The Effect of Faults on Plasma Particle Detector Data Reduction

Michael L. Rilee, Scott A. Boardsen, and Maharaj K. Bhat
Emergent ITI, NASA Goddard Space Flight Center, Mailstop 931, Greenbelt, MD 20770
Phone: (301) 286-4743, email: Michael.L.Rilee.1@gsfc.nasa.gov

Steven A. Curtis
NASA Goddard Space Flight Center, Code 695, Greenbelt, MD 20770
Phone: (301) 286-9188, email: Steven.A.Curtis.1@gsfc.nasa.gov

Abstract—Missions in NASA's Solar-Terrestrial Probe line feature challenges such as multiple spacecraft and high data production rates. An important class of scientific instruments that have for years strained against limits on communications are the particle detectors used to measure space plasma density, temperature, and flow. The Plasma Moment Application (PMA) software is being developed for the NASA Remote Exploration & Experimentation (REE) Program's series of Flight Processor Testbeds. REE seeks to enable instrument science teams to move data analyses such as PMA on board the spacecraft thereby reducing communication downlink requirements. Here we describe the PMA for the first time and examine its behavior under single bit faults in its static state. We find that ~90% of the faults lead to tolerable behavior, while the remainder cause either program failure or nonsensical results. These results help guide the development of fault tolerant, non-hardened flight/science processors.

TABLE OF CONTENTS
1. INTRODUCTION
2. PLASMA MOMENT APPLICATION
3. TESTING METHOD
4. RESULTS
5. CONCLUSIONS

FUTURE missions in NASA's Solar-Terrestrial Probe (STP) line have challenging requirements involving multi spacecraft operations and science data communication. These missions form an important part of NASA Sun-Earth Connections theme [1,2]. One such mission is STP's Magnetospheric Multi Scale (MMS, launch 2007), which consists of five spacecraft each carrying a full complement of science instruments designed to observe the fundamental processes that structure Geospace, the near Earth space environment [3]. To fit within mission resource budgets, MMS science return is purposefully limited by obtaining data at a high rate during "burst mode" operations during limited portions of MMS orbits. For the rest of the orbit, "normal mode" stores low resolution data for later transmission to Earth. By performing some data analyses on board the spacecraft, the science return of missions such as MMS may be dramatically enhanced [4].

For example, the particle detectors used in space physics experiments measure properties, e.g. mass, momentum, and energy, of the particles making up the space environment. For many years these science instruments have been capable of producing more information than can reasonably be downlinked to Earth. Thus communication limitations have restricted the quality of the raw measurements which in turn limit the conclusions that we may draw from our observations. The STP component of NASA's Remote Exploration and Experimentation Project (REE) is developing methods of on board data analysis that will enable STP missions that are currently infeasible due to communication limitations [5].

Before science data can be used to achieve mission goals, important calibrations and reductions must be performed. Results of these analyses require storage that is often thousands of times smaller than the storage required of the original raw data. By performing these analyses on board spacecraft and transmitting the results to Earth one gains the opportunity to reduce science communication requirements by thousands as well. However, performing these analyses in real-time on board spacecraft requires computing capability that is not feasible if obtained by traditional techniques of constructing radiation tolerant systems. The REE Testbed is a prototype flight processor that uses a mixture of hardware and software techniques to mitigate the effects of radiation induced faults while providing supercomputing performance.

In this work, REE/STP presents a software application being developed for the REE Testbed that analyzes particle detector data. The Plasma Moment Application (PMA) calculates plasma (fluid) moments including the space plasma density, bulk flow, pressure, and heat flux. These moments are calculated using spatial discretization techniques reminiscent of finite-element analysis (FEA). These unstructured grid techniques allow the analysis of almost arbitrarily constructed or edited data sets. Model fitting and parameter estimation will be used to address inescapable "blind spots" and other imperfections in particle detector data. PMA represents a first step towards autonomous space science data analysis, and this paper is, to our knowledge, the first description of how a data analysis code developed within the space science community responds to faults.

U.S. Government work not protected by U.S. copyright.

5-2427
Spacecraft spin "z" axis

Instrument normal

Particle Path

Fiducial "x" axis

Instrument "x" axis

Figure 1. Particle detector geometry

2. PLASMA MOMENT APPLICATION

The Plasma Moment Application consists of a model particle spectrometer and an analysis program. The model particle spectrometer produces a set of sensor readings and related information; from these sensor readings, sensor configuration, and spacecraft attitude, a map of the plasma in velocity space is constructed. The analysis program converts the sensor readings to a plasma particle density per unit volume of phase space \((v_x \times \text{position})\). After this conversion, the analysis program fits an anisotropic temperature plasma distribution to the data; this fit is used to patch data gaps and as a check on subsequent analysis. The analysis program then constructs a grid from the distribution of sensor readings which it then uses to calculate statistical moments by numerical quadrature. Though motivated by existing plasma particle spectrometers, these models are implemented at a slightly more abstract level so that they may be specialized to a variety of detectors and measurement strategies.

PMA Spectrometer Model

Our initial spectrometer model is loosely based on the Electron Spectrometer Experiment (ESE) on ISEE-1 [6]. This device is designed to observe the three dimensional electron plasma distribution function as the spacecraft moves through Geospace including the solar wind, magnetosheath, outer magnetopause, and near tail regions. At the output of each of six electrostatic analyzers a two channel electron multiplier registers the influx of electrons from given differential energy ranges and directions. The ESE logarithmically steps through 16 energy levels, dwelling equally at each, and completes one energy sweep every half second. Each electron spectrometer is paired with another, and the pairs are arranged on three mutually orthogonal axes. A single detector is depicted in Figure 1. As the ISEE-1 spacecraft rotates every three seconds, the views of the spectrometers are swept across the sky and a three dimensional picture of the electron and their flow in the vicinity of the spacecraft is recorded every half second.

The PMA Spectrometer Model is not constrained to specifically represent the ISEE-1/ISEE. Many of the parameters that determine the data production rate of the ESE can be changed so that the requirements of other particle observation strategies can be studied. These parameters include: the number of energy steps, the central energy and range of each step, the step rate, and the number of detectors. The effective aperture of the instrument and the solid angle from which particles are admitted can also be specified. At the moment we presume ideal efficiency, in that each entering particle is counted. The spacecraft's, and hence the instrument's, rate of rotation may also be specified. These parameters control how particle velocity space is sampled.

To determine the flux of particles into the instrument, a small set of simple plasma models has been implemented. These models map the particle velocity space to a real valued phase space distribution function. The model we use here is a simple Gaussian, or normal distribution, centered on the mean drift of the plasma relative to the spacecraft, and with a width determined by the plasma temperature. The integration of this particle distribution over a sample region of velocity space yields the average number of particles that enter the detector, neglecting statistical fluctuation. To model the actual number of particles entering the detector we select a random number from a Poisson distribution with the aforementioned average as its mean. This process is the inverse of the data reduction which is described in more detail below.

The preceding two paragraphs have dealt with how velocity space is sampled by a model instrument and how the contents of velocity space are determined. Thus as the instrument proceeds through its observations it builds up a data set of measurements and relevant auxiliary data including time and duration of observation, and instrument configuration and position. These data and metadata are read into and processed by the analysis portion of the PMA application.

PMA Analysis Program

The instrument-modelling portion of the PMA reads a file that contains parameters which define a model spectrometer and plasma and produces a set of files that contain flux measurements and descriptive information. The analysis portion of the PMA reads these files with the objective of calculating the statistical moments of the observed plasma distribution in velocity space. The moments to be found correspond to the plasma density, bulk velocity, temperature, and heat flux. The steps required to transform the measurements into the desired fluid moments are outlined in Table 1.

Portions of the PMA program are not exercised for this work. The calibration step mentioned in Table 1 is usually iterative in nature, and not qualitatively different from the "patching step". The patching step is similarly truncated for the runs presented here. The model used for patching the data is constructed to exercise the linear algebra library; the model thus obtained was saved for later comparison to the results of the moment calculation. However, for this work the actual patch is not constructed.
TABLE 1
OUTLINE OF PMA PROGRAM

1. Input data
2. Transform to working coordinates & units
3. Calibrate
4. Patch holes in data set
5. Construct grid from data
6. Calculate moments
7. Write results

Constructing a patch requires a matrix-vector multiplication and the addition of a “few” modeled data points to the data set. Our experiences with this process will be documented elsewhere, but here the focus is the behavior of PMA under faults, and we feel that an additional matrix-vector multiply will not significantly affect our results.

Here, calibration means accounting for the different sensitivities of the various detectors. These different sensitivities are due to intrinsic variability in these devices as shipped, or differences in how they wear. These differences can be ameliorated by comparing how different detectors respond to a predictable environment, for example, by using the solar wind as a calibration beam. For these instruments the calibration, though important for the analysis of real data, looks like a filter, a model fit, or iterates of these. Because a model fit is required for the patching step, the addition of the calibration step will not add to the present study but will be addressed in future work. Therefore, our models presume ideal (cross) calibration.

Spacecraft often acquire an electric charge, which leads to a spacecraft electric potential \( \phi_{sc} \). Charged particles which enter the detector move through this potential, changing their energies. Therefore, in reconstructing the PDF from the sensor data, we must account for the shift \( \Delta E \) in particle energy \( E \):

\[
\Delta E = q \phi_{sc}.
\]

This energy shift is usually small and the velocities low, so nonrelativistic formulae may be used. Current code simply drops data with \( E < q\phi_{sc} \), and fits the patch to higher energy, \( E > q\phi_{sc} \), data.

Model Fitting

Model fitting is one method that has been used to address irregularities in plasma analyzer data. Particles with low velocities relative to the spacecraft are particularly difficult to measure for two reasons. First, a lower velocity implies a lower flux into the instrument which implies lower count rates and greater uncertainties and fluctuations in particle flux. Second, lower velocity particles are more susceptible to the electric potential of the spacecraft. A spacecraft placed in the Sun’s ultraviolet radiation becomes electrically charged and drives an electric current in the surrounding plasma that “corrupts” measurements of the properties of the space plasma. For these two reasons, measurements of the plasma particle distribution function at lower energies are difficult to interpret and often look like a “hole” in the data. A way around this problem is to “patch” the hole by fitting a model distribution function, e.g. a Maxwellian or Gaussian, to the measurements and then using this model to generate model data in the hole.

In this work, we use a ten parameter generalized Maxwellian (Eq. 2). The parameters are of the following form:

\[
\ln f \approx B_0 + B \cdot v + v^T \cdot B \cdot v,
\]

where \( B \) is an upper triangular \( 3 \times 3 \) matrix; the symbol \( ^T \) denotes transposition. Note that \( \ln f \) is a column vector of data, in the computer scientific sense, indexed by \( v \). Hence the model \( \ln f \) is linear in the parameters \( B = (B_0, B, B) \), Eq. 2 can be rewritten:

\[
\ln f = VB,
\]

where \( V \) is defined to make Eqs. 2 and 3 equivalent. Furthermore, \( V \) depends on the velocities \( v \) at which the data was obtained. We can write the \( j^{th} \) row of \( V \)

\[
\ln f_j = (1, v_j, v_j^T \cdot v_j) B,
\]

where \( j \) indexes the measurements \( f_j \). One can test the patch to \( f \) by checking for kinks or non-monotonicity.

Because the logarithm of the model distribution function is linear, we use a linear least squares fitting procedure implemented with PLAPACK [7] and LAPACK [8]. PLAPACK (Parallel Linear Algebra Package) is an object oriented library that provides dense linear algebra for multiple processor computers using the Message Passing Interface (MPI) [9]. Furthermore, experiments in Software Implemented Fault Tolerance (SIFT) are being performed with PLAPACK, with the plan that applications such as PMA will gain enhanced robustness to faults simply by using a SIFT based version of PLAPACK. After the model parameters have been found, they can be used to patch irregularities as mentioned above, or they can be compared with the results of the plasma moment calculation to be discussed below.

Plasma Moments

The primary function of the PMA is the plasma moment calculation, which proceeds after calibration and patching have resulted in a consistent, complete data set. To allow a great variety of measurement strategies to be studied, as well as to allow easy import of existing particle spectrometry data, the PMA uses a very general method to partition velocity space using techniques from finite element analysis [10].

The fundamental quantities of interest are the plasma or fluid moments of the distribution of particles in the space environment. The density, bulk flow velocity, and the pressure tensor are

\[
n = \int d^3v f(v),
\]
TABLE 2
QUANTITIES ASSOCIATED WITH PDF ESTIMATION.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Interpretation</th>
<th>Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>t</td>
<td>Time</td>
<td></td>
</tr>
<tr>
<td>(\phi, \theta)</td>
<td>Azimuth and co-altitude</td>
<td>t</td>
</tr>
<tr>
<td>n</td>
<td>Detector unit normal</td>
<td>(\theta, \phi, t)</td>
</tr>
<tr>
<td>(v)</td>
<td>Particle velocity</td>
<td>(\theta, \phi, t)</td>
</tr>
<tr>
<td>(v)</td>
<td>Velocity magnitude</td>
<td>(v =</td>
</tr>
<tr>
<td>C</td>
<td>Measured particle counts</td>
<td>(v, \theta, \phi)</td>
</tr>
<tr>
<td>(f)</td>
<td>Distribution function</td>
<td>(C, v, \theta, \phi)</td>
</tr>
</tbody>
</table>

TABLE 3
THE GEOMETRIC FACTOR \(\eta\) CONNECTING THE PDF AND COUNTS.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Interpretation</th>
<th>Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Detector collecting area</td>
<td>t</td>
</tr>
<tr>
<td>(\delta\Omega)</td>
<td>Solid angle subtended</td>
<td>(\delta\Omega = \sin \theta \delta\theta \delta\phi, E)</td>
</tr>
<tr>
<td>(\delta t)</td>
<td>Measurement duration</td>
<td>t</td>
</tr>
<tr>
<td>(E^{-1}\delta E)</td>
<td>Energy bandpass</td>
<td>t</td>
</tr>
<tr>
<td>(\eta)</td>
<td>Integration factor</td>
<td>(\eta = A \delta\Omega \delta t (2E)^{-1}\delta E)</td>
</tr>
</tbody>
</table>

\[
u = \frac{1}{n} \int d^3v f(v) v, \text{ and} \]
\[
P = m \int d^3v f(v) (v - u)(v - u), \]

where we interpret the dyadic product in Einstein's notation thusly: \(ab = aibj\). The integrals are evaluated over the particle velocity space, and \(m\) is the mass of the particle, e.g. electron, proton, ionized Oxygen, being studied. The tensor product is defined by: \(P_{ij} v_j\).

The heat flux tensor is given by:

\[
Q = \frac{m}{2} \int d^3v f(v) v^2 v - \left[\frac{1}{2} u TrP + P \cdot u + \frac{m n}{2} u u^2\right],
\]

where \(u = |u|\). Finally, in practice it is not always necessary to evaluate \(n, u, P, Q\), as written above. For some purposes, integrals of the dyadic powers of \(v\) are adequate, and from these the pressure and heat flux can be calculated.

Another issue involving grid construction, is that GEOMPACK forms a convex-hull of the data points. The convex hull is the minimal convex surface that encloses all of the points. For samplings of velocity space that are essentially non-convex, this can lead to grids (discretizations) that have copious, ill-shaped tetrahedra as mentioned above. The way around this problem that we have pursued is based on the recognition that in spherical coordinates the regions of velocity space samples are essentially non-convex. The resulting grid is a set of tetrahedral volumes that partition velocity space; the vertices of these tetrahedra correspond to velocity space measurements. From there it is a simple matter to evaluate functions defined on the measurements, including the calculation of moments of the plasma particle distribution by numerical quadrature. Integrations over velocity space become summations over the set of tetrahedra.

An advantage of this approach is that nearly arbitrary distributions of measurements may be handled. A key disadvantage is one that also arises in FEA, namely the errors that can be introduced by poorly shaped tetrahedra. Poorly shaped tetrahedra lead to difficulties both in grid construction and numerical errors during calculation. Nevertheless, we feel that this FEA inspired approach allows one to extract information from the data set with minimal transformation because the data is left "in place", and not accumulated into bins or scattered onto a grid. Thus our approach will provide the best time resolution possible for a given data set because there is no grid to be filled with data: the data defines the grid. Furthermore, our approach allows particle spectrometer developers to start to consider reactive, adaptive measurement strategies that allow the instrument to focus on features within velocity space, obtaining higher time and velocity space resolution.
The location of the data points in the angular dimensions are trivially bound, while the magnitude of the velocity has given upper and lower limits. Therefore, the data points lie within a rectangular volume in spherical coordinates. As long as there is a sufficient density of points within this volume, one obtains good results for both the automated discretization and the numerical calculation. Further details regarding the automated discretization are beyond the scope of this work and will be presented elsewhere, including a discussion of issues relating to troublesome degenerate points.

Let denote the direction and velocity state of the detector during measurement \(i\): \(\xi_i = \{v_i, \theta_i, \phi_i\}\). With this definition we write the estimate distribution function at \(\xi_i\) as \(f_i\). We can then evaluate the integrals Eq. 5–8 by numerical quadrature because we know the integrands as functions of the data: \(\xi_i\) and \(f_i\). After the triangulation, a list of tetrahedra is constructed. The volume of each tetrahedra is determined by its four vertices determined by \(\xi_i\). We determine the tetrahedra’s contribution to the integrals by linearly interpolating the appropriate functions of \(f_i\) and \(\xi_i\) from the nodes.

3. Testing Method

Application Form

In this paper we consider the effect of faults on this data analysis scheme. A goal of the REE effort is to allow instrument operators or scientists to run data analysis programs on board the spacecraft. Furthermore, REE seeks to provide a computing environment, based on commercial-off-the-shelf (COTS) components, augmented by hardware implemented fault tolerance in conjunction with SIFT (H/SIFT), that otherwise closely resembles the computing environment used by the ground-based scientist. Scientist generated analysis programs are to run on the REE Flight Processor and will suffer faults. Therefore, most of the PMA code was constructed using Fortran 90 and 95 with enough C and shell scripting to allow for library access and run management (Table 4). Much arithmetic was performed using double precision floating point.

To start to determine how faults affect such science programs, we have subjected the PMA to faults and have studied their effects. REE is in the process of bringing a hardware and software based fault injector on line; this fault injector will allow faults to be injected into various parts of the REE Flight Processor Testbed, including memory, cache, registers, etc. Currently, however, we do not have ready enough access to the run-time state of the PMA and processor to allow extensive testing of the dynamic portions of the PMA process.

On the other hand, we have full access to the static state of the program. Therefore to start our testing we have pursued the following simple procedure. The PMA analysis executable and model data were archived without compression. This archive was copied and then a randomly chosen bit in the copy was flipped. Every bit had an equal chance to be chosen. The PMA and data were then extracted from the archive and the analysis was run. The run-time environment for the tests used here is detailed in Table 4. The size of the executable and data set are given in Table 5; this shows the relative sizes of the two parts of the application, which implies that the executable likely suffered about 50% more faults than the data. No special effort was expended to reduce executable size or enhance run-time robustness, and the data set used did not model any particular instrument.

---

Table 4

<table>
<thead>
<tr>
<th>APPLICATION ENVIRONMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apple PowerBook 2000</td>
</tr>
<tr>
<td>384 MB RAM</td>
</tr>
<tr>
<td>Yellow Dog Linux Champion Server v1.2</td>
</tr>
<tr>
<td>Kernel version 2.2.15pre19</td>
</tr>
<tr>
<td>PLAPACK 2.08</td>
</tr>
<tr>
<td>GEOMPACK version 3</td>
</tr>
<tr>
<td>LAPACK version 2.0 (Very few routines used)</td>
</tr>
<tr>
<td>Message Passing Interface MPICH 1.2.0 &amp; p4 [15,9,16]</td>
</tr>
<tr>
<td>ABSOFT Fortran 90 version 3.0 &amp; bundled IMSL</td>
</tr>
<tr>
<td>GNU gcc 2.95.2</td>
</tr>
<tr>
<td>Optimization enabled for C &amp; Fortran compilers</td>
</tr>
</tbody>
</table>

---

1 By static state we mean that state available before run-time. Dynamic state is complementarily defined.
errors. Over three fourths of the serious errors were caused by bits flipped in the executable. Package level errors, e.g. PLAPACK or higher level MPI/user abort, are due mainly to data corruption. Error induced “user aborts” are evenly split between data and executable corruption. For segmentation violations caught in MPI, faults in the executable outnumber faults in the data by nine to one. The faults that produced the 16 intolerable results recorded in the Table 6 did not cause the PMA to crash; consistent with this is that 15 out of those 16 faults occurred in data.

4. RESULTS

The results of a series of runs is tabulated in Table 6. The zeroth moment (density) was examined for discrepancies from the non-corrupted control run. Two tenths of the runs either did not successfully complete, or produced a density that differed from true. However, half of the non-negligibly corrupted results, i.e. one tenth of the runs, produced densities that were within 0.5% of true and are therefore usable. Similar results are obtained if the first velocity moment, i.e. the bulk flow velocity $u$, is studied.

The types of errors are summarized in Table 7. Of the 88 errors that halted the program, just under half seem to have caught within the “lower” levels of MPICH and its p4 substrate. Another 15% were caught by the operating system (OS), which never faltered during these tests. Just under 25% of the serious errors occurred in PLAPACK or higher levels of MPI, e.g. incorrect PLAPACK arguments or faulty MPI communication parameters, etc. A number of Fortran run-time errors also occurred, though about half of them seem due to corruption of the archive and thus are not directly relevant to the current problem.

Table 8 shows the source of faults that caused the 88 serious errors. Over three fourths of the serious errors were caused by bits flipped in the executable. Package level errors, e.g. PLAPACK/DLINRG, are due mainly to data corruption. Error induced “user aborts” are evenly split between data and executable corruption. For segmentation violations caught in MPI, faults in the executable outnumber faults in the data by nine to one. The faults that produced the 16 intolerable results recorded in the Table 6 did not cause the PMA to crash; consistent with this is that 15 out of those 16 faults occurred in data.

In this work we have shown that a data analysis program written by a scientist for a COTS parallel computer and development system can have reasonable performance in the presence of faults to its static state. This is partially due to the robustness of double precision floating point arithmetic to single bit errors. Another reason for the application’s robustness is that PMA reduces the data to statistical moments, which reduces the overall impact of data errors. Furthermore, PMA operates as a filter that undergoes only a limited amount of iteration. Therefore, there is not a great deal of time for faults and their consequences to accumulate. Fresh, new processes will continually start and limit the effect of faults to old, expired processes.

Applications that require long lifetimes, servers, operating systems, system monitors, mission planner/schedulers may drive fault tolerance development, but science analyses seem from the start more robust. However, for science analyses that may affect mission operations, stringent real-time requirements on system or instrument health and safety may drive higher levels of science application robustness than we have observed with our experiments with the PMA. Certainly, given the inherent robustness of PMA, achieving these higher levels of reliability seems feasible within the REE technology development framework.

The logical continuation of this work is to perform a similar series of tests using the fault injection capabilities of the REE Testbed. With that system we will be able to determine how PMA responds to errors in its dynamic state. We doubt that the results will differ greatly from what we have found here, though there may be a somewhat wider range of consequences due to the broader range of faults. Certainly, we will gain a better appreciation for the behavior of the PMA when other parts of the computing subsystem suffer faults, because that is something we are currently not able to test. Our results point out that the REE emphasis on hardware, software libraries, and OS fault tolerance is well placed. PMA will be used to test the fault tolerant versions of these COTS components for space borne
applications.

6. ACKNOWLEDGEMENTS

Acknowledged is support from the Remote Exploration and Experimentation Project of NASA's High Performance Computing and Communications Program which is funded by the Office of Space Science with the work performed under contract NAS5-32350. A. Figueroa-Viñas played an important role during the formulation of this work.

REFERENCES