunha, B. A. *Clin. Microbiol. Infect.* **2002**, *8*, 489-503.
- (28) Jane's Information Group *Chemical-Biological Defense Guidebook*; Jane's Information Group, 2000.
- (29) Taborsky, V. *FAO Agricultural Services Bulletin* **1992**.
- (30) Zimlich, W. C.; Ding, J. Y.; Busick, D. R.; Moutvic, R. R.; Placke, M. E.; Hirst, P. H.; Pitcairn, G. R.; Malik, S.; Newman, S. P.; Macintyre, F.; Miller, P. R.; Shepherd, M. T.; Lukas, T. M. *Respiratory Drug Delivery VII* **2000**, 241-246.
- (31) Ludvigson, L.; Dover, N.; Horn, J.; Pitesky, M.; Tobias, H.; Ferguson, D.; Resnekov, O.; Woods, B.; Gard, E. In *Detection of Distinct Biomarkers During Bacterial Spore Development Using Bio Aerosol Mass Spectrometry (BAMS)*; American Society of Microbiology Annual Meeting: New Orleans, LA, 2004.
- (32) Errington, J. *Microbiol. Rev.* **1993**, *57*, 1-33.
- (33) Holler, S.; Pan, Y. L.; Chang, R. K.; Bottiger, J. R.; Hill, S. C.; Hillis, D. B. *Opt. Lett.* **1998**, *23*, 1489-1491.
- (34) Westphal, A. J.; Price, P. B.; Leighton, T. J.; Wheeler, K. E. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 3461-3466.
- (35) Driks, A. *Trends Microbiol.* **2002**, *10*, 251-254.
- (36) Driks, A. *Microbiol. Mol. Biol. Rev.* **1999**, *63*, 1-+.
- (37) Madigan, M. T.; Martinko, J. M.; Parker, J. *Brock Biology of Microorganisms*, 10 ed.; Pearson Education, Inc.: Upper Saddle River, NJ, 2003.
- (38) Hathout, Y.; Setlow, B.; Cabrera-Martinez, R. M.; Fenselau, C.; Setlow, P. *Appl. Environ. Microbiol.* **2003**, *69*, 1100-1107.
- (39) Mock, M.; Fouet, A. *Annu. Rev. Microbiol.* **2001**, *55*, 647-671.
- (40) Peters, C. J.; Hartley, D. M. *Lancet* **2002**, *359*, 710-711.
- (41) U.K. Health Protection Agency *Deliberate Release*; [http://www.hpa.org.uk/infections/topics\\_az/deliberate\\_release/menu.htm](http://www.hpa.org.uk/infections/topics_az/deliberate_release/menu.htm), 2003.
- (42) Dennis, D. T.; Inglesby, T. V.; Henderson, D. A.; Bartlett, J. G.; Ascher, M. S.; Eitzen, E.; Fine, A. D.; Friedlander, A. M.; Hauer, J.; Layton, M.; Lillibridge, S. R.; McDade, J. E.; Osterholm, M. T.; O'Toole, T.; Parker, G.; Perl, T. M.; Russell, P. K.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **2001**, *285*, 2763-2773.
- (43) Borio, L.; Inglesby, T.; Peters, C. J.; Schmaljohn, A. L.; Hughes, J. M.; Jahrling, P. B.; Ksiazek, T.; Johnson, K. M.; Meyerhoff, A.; O'Toole, T.; Ascher, M. S.; Bartlett, J.; Breman, J. G.; Eitzen, E. M.; Hamburg, M.; Hauer, J.; Henderson, A.; Johnson, R. T.; Kwik, G.; Layton, M.; Lillibridge, S.; Nabel, G. J.; Osterholm, M.

- T.; Perl, T. M.; Russell, P.; Tonat, K. *JAMA-J. Am. Med. Assoc.* **2002**, *287*, 2391-2405.
- (44) Hill, S. C.; Pinnick, R. G.; Nachman, P.; Chen, G.; Chang, R. K.; Mayo, M. W.; Fernandez, G. L. *Appl. Optics* **1995**, *34*, 7149-7155.
- (45) Hill, S. C.; Pinnick, R. G.; Niles, S.; Pan, Y. L.; Holler, S.; Chang, R. K.; Bottiger, J.; Chen, B. T.; Orr, C. S.; Feather, G. *Field Anal. Chem. Technol.* **1999**, *3*, 221-239.
- (46) Reyes, F. L.; Jeys, T. H.; Newbury, N. R.; Primmerman, C. A.; Rowe, G. S.; Sanchez, A. *Field Anal. Chem. Technol.* **1999**, *3*, 240-248.
- (47) Eversole, J. D.; Hardgrove, J. J.; Cary, W. K.; Choulas, D. P.; Seaver, M. *Field Anal. Chem. Technol.* **1999**, *3*, 249-259.
- (48) Chen, G.; Nachman, P.; Pinnick, R. G.; Hill, S. C.; Chang, R. K. *Opt. Lett.* **1996**, *21*, 1307-1309.
- (49) Nachman, P.; Chen, G.; Pinnick, R. G.; Hill, S. C.; Chang, R. K.; Mayo, M. W.; Fernandez, G. L. *Appl. Optics* **1996**, *35*, 1069-1076.
- (50) Kaye, P. H.; Barton, J. E.; Hirst, E.; Clark, J. M. *Appl. Optics* **2000**, *39*, 3738-3745.
- (51) Hairston, P. P.; Ho, J.; Quant, F. R. *J. Aerosol. Sci.* **1997**, *28*, 471-482.
- (52) Agranovski, V.; Ristovski, Z.; Hargreaves, M.; Blackall, P. J.; Morawska, L. *J. Aerosol. Sci.* **2003**, *34*, 301-317.
- (53) Agranovski, V.; Ristovski, Z.; Hargreaves, M.; Blackall, P. J.; Morawska, L. *J. Aerosol. Sci.* **2003**, *34*, 1711-1727.
- (54) Pan, Y. L.; Holler, S.; Chang, R. K.; Hill, S. C.; Pinnick, R. G.; Niles, S.; Bottiger, J. R. *Opt. Lett.* **1999**, *24*, 116-118.
- (55) Delehanty, J. B.; Ligler, F. S. *Anal. Chem.* **2002**, *74*, 5681-5687.
- (56) Cockerill, F. R.; Smith, T. F. *Asm News* **2002**, *68*, 77-83.
- (57) Fitch, J. P.; Gardner, S. N.; Kuczmariski, T. A.; Kurtz, S.; Myers, R.; Ott, L. L.; Slezak, T. R.; Vitalis, E. A.; Zemla, A. T.; McCready, P. M. *Proceedings of the Ieee* **2002**, *90*, 1708-1721.
- (58) Fitch, J. P.; Raber, E.; Imbro, D. R. *Science* **2003**, *302*, 1350-1354.
- (59) McBride, M. T.; Masquelier, D.; Hindson, B. J.; Makarewicz, A. J.; Brown, S.; Burris, K.; Metz, T.; Langlois, R. G.; Tsang, K. W.; Bryan, R.; Anderson, D. A.; Venkateswaran, K. S.; Milanovich, F. P.; Colston, B. W. *Anal. Chem.* **2003**, *75*, 5293-5299.
- (60) Basile, F.; Beverly, M. B.; Voorhees, K. J.; Hadfield, T. L. *Trac-Trends Anal. Chem.* **1998**, *17*, 95-109.
- (61) Basile, F.; Beverly, M. B.; Abbas-Hawks, C.; Mowry, C. D.; Voorhees, K. J.; Hadfield, T. L. *Anal. Chem.* **1998**, *70*, 1555-1562.
- (62) Griest, W. H.; Wise, M. B.; Hart, K. J.; Lammert, S. A.; Thompson, C. V.; Vass, A. A. *Field Anal. Chem. Technol.* **2001**, *5*, 177-184.
- (63) Barshick, S. A.; Wolf, D. A.; Vass, A. A. *Anal. Chem.* **1999**, *71*, 633-641.
- (64) Fenselau, C.; Demirev, P. A. *Mass Spectrom. Rev.* **2001**, *20*, 157-171.
- (65) Van Baar, B. L. M. *Fems Microbiol. Rev.* **2000**, *24*, 193-219.
- (66) Lay, J. O. *Mass Spectrom. Rev.* **2001**, *20*, 172-194.
- (67) Xu, M.; Basile, F.; Voorhees, K. J. *Anal. Chim. Acta* **2000**, *418*, 119-128.



- (68) Anderson, C. W.; Carlson, M. A. *Johns Hopkins APL Tech. Dig.* **1999**, *20*, 352-362.
- (69) Scholl, P. F.; Leonardo, M. A.; Rule, A. M.; Carlson, M. A.; Antoine, M. D.; Buckley, T. J. *Johns Hopkins APL Tech. Dig.* **1999**, *20*, 343-351.
- (70) McLoughlin, M. P.; Allmon, W. R.; Anderson, C. W.; Carlson, M. A.; DeCicco, D. J.; Evancich, N. H. *Johns Hopkins APL Tech. Dig.* **1999**, *20*, 326-334.
- (71) Cornish, T. J.; Bryden, W. A. *Johns Hopkins APL Tech. Dig.* **1999**, *20*, 335-342.
- (72) Suess, D. T.; Prather, K. A. *Chem. Rev.* **1999**, *99*, 3007-3035.
- (73) Sinha, M. P.; Platz, R. M.; Vilker, V. L.; Friedlander, S. K. *Int. J. Mass Spectrom. Ion Process.* **1984**, *57*, 125-133.
- (74) Sinha, M. P.; Platz, R. M.; Friedlander, S. K.; Vilker, V. L. *Appl. Environ. Microbiol.* **1985**, *49*, 1366-1373.
- (75) Gieray, R. A.; Reilly, P. T. A.; Yang, M.; Whitten, W. B.; Ramsey, J. M. *J. Microbiol. Methods* **1997**, *29*, 191-199.
- (76) Parker, E. P.; Trahan, M. W.; Wagner, J. S.; Rosenthal, S. E.; Whitten, W. B.; Gieray, R. A.; Reilly, P. T. A.; Lazar, A. C.; Ramsey, J. M. *Field Anal. Chem. Technol.* **2000**, *4*, 31-42.
- (77) Johnston, M. V.; Wexler, A. S. *Anal. Chem.* **1995**, *67*, A721-A726.
- (78) Murphy, D. M.; Thomson, D. S. *Aerosol Sci. Technol.* **1995**, *22*, 237-249.
- (79) Hinz, K. P.; Kaufmann, R.; Spengler, B. *Anal. Chem.* **1994**, *66*, 2071-2076.
- (80) Hinz, K. P.; Kaufmann, R.; Spengler, B. *Aerosol Sci. Technol.* **1996**, *24*, 233-242.
- (81) Lazar, A.; Reilly, P. T. A.; Whitten, W. B.; Ramsey, J. M. *Environ. Sci. Technol.* **1999**, *33*, 3993-4001.
- (82) Gard, E.; Mayer, J. E.; Morrical, B. D.; Dienes, T.; Fergenson, D. P.; Prather, K. A. *Anal. Chem.* **1997**, *69*, 4083-4091.
- (83) Mallina, R. V.; Wexler, A. S.; Johnston, M. V. *J. Aerosol. Sci.* **1999**, *30*, 719-738.
- (84) Salt, K.; Noble, C. A.; Prather, K. A. *Anal. Chem.* **1996**, *68*, 230-234.
- (85) He, L.; Murray, K. K. *J. Mass Spectrom.* **1999**, *34*, 909-914.
- (86) Jackson, S. N.; Murray, K. K. *Anal. Chem.* **2002**, *74*, 4841-4844.
- (87) Stowers, M. A.; van Wuijckhuijse, A. L.; Marijnissen, J. C. M.; Scarlett, B.; van Baar, B. L. M.; Kientz, C. E. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 829-833.
- (88) Valegard, K.; Liljas, L.; Fridborg, K.; Unge, T. *Nature* **1990**, *345*, 36-41.
- (89) Friedli, G.-L., University of Surrey, 1996.
- (90) May, K. R. *J. Aerosol. Sci.* **1973**, *4*, 235-243.
- (91) Kinney, P. D.; Pui, D. Y. H.; Mulholland, G. W.; Bryner, N. P. *J. Res. Natl. Inst. Stand. Technol.* **1991**, *96*, 147-176.
- (92) Mainelis, G.; Willeke, K.; Baron, P.; Reponen, T.; Grinshpun, S. A.; Gorny, R. L.; Trakumas, S. *J. Aerosol. Sci.* **2001**, *32*, 1087-1110.
- (93) Terzieva, S.; Donnelly, J.; Ulevicius, V.; Grinshpun, S. A.; Willeke, K.; Stelma, G. N.; Brenner, K. P. *Appl. Environ. Microbiol.* **1996**, *62*, 2264-2272.
- (94) Reponen, T.; Willeke, K.; Ulevicius, V.; Grinshpun, S. A.; Donnelly, J. *Aerosol Sci. Technol.* **1997**, *27*, 405-421.
- (95) Ulevicius, V.; Willeke, K.; Grinshpun, S. A.; Donnelly, J.; Lin, X. J.; Mainelis, G. *Aerosol Sci. Technol.* **1997**, *26*, 175-190.
- (96) Noble, C. A.; Prather, K. A. *Environ. Sci. Technol.* **1996**, *30*, 2667-2680.

- (97) Karataev, V. I.; Mamyrin, B. A.; Shmikk, D. V. *Soviet Physics - Technical Papers* **1972**, *16*, 1177-1111-1179.
- (98) Mamyrin, B. A.; Karataev, V. I.; Shmikk, D. V.; Zagulin, V. A. *Soviet Physics - JETP* **1973**, *37*, 45-48.
- (99) Wiley, W. C.; McLaren, I. H. *The Review of Scientific Instruments* **1955**, *26*, 1150-1157.
- (100) Wiza, J. L. *Nuclear Instruments & Methods* **1979**, *162*, 587-601.
- (101) Nudelman, R.; Bronk, B. V.; Efrima, S. *Appl. Spectrosc.* **2000**, *54*, 445-449.
- (102) Marchand, P.; Marmet, L. *Rev. Sci. Instrum.* **1983**, *54*, 1034-1041.
- (103) Song, X. H.; Hopke, P. K.; Ferguson, D. P.; Prather, K. A. *Anal. Chem.* **1999**, *71*, 860-865.
- (104) Stein, S. E.; Scott, D. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 859-866.
- (105) Steele, P. T.; Tobias, H. J.; Ferguson, D. P.; Pitesky, M. E.; Horn, J. M.; Czerwieniec, G. A.; Russell, S. C.; Lebrilla, C. B.; Gard, E. E.; Frank, M. *Anal. Chem.* **2003**, *75*, 5480-5487.
- (106) Feldhaus, D.; Menzel, C.; Berkenkamp, S.; Hillenkamp, F.; Dreisewerd, K. *J. Mass Spectrom.* **2000**, *35*, 1320-1328.
- (107) Berkenkamp, S.; Menzel, C.; Hillenkamp, F.; Dreisewerd, K. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 209-220.
- (108) Westmacott, G.; Ens, W.; Hillenkamp, F.; Dreisewerd, K.; Schurenberg, M. *Int. J. Mass Spectrom.* **2002**, *221*, 67-81.
- (109) Noble, C. A.; Prather, K. A. *Aerosol Sci. Technol.* **1998**, *29*, 294-306.
- (110) Wenzel, R. J.; Prather, K. A. *Rapid Commun. Mass Spectrom.* **2004**, *18*, 1525-1533.
- (111) Ferguson, D. P.; Pitesky, M. E.; Tobias, H. J.; Steele, P. T.; Czerwieniec, G. A.; Russell, S. C.; Lebrilla, C. B.; Horn, J. M.; Coffee, K. R.; Srivastava, A.; Pillai, S. P.; Shih, M. T. P.; Hall, H. L.; Ramponi, A. J.; Chang, J. T.; Langlois, R. G.; Estacio, P. L.; Hadley, R. T.; Frank, M.; Gard, E. E. *Anal. Chem.* **2004**, *76*, 373-378.
- (112) Eversole, J. D.; Cary, W. K.; Scotto, C. S.; Pierson, R.; Spence, M.; Campillo, A. *J. Field Anal. Chem. Technol.* **2001**, *5*, 205-212.
- (113) Hill, S. C.; Pinnick, R. G.; Niles, S.; Fell, N. F.; Pan, Y. L.; Bottiger, J.; Bronk, B. V.; Holler, S.; Chang, R. K. *Appl. Optics* **2001**, *40*, 3005-3013.
- (114) Auerbach, J. M.; Karpenko, V. P. *Appl. Optics* **1994**, *33*, 3179-3183.
- (115) Hoffnagle, J. A.; Jefferson, C. M. *Appl. Optics* **2000**, *39*, 5488-5499.
- (116) Ih, C. S. *Appl. Optics* **1972**, *11*, 694-695.
- (117) Kanzler, K. *Proceedings of SPIE*, San Diego, CA 2001.
- (118) Rohlfing, A.; Menzel, C.; Kukreja, L. M.; Hillenkamp, F.; Dreisewerd, K. *Journal of Physical Chemistry B* **2003**, *107*, 12275-12286.
- (119) Dreisewerd, K.; Schurenberg, M.; Karas, M.; Hillenkamp, F. *Int. J. Mass Spectrom. Ion Process.* **1995**, *141*, 127-148.
- (120) Dean, D. H.; Douthit, H. A. *J. Bacteriol.* **1974**, *117*, 601-610.
- (121) Tuminello, P. S.; Arakawa, E. T.; Khare, B. N.; Wrobel, J. M.; Query, M. R.; Milham, M. E. *Appl. Optics* **1997**, *36*, 2818-2824.
- (122) Mätzler, C. *MATLAB Functions for Mie Scattering and Absorption*, 2 ed.; University of Bern: Bern, Switzerland, 2002.

- (123) Gurton, K. P.; Ligon, D.; Kvavilashvili, R. *Appl. Optics* **2001**, *40*, 4443-4448.
- (124) Knochenmuss, R. *J. Mass Spectrom.* **2002**, *37*, 867-877.
- (125) Campbell, S.; Beauchamp, J. L.; Rempe, M.; Lichtenberger, D. L. *Int. J. Mass Spectrom. Ion Process.* **1992**, *117*, 83-99.
- (126) Knochenmuss, R.; Zenobi, R. *Chem. Rev.* **2003**, *103*, 441-452.
- (127) Becker, C. H.; Wu, K. J. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 883-888.
- (128) Lockyer, N. P.; Vickerman, J. C. *Int. J. Mass Spectrom.* **2000**, *197*, 197-209.
- (129) Lockyer, N. P.; Vickerman, J. C. *Int. J. Mass Spectrom.* **1998**, *176*, 77-86.
- (130) Grun, C.; Heinicke, R.; Weickhardt, C.; Grottemeyer, J. *Int. J. Mass Spectrom.* **1999**, *187*, 307-318.
- (131) Aicher, K. P.; Wilhelm, U.; Grottemeyer, J. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1059-1068.
- (132) Weinkauff, R.; Aicher, P.; Wesley, G.; Grottemeyer, J.; Schlag, E. W. *J. Phys. Chem.* **1994**, *98*, 8381-8391.
- (133) Cabalo, J.; Zelenyuk, A.; Baer, T.; Miller, R. E. *Aerosol Sci. Technol.* **2000**, *33*, 3-19.
- (134) Morrical, B. D.; Fergenson, D. P.; Prather, K. A. *J. Am. Soc. Mass Spectrom.* **1998**, *9*, 1068-1073.
- (135) Zelenyuk, A.; Cabalo, J.; Baer, T.; Miller, R. E. *Anal. Chem.* **1999**, *71*, 1802-1808.
- (136) Woods, E.; Smith, G. D.; Dessiaterik, Y.; Baer, T.; Miller, R. E. *Anal. Chem.* **2001**, *73*, 2317-2322.
- (137) Woods, E.; Smith, G. D.; Miller, R. E.; Baer, T. *Anal. Chem.* **2002**, *74*, 1642-1649.
- (138) Russell, S. C.; Czerwieniec, G.; Lebrilla, C.; Tobias, H.; Fergenson, D. P.; Steele, P.; Pitesky, M.; Horn, J.; Srivastava, A.; Frank, M.; Gard, E. E. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 900-909.
- (139) Ullom, J. N.; Frank, M.; Gard, E. E.; Horn, J. M.; Labov, S. E.; Langry, K.; Magnotta, F.; Stanion, K. A.; Hack, C. A.; Benner, W. H. *Anal. Chem.* **2001**, *73*, 2331-2337.
- (140) Dreisewerd, K.; Berkenkamp, S.; Leisner, A.; Rohlfing, A.; Menzel, C. *Int. J. Mass Spectrom.* **2003**, *226*, 189-209.
- (141) Gluckmann, M.; Karas, M. *J. Mass Spectrom.* **1999**, *34*, 467-477.
- (142) Spengler, B.; Kirsch, D. *Int. J. Mass Spectrom.* **2003**, *226*, 71-83.
- (143) Karas, M.; Bahr, U.; Fournier, I.; Gluckmann, M.; Pfenninger, A. *Int. J. Mass Spectrom.* **2003**, *226*, 239-248.
- (144) Zhigilei, L. V.; Yingling, Y. G.; Itina, T. E.; Schoolcraft, T. A.; Garrison, B. J. *Int. J. Mass Spectrom.* **2003**, *226*, 85-106.
- (145) Schoolcraft, T. A.; Constable, G. S.; Jackson, B.; Zhigilei, L. V.; Garrison, B. J. *Nucl. Instrum. Methods Phys. Res. Sect. B-Beam Interact. Mater. Atoms* **2001**, *180*, 245-250.
- (146) Schoolcraft, T. A.; Constable, G. S.; Zhigilei, L. V.; Garrison, B. J. *Anal. Chem.* **2000**, *72*, 5143-5150.