

Development of a miniaturised particle radiation monitor for Earth orbit

Edward Frank Mitchell

Department of Physics

Imperial College London

A thesis submitted for the degree of Doctor of Philosophy

May 2013

Abstract

Geometry and algorithm design for a novel highly miniaturised radiation monitor (HMRM) for spacecraft in medium Earth orbit are presented. The HMRM device comprises a telescopic configuration of application-specific active pixel sensors enclosed in a titanium shield, with an estimated total mass of 52 g and volume of 15 cm³. The monitor is intended to provide real-time dosimetry and identification of energetic charged particles in fluxes of up to 10⁸ cm⁻² s⁻¹ (omnidirectional). Achieving this capability with such a small instrument could open new prospects for radiation detection in space.

The methodology followed for the design and optimisation of the particle detector geometry is explained and analysis algorithms — for real-time use within the monitor and for post-processing reconstruction of spectra — are presented. Simulations with the Geant4 toolkit are used to predict operational results in various Earth orbits. Early test results of a prototype monitor, including calibration of the pixel sensors, are also reported.

Declaration

All material in this document which is not my own work has been properly attributed.

—Edward Mitchell

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Acknowledgements

I am very grateful to Henrique Araújo for his expert supervision of this work and to Doug Griffin for his enthusiastic support of my project contributions. I also wish to acknowledge additional helpful advice from Tim Sumner.

Vital assistance at RAL has been provided by Renato Turchetta, Nicola Guerrini, Sev Gunes-Lasnet, Simon Woodward and Olly Poyntz-Wright, while guidance from Alessandra Menicucci and Eamonn Daly at ESA has been greatly appreciated.

I would also like to thank my family for their ongoing support and encouragement.

Contents

1	Introduction	12
2	Distribution, interactions and detection of particles in Earth orbit	15
2.1	The energetic particle environment	15
2.2	Interactions of charged particles with matter	23
2.3	Radiation damage	28
2.4	Space radiation monitors	33
2.5	Summary	44
3	Monitor geometry design and optimisation	45
3.1	Monitor design methodology	45
3.2	Monitor architecture	46
3.3	Geometry optimisation methodology	52
3.4	Aperture window material selection	57
3.5	Sensor telescope optimisation	63
3.6	Outer casing and aperture	67
3.7	Monitor design implementation: HMRM version 1.0	72
3.8	HMRM geometric response functions	75
3.9	Reduction of the number of sensors	78
4	Radiation monitoring and event identification algorithms	81
4.1	Introduction to HMRM operation	81
4.2	Physics algorithm	86
4.3	Event identification algorithm	91
4.4	HMRM data tables	105
4.5	Summary	109
5	HMRM simulations	110
5.1	Monte Carlo validation	110
5.2	HMRM orbit simulations	122
5.3	Simulated charge sharing investigation	127
5.4	Summary	139
6	Reconstruction of particle energy spectra	140
6.1	The inverse problem	140
6.2	Matrix method	141

6.3	New method	149
6.4	Reconstruction in the presence of sensor noise	153
6.5	Preliminary reference orbit results	155
6.6	Future development	159
6.7	Summary	160
7	Preliminary testing of the HMRM active pixel sensors	161
7.1	The HMRM ASIC	161
7.2	Sensor noise and signal model	165
7.3	Calibration overview	170
7.4	Data acquisition system	171
7.5	Initial tests	173
7.6	Dark calibration	177
7.7	Summary	189
8	Conclusion	190
	Appendices	201
A	List of acronyms	201
B	HMRM ASIC high-level design	203
C	Geometry optimisation results	205
D	HMRM physics algorithm flowchart	207
E	Orbit simulation ID table	213
F	Reconstructions of mean particle spectra in orbits B, C and E	214

List of Figures

2.1	Mean trapped electron energy spectra for the HMRM reference orbits	18
2.2	Mean trapped proton energy spectra for the HMRM reference orbits	18
2.3	Omnidirectional integral fluxes for electrons and protons in orbits A and B .	19
2.4	Omnidirectional integral fluxes for electrons and protons in orbits C, D and E	19
2.5	Proton SEP energy spectra	21
2.6	Mean solar proton spectra in different Earth orbits	22
2.7	Modelled GCR spectra for protons, helium, carbon and oxygen ions	23
2.8	Mean CSDA ranges in common materials for electrons and protons	25
2.9	Electron and proton stopping powers in aluminium	27
2.10	Landau energy loss distribution for protons	28
2.11	Simplified diagram of the GPB proton monitor	40
3.1	Conceptual diagram for the HMRM geometry	50
3.2	Pile-up probability as a function of mean particle incidence rate	51
3.3	HMRM Geant4 optimisation geometry	55
3.4	Diagram of the circuit track layout for one HMRM ASIC sensor pixel	56
3.5	The spectral reflectivity of pure aluminium	59
3.6	Calculated spectral transmissivity of three thicknesses of aluminium foil . . .	59
3.7	Estimates of the sensor interactive quantum efficiency	60
3.8	Estimated mean pixel signal due to sunlight for an aluminium foil window . .	61
3.9	Fractional isotropic fluences detected by S1 with a polyimide window	62
3.10	Fractional isotropic fluences detected by S1 with an aluminium window . . .	62
3.11	Window transmission energies (50%) for electron and proton fluences	63
3.12	Diagram showing the arrangement of ASICs D1 and D2	64
3.13	Fractional electron and proton isotropic fluences detected by S1 and S2 . . .	65
3.14	Distribution of the electron exit distance from the beam axis.	67
3.15	Distribution of the kinetic energy of electrons which exit the shield sample. .	67
3.16	Illustration of the entrance aperture definitions	68
3.17	The HMRM version 1.0 Geant4 geometry model	73
3.18	Comparison of the Geant4 geometry model and electromechanical CAD . . .	73
3.19	Cross-section view of the HMRM Geant4 geometry	74
3.20	Photograph of the first HMRM prototype unit.	74
3.21	Energy dependent effective area of S1 for electrons and protons	76
3.22	Energy dependent effective area for the four detection modes	76
3.23	Predicted mean particle incidence rate on S1 in the five reference orbits . . .	77

3.24	Predicted pile-up probability on S1 in the five reference orbits	77
3.25	Azimuthal dependence of the HMRM unidirectional effective area for C2 mode protons	79
3.26	Elevation angle dependence of the HMRM unidirectional effective area for protons	79
3.27	Elevation angle dependence of the HMRM unidirectional effective area for C1 mode electrons	80
3.28	Elevation angle dependence of the HMRM unidirectional effective area for elec- trons	80
4.1	Photograph of the HMRM FPGA mounted on the monitor central PCB section	83
4.2	The omnidirectional integral proton flux exceeding 1 MeV as a function of time in two orbits	85
4.3	The particle count rate obtained by counting empty readout frames	88
4.4	The mean expected number of empty frames as a function of pile-up probability	89
4.5	Energy deposit probability distributions in a 12 μm silicon layer	92
4.6	Mean energy deposits in a 12 μm silicon layer due to protons, electrons and alpha particles	93
4.7	Simulated energy deposit distributions in S1 due to isotropic monoenergetic fluences	94
4.8	Simulated energy deposits in S1 due to isotropic fluences	95
4.9	Simulated energy deposits in S1 due to isotropic fluences (C1 mode)	96
4.10	Coincident energy deposits in HMRM sensors S1 and S2 for C2 mode events .	97
4.11	Distribution of coincident energy deposits in S1 and S2 for C2 mode events .	97
4.12	Distribution of coincident energy deposits in S2 and S3 for C3 mode events .	98
4.13	Distribution of coincident energy deposits in S3 and S4 for C4 mode events .	99
4.14	Omnidirectional response to electrons and protons for two ID table channels .	103
4.15	Omnidirectional response of channel 28, a pure proton channel	103
4.16	Distribution of coincident energy deposits in S1 and S2 for C2 mode events .	104
4.17	Estimated misidentification probability for events involving a single sensor . .	106
4.18	Conceptual diagram of the relation of threshold levels and digitised pixel values	107
4.19	Reconstructed energy deposit as a function of the original value	108
5.1	Directional backscatter from aluminium	113
5.2	Backscatter data for 35 keV electrons	114
5.3	Mean energy losses of electrons passing through thin aluminium layers	115
5.4	Energy losses of 5.49 MeV alpha particles passing through thin aluminium layers	115
5.5	Angular scattering distribution widths of a 158.6 MeV proton beam	116
5.6	Depth-dose curve in aluminium for a 1.03 MeV electron beam	117
5.7	Depth-dose curves in aluminium for electron beams	118
5.8	Depth-dose curves in high density polyethylene due to ion beams	118
5.9	Depth-dose curves in high density polyethylene due to particle beams	119
5.10	Energy deposit data for electrons incident on silicon sensors	120
5.11	Most probable energy deposit data for protons incident on silicon sensors . .	120
5.12	The FWHM of the energy deposit distributions for protons incident on silicon sensors	121

5.13	Energy deposit data for protons of various energies	121
5.14	Most probable energy deposit values for protons	122
5.15	The omnidirectional integral fluxes for electrons and protons in orbit A . . .	123
5.16	Simulated count rate results for reference orbit A	124
5.17	Simulated dose rate results for reference orbit A	125
5.18	Comparison of measured and ‘true’ count rate	126
5.19	Simulated event count rates for four ID table channels	127
5.20	Simulated event identification counts in all channels	127
5.21	Conceptual diagram of a cross-section through a CMOS APS	128
5.22	Central portion of the charge diffusion simulation geometry	131
5.23	Simulated inward net current and cumulative charge collected for a diode . .	132
5.24	The charge fraction collected by the nine test diodes after 5 μ s	133
5.25	Charge sharing response function	134
5.26	The total fractional charge collected by nine diodes	134
5.27	The time taken for a diode to collect 95% of its long-term charge	134
5.28	Simulated distribution of pixel charge shares	136
5.29	Simulated distribution of pixel charge shares with noise	137
5.30	Distribution of the pixel cluster sums for 5.9 keV photon exposure	138
6.1	HMRM channel counts allocated in an E^{-1} differential flux	143
6.2	Matrix method spectrum reconstruction results for electrons	144
6.3	Matrix method spectrum reconstruction results for protons	144
6.4	Matrix method spectrum reconstruction results for electrons and protons . .	145
6.5	Matrix method spectrum reconstruction results for protons and electrons . .	146
6.6	Combined fractional RMS errors for proton and electron E^{-1} spectra	146
6.7	Fractional RMS errors for E^{-1} energy spectra	147
6.8	HMRM channel counts for a 110 s simulated exposure in reference orbit E . .	148
6.9	Matrix method spectrum reconstruction result for orbit E with six energy bins	149
6.10	Matrix method spectrum reconstruction result for orbit E with three energy bins	150
6.11	Illustration of spectrum modelling and response function modulation	151
6.12	New method spectrum reconstruction result for orbit E	153
6.13	Counts allocated to channels 15–32 in response to a simulated E^{-1} differential flux	154
6.14	Fractional reconstruction error for E^{-1} spectra	154
6.15	Comparison of the orbit A mean integral energy spectra with the peak flux spectra	156
6.16	Total integral fluxes of electrons and protons in reference orbit A	156
6.17	Electron spectrum reconstruction for orbit A	157
6.18	Proton spectrum reconstruction for orbit A	157
6.19	Electron spectrum reconstruction for orbit D	158
6.20	Proton spectrum reconstruction for orbit D	158
7.1	Photograph of an HMRM ASIC mounted on the monitor front PCB	162
7.2	Simplified circuit diagram of a 3T pixel	162
7.3	Simplified circuit diagram of a 4T pixel	163

7.4	Conceptual diagram of threshold levels and digitised pixel values	169
7.5	On-screen view of the HMRM data acquisition software	172
7.6	Oscilloscope capture showing FPGA-generated clock and data	173
7.7	Predicted temperature profile of the HMRM spacecraft panel	175
7.8	HMRM transparent readout for sensors 2 and 3	176
7.9	Threshold scan for comparator level L1 of sensor 2.	177
7.10	L2 threshold scan for sensor 2 calibration.	178
7.11	Error function fits to the L2 scan data for pixels 53 and 305	179
7.12	Distribution of mean pixel signals for sensor 2	179
7.13	Mean pixel signals measured for S2 in darkness with zero trims	180
7.14	Pixel dark noise, measured in the absence of radiation	181
7.15	Linear fit to the pixel mean signal as a function of trim setting	182
7.16	The distribution of trim gains for S2	182
7.17	Predicted pixel offset for S2 after trimming	183
7.18	Dark signal probability distribution for S2, using a comparator software trim	184
7.19	Results of the L2 threshold scan for sensor 2	186
7.20	Mean dark pixel signals measured using a L2 threshold scan for S2	186
7.21	Pixel dark noise map and histogram for S2, after applying trims.	187
7.22	Combined pixel probability distribution for S2	187
7.23	Pixel offset and noise distributions for sensor 3	188
B.1	HMRM ASIC high-level design	204
D.1	FPGA physics algorithm ‘main function’	208
D.2	FPGA physics algorithm: pixel signal operations.	208
D.3	FPGA physics algorithm: rate monitoring	209
D.4	FPGA physics algorithm: dose calculation	210
D.5	FPGA physics algorithm: event identification	211
F.1	Electron spectrum reconstruction for orbit B	215
F.2	Proton spectrum reconstruction for orbit B	215
F.3	Electron spectrum reconstruction for orbit C	216
F.4	Proton spectrum reconstruction for orbit C	216
F.5	Electron spectrum reconstruction for orbit E	217
F.6	Proton spectrum reconstruction for orbit E	217

List of Tables

2.1	Reference Earth orbits selected for the HMRM project.	16
3.1	Composition of the simulated HMRM ASIC sensor	57
3.2	Detection thresholds (50% transmission) for particle flux reaching the four sensors	65
3.3	Particle scattering and transmission data for 2 MeV electrons incident on material layers	66
3.4	Calculated effective areas for S1–4 and C1–4 using two aperture definitions .	69
3.5	Parameter values for the two proposed aperture definitions	70
3.6	HMRM detector geometry design recommendations.	72
4.1	Provisional ID table for the four-sensor HMRM	102
4.2	Provisional comparator level choices	108
4.3	Provisional pixel energy weightings	108
7.1	Truth table for creation of the pixel values by digital encoder	164
7.2	Nominal inter-level gains for the HMRM ASIC comparator levels.	169
7.3	Overall power consumed by the HMRM	174
7.4	Chosen relative trim settings for HMRM sensor 2	184
C.1	Details of the geometry optimisation variations	205
C.2	Identification purities and efficiencies for protons	206
C.3	Identification purities and efficiencies for electrons	206
D.1	Externally specified physics algorithm parameters	212
D.2	Internally generated physics algorithm variables	212
E.1	ID table used for the orbit A simulation	213

Chapter 1

Introduction

The Earth orbit environment is populated by a great variety of energetic charged particles. The interactions of these particles with spacecraft systems and payloads lead to the cumulative degradation of components and to single-event errors in electronic devices. These may result in system malfunction, reduction of mission lifetime and even loss of a spacecraft. Safety concerns for human space habitation and exploration pose even greater challenges from this perspective.

The data provided by monitoring devices help in assessing these risks and in correlating radiation effects with the radiation environment. This may result in improved mission planning, recommendations for spacecraft design and introduces the possibility of real-time alerting.

Monitors in current use may be divided into two broad categories: scientific payloads and support instruments. The former have good particle measurement capability, with large mass ($\gg 1$ kg) and power requirements ($\gg 1$ W). Smaller support instruments, however, have limited functionality (e.g. dosimetry) and offer little or no particle discrimination. Damage effects depend strongly on particle species and energy, meaning that particle identification would be an important advantage for these devices. The development of a small, accurate instrument suitable for widespread use on satellites in Earth orbit could therefore open new prospects for radiation detection in space. This thesis describes many aspects of the development of the Highly Miniaturised Radiation Monitor (HMRM), an innovative particle detector intended for spacecraft radiation environment monitoring.

In April 2009, the European Space Agency (ESA) invited tenders for a contract to develop a small particle radiation monitor for use in medium Earth orbit [1]. The programme covers Phases A and B of the design process which are, respectively, the feasibility study and prototype construction stages. The successful bid to design the HMRM was led by the Space Science Department of the STFC Rutherford Appleton Laboratory (hereafter RAL Space), in partnership with RAL Technology Department and the High Energy Physics Group at Imperial College London (Imperial); the project started formally in February 2010. Imperial has made extensive contributions to the technical design, Monte Carlo and functional simulations, on-board and offline algorithm design and hardware testing of this device. I have made

major contributions to all of these areas and have undertaken most of the work described in this thesis, which is organised as follows.

Chapter 2 provides an introduction to the energetic particle environment in Earth orbit and summarises how these particles interact with matter, causing damage to spacecraft components. Five reference Earth orbits are selected to provide specific examples of possible HMRM use cases; these are simulated at several stages of the development process to assess monitor performance. Space radiation monitors in past and present use are also reviewed, covering a wide range of instrument sizes, detection methods and capabilities. This information is used to draw conclusions and implications for the design of the HMRM.

The hardware design of the HMRM is discussed in Chapter 3. The monitor detects and identifies particles by measuring their energy loss in a telescope of thin silicon active pixel sensors. This technology, developed originally for High Energy Physics experiments, promises unique advantages for miniaturised space-based applications. The detector geometry (the layout and principal characteristics of the sensor telescope, casing and aperture) of this conceptual design was optimised with respect to particle identification and counting ability. This process, described in this chapter, resulted in a set of design recommendations to which I was a leading contributor. The monitor electromechanical design, developed by RAL Space, is also presented together with its implementation as a Geant4 geometry model for particle Monte Carlo simulation.

Chapter 4 describes the radiation monitoring analysis algorithms which are executed within the HMRM, and assesses the potential for particle identification. Probabilistic identification is achieved by measuring the particle energy loss rate in multiple sensors, with fast analysis performed on an event-by-event basis. Particle count rates and ionising dose rates are also measured; these data are made available in small data packets suitable for rapid appraisal. The monitoring and identification algorithm was delivered to RAL Space for implementation in the prototype monitor (together with the device control and communication functions).

Chapter 5 considers the simulated aspects of the HMRM project in further detail. The Geant4 Monte Carlo physics implementation used within the project is validated by comparison with experimental particle data, showing excellent agreement. Using the detailed monitor simulation geometry, the HMRM operation is modelled in a time-dependent Earth orbit environment, confirming that the design is likely to provide useful monitoring data. Finally, the charge diffusion process in a pixel sensor is investigated using a new simulation model, giving insight into the behaviour of the HMRM sensors in response to ionising energy deposits.

Methods for reconstructing the incident particle energy spectra, using simulated monitor data, are explored in Chapter 6. After an introduction to the problem, a commonly used deconvolution method is applied to the HMRM data and its efficacy evaluated. Deficiencies in this scheme are identified, leading to the presentation of an improved method which was developed for the HMRM project. This is applied to the reconstruction of several modelled Earth orbit particle flux environments, demonstrating improved reconstruction ability. Further enhancements to the procedure, to be undertaken in future work, are also proposed.

Finally, Chapter 7 presents the results of initial tests and calibration procedures for the HMRM application specific integrated circuit (ASIC) sensors, which began in December 2012. I contributed to this experimental work as part of a team and was a leading contributor to the design of the calibration technique and data acquisition/analysis software. After an introduction to the ASIC high-level design, a pixel noise and mean signal model are introduced. These models are used to guide the calibration process, whereby pixel noise and pedestal values are extracted. Compensation of the pedestal dispersion fixed pattern noise is demonstrated by the successful application of hardware trim settings.

Testing of the HMRM prototype continues; one unit is currently undergoing integration on the TechDemoSat-1 spacecraft, with a flight opportunity in 2013. Future phases of the programme will address design iteration and spaceflight qualification, with completion predicted for 2015–2016. A successful HMRM development and qualification process may pave the way for widespread use of this novel technology in many commercial and scientific settings.

Chapter 2

Distribution, interactions and detection of particles in Earth orbit

In this chapter the general characteristics of the various particle populations found in Earth orbits are described. Relevant interaction and energy loss mechanisms for these particles are discussed, together with a summary of their damaging effects on spacecraft. Finally, various particle sensor technologies and existing radiation monitors are compared and evaluated, providing guidance for the design of the HMRM.

2.1 The energetic particle environment

Due to the wide variety of particles present near the Earth, and the need to model this environment effectively, it is convenient to split the population into a number of discrete components. These are organised approximately by particle location or origin, although the energy ranges and spatial distributions of these components often overlap. The three considered here are magnetospheric particles, solar particles and galactic cosmic rays (GCRs). Each of these, summarised in the following sections, includes multiple particle species which span many orders of magnitude in energy.

2.1.1 HMRM reference orbits

To help assess the relative importance of the different particle components, five reference orbits have been selected for analysis within the HMRM project [2]: a low Earth orbit (LEO), three medium Earth orbits (MEO) and a geostationary orbit (GEO). These should be representative of the radiation environment experienced by the overwhelming majority of spacecraft and have been referred to during the monitor design and simulation phases. Table 2.1 lists the parameters of these orbits, which are motivated as follows:

1. **Orbit A** is a sun synchronous orbit (SSO), where altitude and inclination evolve such

that satellites ascend or descend over a point on Earth at the same mean solar time, thereby maintaining an almost constant solar illumination angle. This is typically a near-polar retrograde LEO (98° inclination) at 600–800 km altitude (here chosen to be 700 km). A constant solar illumination angle is useful for visible and infrared imaging, for example, with weather and remote sensing satellites.

2. **Orbit B** is a highly elliptical orbit like that followed by some space science missions (e.g. FAST [3]) and is the only eccentric orbit considered. Perigee and apogee altitudes of 400 km and 4,000 km have been selected, with the apogee occurring over the North Pole.
3. **Orbit C** is circular at 10,000 km altitude; this is an attractive MEO for remote sensing purposes as it represents a compromise between LEO (high spatial resolution with low temporal resolution) and GEO (low spatial resolution but high temporal resolution).
4. **Orbit D** represents navigation constellations such as the ESA Galileo system (altitude 23,222 km, 56° inclination) and the similarly located Global Positioning System (GPS) satellites at 20,200 km, 55° inclination.
5. **Orbit E** is a geostationary orbit (GEO) at 35,786 km and is an important location for communication and meteorological satellites.

Table 2.1: Reference Earth orbits selected for the HMRM project.

Orbit	Type	Altitude / km	Inclination / deg	Period / hours
A	LEO	700	98	1.65
B	MEO	400 - 4,000	83	2.19
C	MEO	10,000	0	5.79
D	MEO	23,222	56	14.07
E	GEO	35,786	0	23.93

2.1.2 Magnetospheric particles

The particle radiation environment in Earth orbit is dominated by protons and electrons trapped in the Earth’s magnetosphere. These particles are mainly of solar origin, while some are produced from the decay of neutrons released in GCR interactions with the neutral atmosphere [4]. They are confined to ‘belt’ regions due to their movement in the Earth’s magnetic field, which is often divided into three characteristic motions:

1. **Circular motion about the magnetic field direction.** The Lorentz force on the charged particle causes circular motion in the plane perpendicular to the magnetic field. A component of velocity parallel to the field creates the general case of a helical trajectory.
2. **Bounce motion between the poles.** The approximate dipole configuration of the Earth’s field subjects the particle to an increasing field strength as its helical path takes it closer to either magnetic pole. As the gyroradius shrinks, conserving the magnetic moment, the parallel component of velocity is reduced and eventually reversed at the

‘mirror point’, reflecting the particle towards the other pole. This may occur at low enough altitudes for the particle to be absorbed by the neutral atmosphere before it can be reflected.

3. **East-west drift.** Gradient, curvature and gravitational plasma drifts act in the east-west direction. The net effect is a movement of negative charges to the east and positive charges to the west (producing an electric ring current).

The stably trapped population consists predominantly of protons with energies up to several hundred MeV and electrons up to ~ 10 MeV. The proton belt intensity peaks near altitudes of one Earth radius ($1 R_E$) with omnidirectional integral fluxes of more than $10^5 \text{ cm}^{-2} \text{ s}^{-1}$; this reduces with altitude to $\sim 1 \text{ cm}^{-2} \text{ s}^{-1}$ at $3 R_E$.

The electron spatial distribution is different to that of the proton, with two intensity peaks around altitudes of $0.5 R_E$ and $4 R_E$, where integral fluxes exceed $10^8 \text{ cm}^{-2} \text{ s}^{-1}$. An intensity minimum (the ‘slot’) exists between these at $\sim 2 R_E$. The cusps of the outer electron belt follow the magnetic field lines to low altitudes at high latitudes. This causes a region of increased electron intensity in low Earth orbit at latitudes of 60° – 80° , known as the ‘polar horns’.

While proton fluxes are not significantly greater in the polar horns, a region of increased proton flux exists at approximately $30^\circ \text{ S } 30^\circ \text{ W}$ (to the south-east of Brazil). This location is known as the South Atlantic Anomaly (SAA) and occurs because the magnetic dipole axis of the Earth’s field is offset from the Earth’s centre. The proton belt therefore reaches a lower altitude over that part of the Earth’s surface.

In addition to electrons and protons, trapped heavy ions have been detected at altitudes of $\sim 1 R_E$, believed to be decelerated anomalous cosmic ray ions [5]. The intensities of these ions are, however, several orders of magnitude lower than those of trapped protons in this region.

The NASA AP-8 and AE-8 models

The typical trapped particle energy spectra expected in specific orbits may be obtained from the empirical NASA models AP-8 [6] and AE-8 [7] for protons and electrons, respectively. The spectra are provided as a function of the magnetic field coordinates for protons in the range 0.1 – 400 MeV and electrons in the range 0.04 – 7 MeV. These are static, omnidirectional models, whereas experimental data reveal significant temporal and angular variability. However, they remain widely used references due to their comprehensive spatial and energy coverage.

As well as variations with spatial location, trapped particle populations vary in time, both over short timescales of hours or days (in relation to transient solar events) and over years (due to changes in the Earth’s magnetic field and the ~ 11 year solar activity cycle). These variations are not modelled beyond a simple choice of solar maximum or minimum: as a consequence the modelled spectra can only be interpreted as a best-estimate for a single point in time and are usually treated as having an uncertainty of a factor of two [8].

The AP-8 and AE-8 modelled long-term mean energy spectra for the five reference orbits

at solar maximum are shown in Figs. 2.1 and 2.2. These data were obtained via the ESA Space Environment and Information System (SPENVIS) [8, 9]. Proton fluxes are consistently lower than electron fluxes in these orbits. The relative energy ranges are also different: while electron spectra extend to relativistic energies, most protons are non-relativistic. Of particular note is orbit C, which experiences a very harsh environment with large total fluxes of both electrons and protons. For this reason, and despite its optimal location from an Earth coverage perspective, this orbit is rarely used at present.

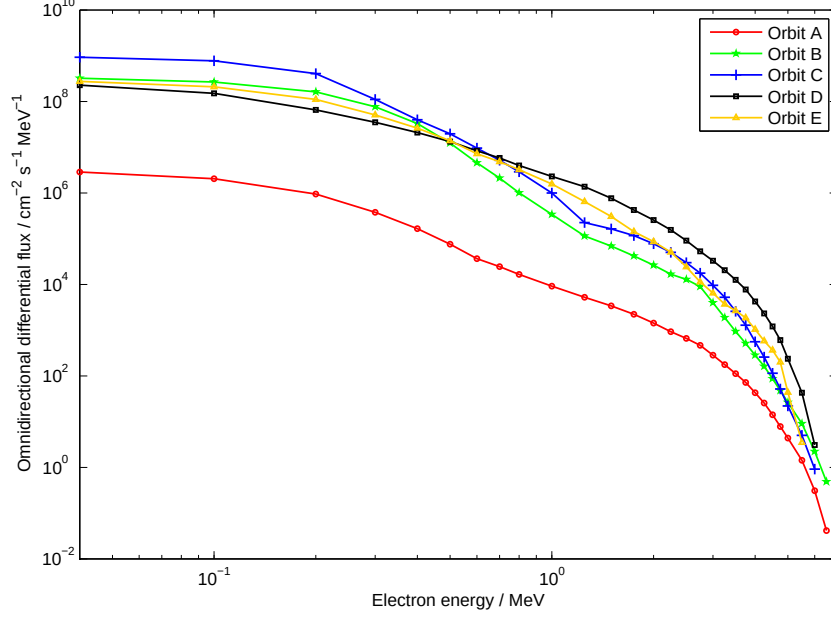


Figure 2.1: Mean trapped electron energy spectra for the HMRM reference orbits, derived from the AE-8 model for solar maximum.

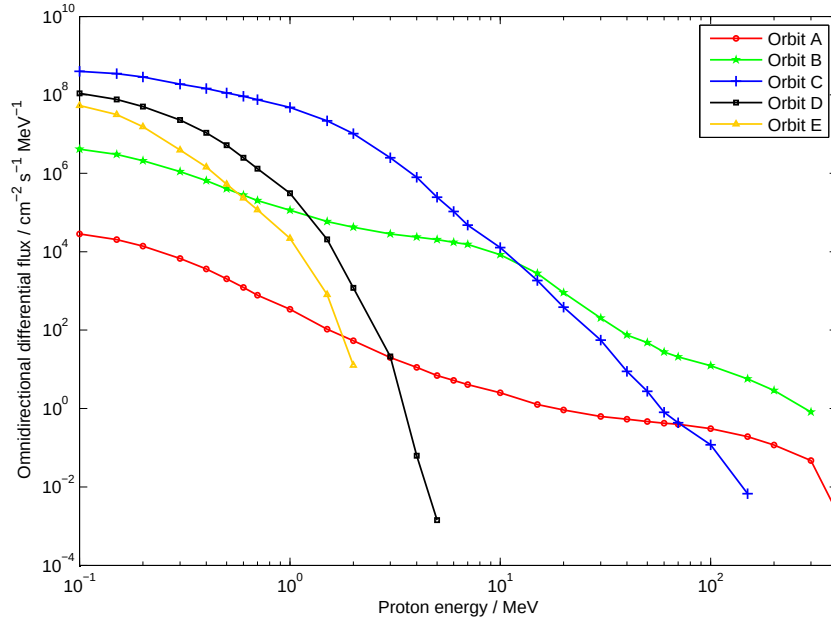


Figure 2.2: Mean trapped proton energy spectra for the HMRM reference orbits, derived from the AP-8 model for solar maximum.

MEO environments are the most challenging as they are conditioned by the complex shape of the magnetosphere and include the most intense regions of the radiation belts. The distribution of the geomagnetic field lines is such that spacecraft will experience charged particle populations which vary widely with altitude, latitude and longitude. Given this diversity, it is also important to study the changing radiation environment experienced *within* these orbits, rather than considering orbit-averaged values on their own. For example, Figs. 2.3 and 2.4 show the AE-8 and AP-8 modelled trapped electron and proton fluxes above selected energies, as functions of orbital time; a worst-case orbit is shown for each (selected to pass through the SAA and polar horns if applicable). Therefore, although these models do not assess transient effects in the environments themselves, rapid changes in flux are experienced due to the spacecraft's passage through the radiation belt regions.

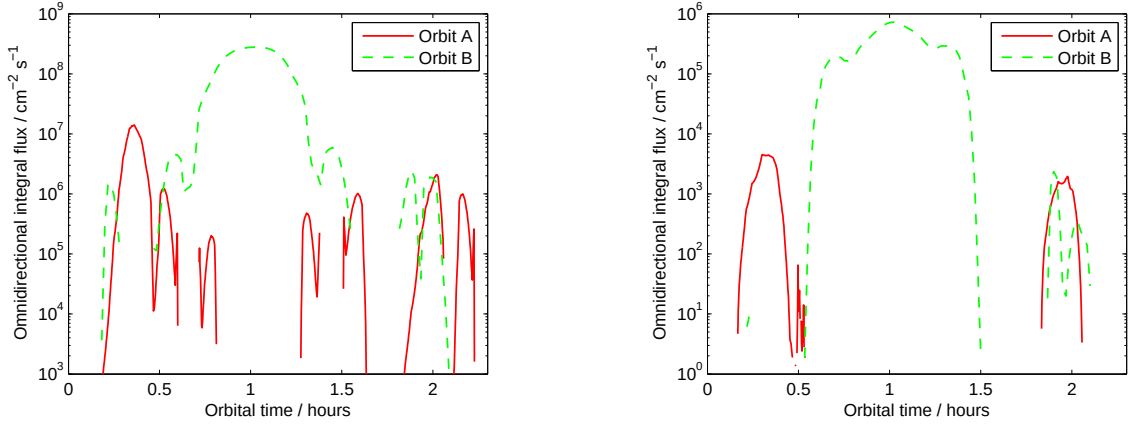


Figure 2.3: Omnidirectional integral fluxes for electrons with energies greater than 40 keV (left) and protons with energy greater than 1 MeV (right) for reference orbits A and B, according to the AE-8 and AP-8 models.

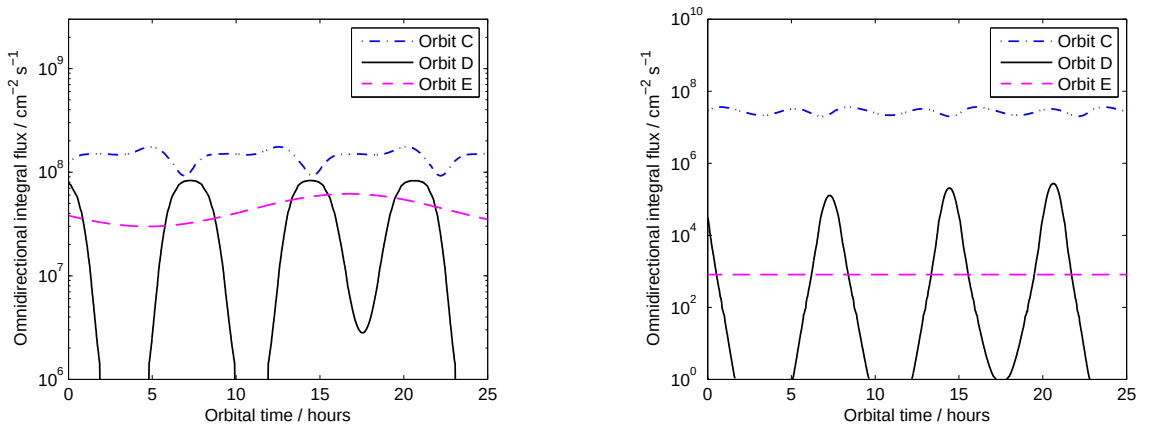


Figure 2.4: Omnidirectional integral fluxes for electrons with energies greater than 40 keV (left) and protons with energy greater than 1 MeV (right) for reference orbits C, D and E, according to the AE-8 and AP-8 models.

Anisotropy

Although the AP-8/AE-8 models supply omnidirectional integral flux values, trapped particles (and especially protons) may display significant anisotropy. One important cause is the ‘east-west effect’ which results from particle absorption in the neutral atmosphere at low altitudes. The magnetic gyration radius of protons may be of order 100 km, which is comparable to the atmospheric density scale height. Particle absorption depends on the altitude of the gyroradius centre and therefore upon whether this point is to the east or west of the observer. The resulting preferential absorption causes a flux variation in the azimuthal plane of factors of two to four [10].

A further anisotropic effect results from the distribution of particle pitch angles (angle between the velocity vector and the magnetic field) which determines the lowest altitude the particle may reach [11].

2.1.3 Solar energetic particles

Charged particles are emitted continuously from the sun (known as the ‘solar wind’) at velocities of approximately 400 km s^{-1} and 800 km s^{-1} at solar minimum and maximum, respectively [12]. Aside from surface charging, these low energies (of order 1 keV) are of little concern for spacecraft operation. However, transient solar energetic particle (SEP) events also occur periodically, where large fluxes of higher energy particles are released. These are associated with solar flare eruptions from the solar surface and with large-scale releases of magnetised plasma from the sun, known as coronal mass ejections (CMEs).

SEPs deriving from flares and CMEs (termed *impulsive* and *gradual*, respectively) have been characterised using space- and ground-based particle detectors in the following ways:

1. **Impulsive SEPs** evolve over timescales of hours and are characterised by low ion intensities and relative ion abundances which differ from those of the corona, while protons typically outnumber electrons by a factor of less than 10 [13]. Particle acceleration is thought to occur at the sun via magnetic reconnection, limiting the particle energy to less than $\sim 50 \text{ MeV/n}$. These events are only detected at Earth if the event originates in the eastern hemisphere of the sun, such that there is a direct connection through the interplanetary magnetic field.
2. **Gradual SEPs** typically evolve over periods of days. Particle acceleration is hypothesised to occur in the interplanetary medium between sun and Earth, driven by turbulence in magnetic shock fronts [14]. This allows acceleration to $\sim \text{GeV/n}$ energies and thus poses a greater threat to spacecraft. These SEPs match the coronal ion abundances, while protons outnumber electrons by a factor of more than 10, and may be detected from any origin on the sun.

Individual SEPs may be described by time-dependent energy spectra obtained from spacecraft data. For example, Fig. 2.5 shows two examples of large SEP events detected by satellites in Earth orbit: a peak proton spectrum from February 1984 (Zhang *et al.* [12]) and a series of three spectra (at ten minute intervals) from January 2005 (Wang [15]). The latter event

(measured by GOES-11 in geostationary orbit) is estimated to be the largest in a 10–20 year period and shows a peak flux more than ten times the magnitude of the former event. These data also show that proton spectra are harder prior to the event peak and softer afterwards, due to the different particle travel times.

Total fluences in individual SEP events can vary by a factor of 10^4 ; this means that the total exposure of a spacecraft over its operational life will be dominated by a small number of rare, extreme events [16]. Larger SEPs are therefore important as they will determine the overall total dose received.

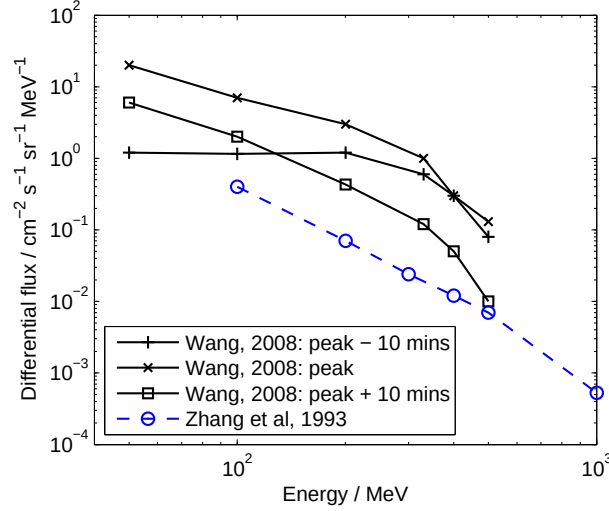


Figure 2.5: Proton energy spectra for a large SEP event at ten minute intervals [15] and a peak spectrum for a separate event (data provided in [12]).

The relative exposure to solar particles depends on the extent of geomagnetic shielding, which is determined by the altitude and latitude of the orbit. Higher altitude orbits (such as geostationary) and higher latitude orbits (such as polar low Earth orbits) are most exposed, while the penetration of the particle increases with magnetic rigidity. The shielding effect is illustrated by the modelled mean solar proton spectra shown in Fig. 2.6. These spectra are derived from an extension of the JPL-91 models [17] by Rosenqvist *et al.* [18] and rely on a complete list of solar proton event fluences between 1974 and 2002. The derived spectra are 99% confidence levels for the flux not being exceeded and are solar-cycle averages.

In addition to predictions for the reference orbits, a series of equatorial orbits are shown to demonstrate the progressive shielding of all but the highest energy particles. Smaller reductions are obtained for the reference orbits since most are either polar (A and B) or high altitude (D and E) and thus have little shielding. Orbit C (not shown) is an exception: negligible solar flux is predicted at this altitude due to its equatorial inclination. This contrasts with its exposure to trapped particles, for which it is the most hazardous reference orbit.

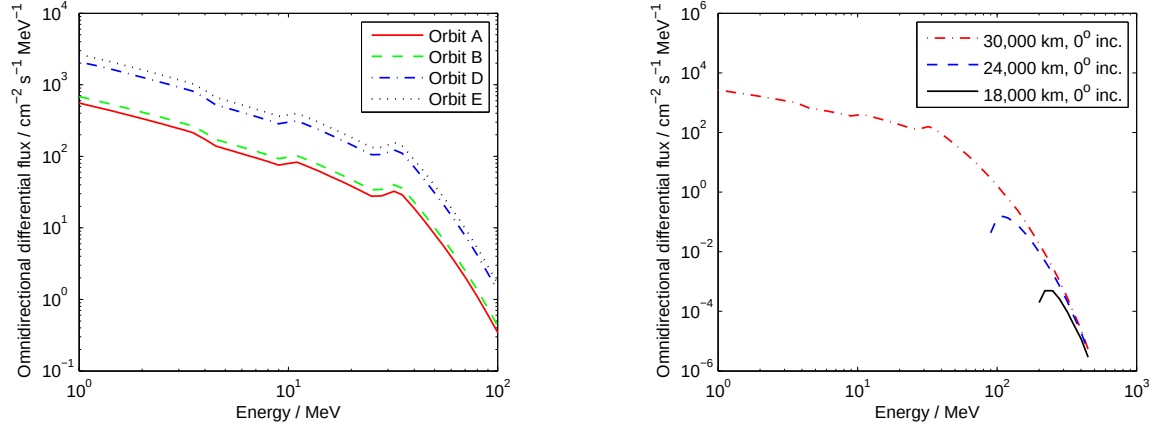


Figure 2.6: Mean solar proton spectra in different Earth orbits. Left: reference orbits A, B, D and E (orbit C is predicted to have negligible solar proton flux). Right: predictions for a series of equatorial orbits, showing the variation in magnetospheric shielding.

2.1.4 Galactic cosmic rays

GCRs are energetic particles originating in our own galaxy which are accelerated in astrophysical shock fronts such as supernova remnants [19]. They comprise approximately 85% protons, 12% helium nuclei (^3He and ^4He), 1% heavier nuclei and 2% positrons/electrons, with all nuclei fully ionised [20]. This composition includes the original accelerated particles as well as the products of fragmentation reactions of the heavier nuclei (known as the primary and secondary components).

Measurements have shown that the cosmic ray flux is highest at solar activity minimum, when the interplanetary magnetic field is weakest [21]. At solar maximum, the increased magnetic field provides greater shielding of the solar system, decreasing GCR flux near Earth. Furthermore, interactions with the solar wind cause a decrease in flux at energies below ~ 500 MeV/n. This is observable in the modelled energy spectra provided in Fig. 2.7, which also shows the power law trend at higher energies (with an index of approximately -2.7).

The omnidirectional integral flux of GCRs is estimated as $1.9 \text{ cm}^{-2} \text{ s}^{-1}$ and $4.2 \text{ cm}^{-2} \text{ s}^{-1}$ for solar maximum and minimum, respectively. Therefore, while these high energy particles cannot be blocked by spacecraft shielding, their incidence rate is far lower than the other particle components considered here.

Anomalous cosmic rays, mentioned in Section 2.1.2, are distinct in that they are singly charged and have lower energy (generally up to 30 MeV). These are thought to be due to the ionisation of neutral gas at the boundary of the solar system by UV radiation. After acceleration at the termination shock, they may propagate towards the Earth [22, 23]. Although their smaller charge increases their rigidity (and hence magnetospheric penetration), their relatively small flux makes them a minor contribution to the overall radiation environment.

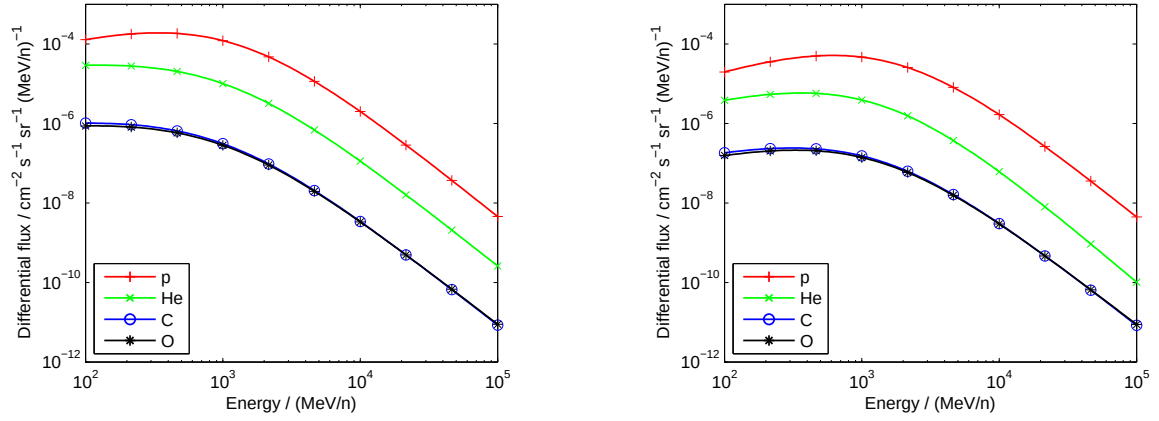


Figure 2.7: Modelled GCR spectra for protons, helium, carbon and oxygen ions, at solar minimum (left) and solar maximum (right) above the Earth’s atmosphere. These spectra are a parameterised function fit to experimental data [12].

2.2 Interactions of charged particles with matter

The overview of the particle radiation environment in Earth orbit has shown that electrons with energies up to ~ 6 MeV and protons/ions up to ~ 500 MeV/n constitute the majority of the flux. To understand how these particles may affect objects in orbit, and how they may be detected, a brief summary of their interactions and energy loss mechanisms in matter is presented.

An energetic charged particle passing through matter undergoes a series of random, independent interactions with the atomic electrons and nuclei of the target material. These are dominated by electronic (‘collisional’) interactions with the electrons which may cause ionisation or atomic excitation. Since the energy loss at each interaction is typically much less than the total kinetic energy of the projectile, the overall effect is a quasi-continuous transfer of energy to the target medium.

As well as energy loss, Coulomb scattering (especially from the target nuclei) causes deviation of the projectile trajectory. The cumulative effect of many such small angle deflections as the particle crosses a material layer (multiple scattering) is accurately described by Molière theory [24]. In addition, single-event, wide-angle scattering (Rutherford scattering) may occur.

Coulomb interactions may leave an atom in an excited state; subsequent de-excitation results in the emission of X-rays or Auger electrons. Larger energy transfers resulting in ionisation can release an electron with sufficient energy to cause further ionisation (a delta ray). These delta rays may travel some distance before depositing their energy: together with radiative losses, this means that the *local* energy deposit may be less than the energy lost by the primary particle.

Although collisional losses are most important at the particle energies found in Earth orbit, other processes also contribute. Bremsstrahlung losses for electrons are approximately pro-

portional to the particle energy, whereas collisional losses rise only logarithmically [25]. As a result, bremsstrahlung dominates at energies above a critical energy, generally of order 10 MeV in common materials. Above this energy, electron energy loss in a thick absorber proceeds via an electromagnetic (EM) cascade, comprising multiple pair production and bremsstrahlung processes. As well as increasing with energy, radiative losses increase with the atomic number of the target material. Protons/ions, meanwhile, show negligible bremsstrahlung losses at the energies of interest.

In contrast, other radiative processes, such as Cherenkov and transition radiation, cause negligible energy loss (e.g. Cherenkov represents $\sim 0.1\%$ of losses for strongly relativistic electrons [26]). These processes are, however, used in some radiation detection schemes, as described further in Section 2.4.4.

High energy projectiles may also participate in nuclear reactions with the target material. These are less probable than ionising interactions with the atomic electrons but can result in larger single-event energy deposits. An important additional result is the change in composition of the projectile flux due to fragmentation and secondary particle production.

Hadronic cascades may develop as a series of nuclear reactions in a thick absorber. These are characterised by the production of many pions (being the lightest hadrons) and other secondary particles. Neutral pions, comprising approximately one third of the yield, decay rapidly to gammas, initiating EM cascades (with shorter characteristic length than that of the hadronic cascade). Meanwhile the hadronic component also loses energy through ionisation and a variety of other processes. The threshold for such cascades is approximately twice the pion rest mass energy (~ 280 MeV) making them relevant only for GCRs and the highest energy trapped protons; moreover, most spacecraft are too thin for full cascade development.

Many of these complex phenomena cannot be described accurately with mean quantities or simple models and Monte Carlo methods are instead most appropriate. However, collisional (ionisation and excitation) energy losses may be characterised by the mean value per unit path length, as a function of particle energy, in a given material. The earliest accurate theoretical descriptions of this quantity, known as the linear stopping power or specific energy loss, were provided by Bethe [27]. Refined versions of this theory are still in use and provide accurate data across a wide range of energies. Different formulations exist for electrons and for heavier particles, due primarily to the difference in mass.

2.2.1 Electrons

The mean collisional energy loss rate, $-\langle \frac{dE}{dx} \rangle$, for an electron projectile with energy E is given by the Bethe mean stopping power formula [28]:

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{2\pi N Z e^4}{m v^2} \left\{ \ln \left[\frac{(\gamma + 1) E^2}{2 I^2} \right] - \left(\frac{2}{\gamma} - \frac{1}{\gamma^2} \right) \ln 2 + \frac{1}{\gamma^2} + \frac{1}{8} \left(1 - \frac{1}{\gamma} \right)^2 - \delta \right\}, \quad (2.1)$$

where m and v are the electron mass and velocity, γ is the Lorentz factor and N and Z are the target atom number density and atomic number. The mean excitation energy, I , must

generally be found through experiment rather than calculated directly. The correction term δ describes the density effect, where the polarisation of the medium lessens the electric field, due to the projectile, experienced by the target atoms [29]. This reduces the stopping power and is most significant (of order 10%) at relativistic energies.

Equation 2.1 is valid for electron projectile energies greater than the atomic binding energies (a few keV, depending on material). At high energies the total electron stopping power is progressively underestimated as radiative losses are not included. For example, radiation constitutes $\sim 1\%$ of energy loss for electrons at 700 keV in aluminium and $\sim 10\%$ at 7 MeV [30].

Integration of the mean energy loss rate to zero particle energy provides an estimate of the mean range of a particle, known as the continuous slowing down approximation (CSDA). Examples of the mean CSDA ranges for electrons in common materials are shown in Fig. 2.8. This, however, is the range along the particle trajectory rather than the forward range through a material. The ratio of these ranges, known as the detour factor, is always less than one due to multiple scattering; for example, a 1 MeV electron has a detour factor of ~ 0.5 in most materials.

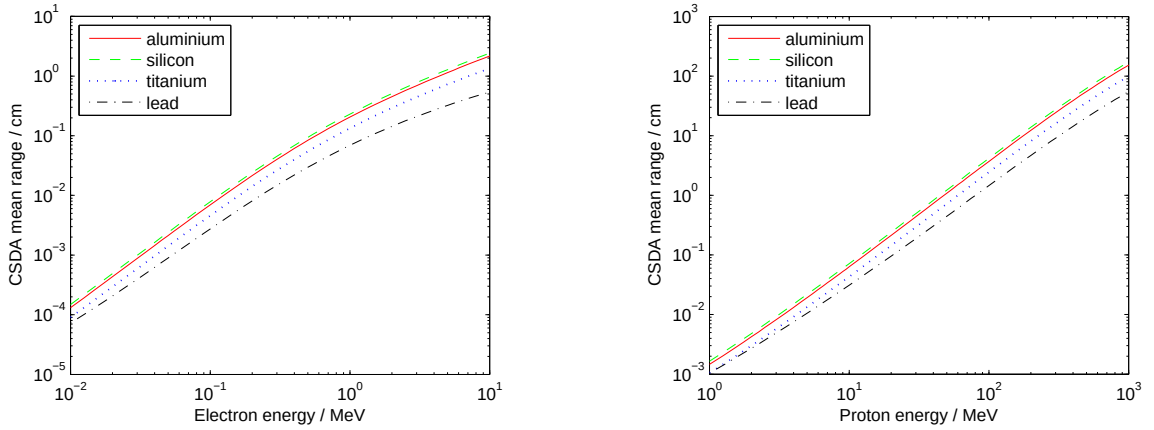


Figure 2.8: Mean CSDA ranges in several common materials for electrons (left) and protons (right).

2.2.2 Protons and ions

The mean specific energy loss for particles more massive than electrons is given by the Bethe-Bloch equation, one form of which is [25]:

$$-\left\langle \frac{dE}{dx} \right\rangle = \frac{e^4 z^2 Z N}{4\pi\epsilon_0^2 m c^2 \beta^2 \rho} \left[\frac{1}{2} \ln \frac{2mc^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta}{2} \right], \quad (2.2)$$

where βc and ze are the projectile velocity and charge and ρ is the target medium mass density. $T_{\max}$ is the maximum energy transfer possible in a single collision, which is dependent on the projectile mass and velocity; the other parameters are as previously described.

This formula accurately describes collisional losses for ions with velocities much greater than those of the orbital electrons of the target atoms ($E \gg 25$ keV/n). Several corrections have

been added to the original Bethe formulation to account for the density effect at high energy (as discussed for electrons) and the Barkas, Bloch and Shell corrections at low energies [31, 32, 33].

For heavy ions especially, charge exchange with the target medium makes application of this formula more problematic below ~ 1 MeV/n, and an effective charge as a function of velocity must be used. Above these energies, the ion may be considered to be fully stripped of electrons.

Multiple scattering is less significant for protons/ions than for electrons due to their greater mass and consequently detour factors are just $\sim 1\%$ at MeV/n energies. The CSDA range (examples shown in Fig. 2.8) is therefore a more useful quantity than for electrons. Radiative losses are also less significant, at the energies of interest, than for electrons (proton radiative losses reach 1% only at GeV energies).

While energy transfer to the atomic electrons dominates the total stopping power, a small but important loss is due to ion-nucleus collisions, known as the non-ionising energy loss (NIEL). The resulting nuclear recoil may lead to atomic displacement in a crystal lattice and material damage; this is discussed further in Section 2.3.

A common feature of Equations 2.1 and 2.2 is the increase in energy loss rate as the projectile velocity decreases, peaking shortly before the particle stops. The dose deposited by a monoenergetic beam at a given depth in a medium will depend on the summed losses of its constituent particles. For protons, and especially heavier ions, this results in a depth-dose curve with a sharp peak near the mean range (the *Bragg peak*). Electron beams, due to their greater straggling and scattering, do not exhibit a Bragg peak.

2.2.3 Electron and proton similarity

For particles of moderate energy, the momentum transfer between projectile and target atom is proportional to the time integral of the Coulomb interaction and thus approximately inversely proportional to the projectile velocity. The stopping power is therefore largely dependent on the velocity rather than the projectile mass (which only enters Equation 2.2 indirectly via $T_{\max}$). In addition, the stopping power depends on the magnitude, rather than the polarity, of the particle charge (z^2 dependence in Equation 2.2). As a consequence, electrons and protons, despite their dissimilar properties, show similar energy loss rates as a function of velocity. This is illustrated by Fig. 2.9 where the stopping power is shown as a function of $(\gamma - 1)$, which gives the kinetic energy as a multiple of the particle rest mass energy. Both curves show a broad minimum at $(\gamma - 1) \approx 2$, the region of minimum ionising particles (MIPs), where stopping powers are $\sim 1.5 \text{ MeV cm}^2 \text{ g}^{-1}$ for many common materials.

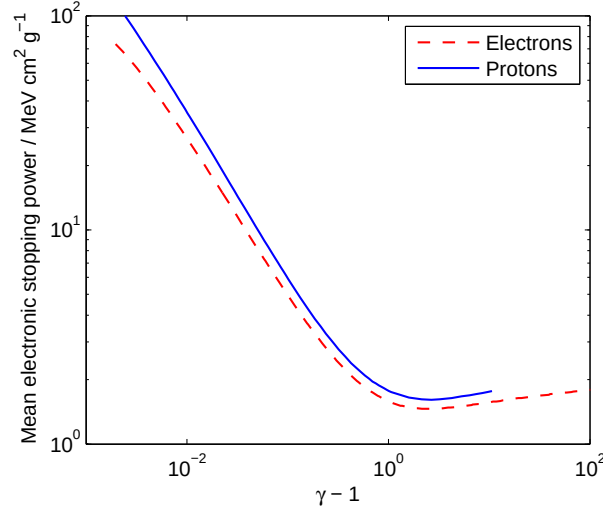


Figure 2.9: Electron and proton stopping powers in aluminium as a function of $(\gamma - 1)$. This quantity gives the particle kinetic energy as a multiple of the rest mass energy.

The small differences in proton and electron stopping powers which do occur are due to the effect on collision kinematics of the differing particle masses and also due to spin and identical particle effects. However, more important differences in the overall energy loss in material layers result from the contrasting multiple scattering behaviour (affecting path length) and fluctuations in dE/dx .

2.2.4 Energy loss fluctuation

The mean stopping power is often a satisfactory description in simple beam dosimetry applications. However, the energy loss process is fundamentally stochastic and is thus best described by a probability distribution; this is especially important when treating single particles. Energy loss fluctuation occurs due to both the probabilistic nature of the interactions themselves and the variance in the number of collisions occurring in a material layer. The latter effect has a greater relative importance for thinner layers.

The distribution of collisional energy transfers is asymmetric, featuring a long tail at high energies, representing rare events with large energy loss. These also prevent accurate calculation of the mean energy loss from a finite experimental dataset. As modelled by the Landau distribution [34] and later modifications [35], the energy loss is instead best quantified by the most probable value and the distribution width. An example of the Landau energy loss distribution for 250 MeV protons in a 1 mm silicon layer is given in Fig. 2.10; the most probable, mean and FWHM values are 330 keV, ~ 390 keV and 89 keV respectively. The Landau description is inadequate, however, for thin absorbers (e.g. thin silicon sensors) where the observed distribution width is greater than that predicted, due to the effects of the binding of atomic electrons [36].

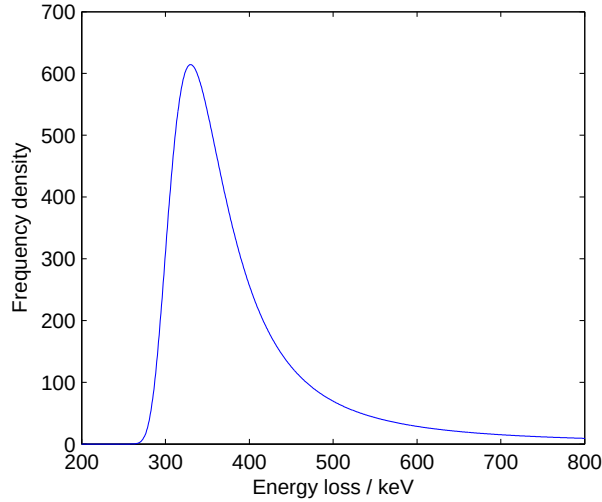


Figure 2.10: Landau energy loss distribution for 250 MeV protons at normal incidence on a 1 mm silicon layer.

Energy loss variance, together with angular scattering, contributes to the energy and range straggling observed for particle beams passing through matter. Mean stopping powers and CSDA approximations are therefore only suitable for the simplest dosimetry applications, while Monte Carlo simulations are required to obtain greater accuracy in realistic situations.

2.3 Radiation damage

While the deposition of energy by particles passing through matter often results in negligible changes to the macroscopic properties of the material, several processes can result in permanent damage. The effects of this damage may become apparent after cumulative exposure to particle fluences over a period of time, or may result from single particle events. When these materials are used in critical applications, such as spacecraft electronic components, the damage may result in subsystem performance degradation or even loss of the spacecraft.

Energetic particle fluences may cause chemical and structural changes to materials, such as colouration, embrittlement and gas evolution [37]. Surface charging and deep charging of dielectrics, and the eventual discharge, may also cause undesirable effects within the spacecraft. These aspects are, however, not discussed further here. The most significant threat to spacecraft from radiation comes from the effect on semiconductor components (principally silicon), as these form the basis of the on-board microelectronics, solar cells and specialised sensors such as imagers.

Radiation damage in semiconductors is often divided into two main categories: bulk and surface effects.

2.3.1 Semiconductor bulk damage

Coulomb or nuclear scattering of an incident projectile from nuclei in the bulk semiconductor constitutes a NIEL. Instead of atomic excitation or ionisation, atomic recoil and displacement from the crystal lattice may result. This depends on the displacement energy threshold which, for silicon, is of order 10 eV (depending on incidence direction with respect to the lattice) [38].

The ability of a particle to impart this energy is dependent on the collision kinematics, determined by the particle mass and energy. For example, a 100 keV electron can transfer a maximum of ~ 5 eV, while a 10 MeV deuteron may transfer ~ 1 MeV [38]. Note, however, that the *mean* value transferred is generally many orders of magnitude smaller. Heavy particles such as protons and ions are most damaging, with electrons contributing only at relativistic energies.

Energy losses below the displacement threshold are transferred to the lattice as phonon excitations. Above the threshold, the displaced atom moves to a metastable interstitial site, leaving a lattice vacancy behind. Known as a Frenkel (or point) defect, this is likely to remain for long time periods, affecting the electrical properties of the device. Energy transfers much greater than the displacement threshold (~ 1 keV) can produce an energetic primary knock-on atom. This has sufficient energy to cause further damage, producing a localised cluster of many defects.

Displacement defects introduce additional energy levels into the semiconductor band structure. Levels occurring mid-bandgap promote band transitions and recombination. Levels close to the band edges cause charge carrier trapping and delayed release, decreasing carrier lifetime and mobility. These levels can act in a similar way to donor/acceptor dopants and thus interfere with the intended semiconductor doping. Large irradiations have even been found to cause type inversion in high resistivity n-type silicon, producing p-type behaviour [39].

2.3.2 Semiconductor surface damage

While lattice displacement in the bulk semiconductor is due to non-ionising energy losses, other forms of damage may result from ionising losses. These may be transient effects, described further in Section 2.3.3, or more permanent changes classed as surface damage. Two examples are ionisation in surface oxide layers and the trapping of charge at interfaces. Since both of these effects are caused by ionising losses, all charged particles can contribute and the severity scales with the magnitude of the received dose.

Electron-hole pairs generated in silicon dioxide (requiring an average energy of ~ 18 eV [40]) may diffuse or drift to interfaces with doped silicon or metallic contacts, where some are removed. The remainder can accumulate at the interface, where lattice imperfections and dangling bonds provide trapping sites. This accumulation of charge affects the characteristics of the electronic component. A further possibility is the creation of parasitic current paths within the semiconductor device which may be self-sustaining if a positive feedback effect results.

2.3.3 Radiation damage effects in semiconductor components

Displacement and surface damage in semiconductors are typified by an increase in leakage currents, changes in carrier concentrations and a reduction in carrier lifetimes. A summary of the effects on three important classes of semiconductor device which find widespread application in space (solar cells, CCD image sensors and CMOS microelectronics) is presented below.

Solar cells

Solar cells comprise a large area (many cm^2), thin p-n junction covered by a thin transparent cover glass. Important figures of merit are the open-circuit voltage and closed-circuit current. Spacecraft solar cells are often operated near the point of peak power output and thus radiation-induced degradation may be a major determinant of mission life [8].

Aside from the darkening of cover glasses due to ionising losses (mitigated by careful material selection), displacement damage within the junction space charge region is most important. This causes an increase in the diode saturation current and a drop in open-circuit voltage. Due to the reliance on the diffusion of light-generated charge carriers, the cell efficiency is highly dependent on the minority carrier diffusion length and is therefore decreased by displacement damage.

Of special importance are low energy protons at $\sim 1.5\text{--}3$ MeV, which have mean ranges in silicon corresponding to solar cell junction depths. For example, 2 MeV protons may reach and stop at the junction depth in a typical 50 μm thick cell. This sensitive region thus coincides with the point of highest energy loss rate at the end of the particle's track (Bragg peak). As described in Section 2.1, protons at these energies are among the most numerous in MEO, making this an important hazard.

CCD image sensors

Charge coupled devices (CCDs) are widely used in scientific and space imaging applications due to their high resolution, quantum efficiencies and linearity. However, their reliance on the coupling of pixel charges across many thousands of pixels during readout makes them especially sensitive to decreased charge transfer efficiency (CTE). Bulk silicon damage is the greatest concern for CCD operation in space, with displacements leading to charge trapping, lower carrier mobility and reduced CTE.

An increase in dark current (and thus shot noise) and dark current spikes may also result. 'Random Telegraph Signal' noise occurs due to trapping and release of carriers at defect sites; together, the decreased signal (due to lower CTE) and increased noise lower the device sensitivity. 'Hot pixels' with abnormally large dark currents may also result from defect clustering.

Imagers are generally operated behind a greater depth of material than solar cells, raising the proton and ion energy threshold required for penetration to ~ 10 MeV/n. These

energies, which are particularly damaging, show significant flux enhancements during SEP events.

CMOS Microelectronics

The adoption of complementary metal-oxide-semiconductor (CMOS) technology in space is driven by its rapid development and cost-reduction by the commercial semiconductor industry [41]. In common with other process technologies, it is susceptible to both surface and bulk damage, with contributions from both ionising and non-ionising energy deposits.

Ionisation charge liberated by the passage of a high energy particle may collect on a charge storage node (e.g. in RAM) and be sufficient to change the binary state of the circuit. These reversible single event upsets (SEU), or ‘bit flips’, may be mitigated by triple-majority voting circuits (at the expense of increased complexity) [42]. Increased miniaturisation implies that less charge is needed to flip the bit (node sizes are reduced), although the cross-section for the particle hit is also reduced. Large irradiations are characterised by an increase in leakage current and power consumption, due mostly to surface effects.

A significant effect for CMOS is the shift in MOSFET threshold voltage (V_{th}) due to hole accumulation at the gate insulation layer [43]. Electron-hole pairs are formed in the silicon dioxide by ionising radiation losses; the more mobile electrons can escape into the gate circuit, while the holes become trapped at the silicon-silicon dioxide interface. The presence of this positive charge means that a more negative gate voltage is required to obtain the same effect as previously. Gradual annealing occurs due to electrons tunneling through the oxide and recombining with the holes: a thin oxide layer, as found in deep sub-micron processes, is therefore desirable. As well as being an unwanted effect in general CMOS circuitry, this change in V_{th} is exploited in the RadFET dose monitoring component (discussed further in Section 2.4.5).

Other effects of ionising radiation include single-event latchups, where a parasitic circuit is formed which may lead to large current flow and device degradation or destruction. This can occur in CMOS layouts due to formation of a parasitic bipolar transistor configuration which, together with the CMOS device, produces a pnpn thyristor structure [44]. A transient current pulse from ionising radiation initiates the latchup which then persists through positive feedback until power is removed.

The internal positioning of spacecraft electronics generally provides an omnidirectional shield of a few millimetres of aluminium due to spacecraft and payload structures. Although this implies a proton penetration threshold of order 10 MeV, and thus a decrease in proton primary flux, an increase in the flux of damaging secondary particles may result due to nuclear reactions and GCR ion fragmentation in this material.

2.3.4 Biological radiation damage

Exposure to ionising radiation results in adverse health effects which may be divided into two categories: deterministic and stochastic [25]. The former occurs when a threshold dose is

exceeded, with a severity that increases with dose, leading to cell and tissue damage (e.g. skin erythema and cataract formation). Stochastic effects occur randomly with a probability that depends on the dose received. The severity of illness, however, does not scale with dose and a threshold may not exist. The main example is cancer, the average induction of which is estimated at $\sim 5\%$ per Sv [45]. Lethal ionising doses (a 50% probability of death within 30 days without medical treatment) are estimated as $\sim 2.5\text{--}4.5$ Gy measured internally on the body centre line.

Annual dose equivalent exposure on Earth is 2.4 mSv (mean, whole-body value [46]), of which less than 0.9 mSv is due to GCRs and their secondaries. The recommended mean annual limit for radiation workers is 20 mSv, averaged over a five year period [45].

Human exposure to space radiation, while rare, is a relevant concern due to the continued use of the International Space Station (ISS) in low Earth orbit (approximately 400 km altitude, 52° inclination), and the possibility of future space exploration. Most space flight (in the past and predicted in the near future) is in LEO and trapped radiation is therefore the greatest hazard.

The ISS orbit, according to the AP-8/AE-8 models, experiences electron and proton fluxes of approximately 10% of those of reference orbit A. Proton fluxes are largely confined to the SAA, which is crossed for a period of ~ 5 minutes (above 50% of peak) in a third of orbits (each orbit lasts 93 minutes). At the peak, the proton flux above 100 MeV is of order $10^2 \text{ cm}^{-2} \text{ s}^{-1}$ while the electron flux above 1 MeV is of order $10^6 \text{ cm}^{-2} \text{ s}^{-1}$. The SAA is estimated to account for $\sim 50\%$ of absorbed dose in the ISS [47].

SEPs are generally less important due to increased geomagnetic shielding, although the largest events are significant. For example, the October 1989 SEP (estimated to be a once-per-decade event) caused the measured dose rate in the Mir space station (also at ~ 400 km) to double [48].

Ionising dose rates in the Columbus module of the ISS have been estimated via simulation to be $131 \pm 10 \text{ } \mu\text{Gy day}^{-1}$ [49] and were measured in the ISS US Lab to be in the range $153\text{--}231 \text{ } \mu\text{Gy day}^{-1}$ [50]. Measurements on Mir using a thermoluminescent dosimeter (TLD) gave a long-term mean dose rate of $\sim 162\text{--}508 \text{ } \mu\text{Gy day}^{-1}$ [51]. This was found to vary depending on orbital altitude and position of the dosimeter within the spacecraft. Some of the highest dose rates experienced in Earth orbit occurred during the high altitude Hubble Space Telescope servicing missions with the Space Shuttle, where dose rates often exceeded $2000 \text{ } \mu\text{Gy day}^{-1}$ [48].

The walls of habitable spacecraft have typically $\sim 5 \text{ g cm}^{-2}$ of material, mostly aluminium, giving proton penetration thresholds of 50–70 MeV [52]. While shielding much of the trapped radiation, this provides little attenuation of GCRs, which are estimated to account for $\sim 10\%$ of dose deposited inside the ISS. However, as well as providing shielding, this material may ‘convert’ incident radiation to forms with higher relative biological effectiveness (RBE). Of particular concern is the fragmentation of GCR projectiles and target material, producing knock-on and spallation nucleons, recoil nuclei and high-energy heavy fragments.

The RBE of different particles over different energy ranges remains uncertain and the scaling

of risk with exposure is not fully understood. However, energetic heavy ions are known to be especially damaging due to their large energy loss rates. Biomolecular models indicate that these particles also cause more complex DNA breakage damage and thus a greater risk of cancer induction [53]. Other health effects include persistent inflammation, oxidative damage and acceleration of age-related illnesses.

The considerable risks presented by radiation to devices and people in space mean that radiation monitors are an important addition to any spacecraft.

2.4 Space radiation monitors

A radiation monitor is here defined as an instrument consisting of one or more radiation sensors in an arrangement designed to provide information regarding incident particles. The sensors contain, or are coupled to, a sensitive volume which allows the production of an electrical signal in response to a single incident particle. This usually involves measurement of the particle's ionising energy loss by detecting the generated charge carriers or photons. Non-electronic methods (for example, thermoluminescent dosimeters and photographic emulsions) are of limited use compared to electrical transducers which allow convenient processing, analysis and transmission of data.

Single particle response is a further important requirement, in contrast to long-term integrated dosimetry of a particle flux. Measurement of individual particles introduces the possibility of particle identification (discrimination of species and energy) and thus a more complete characterisation of the radiation environment. For this to be feasible, the sensor response and reset time must be shorter than the time between particle hits and the output signal should vary depending on the particle mass, energy or charge, or a combination of these parameters.

Particle identification ability (rather than simple counting) depends both on the limitations of the sensor technology and on the overall identification 'scheme'. This scheme encompasses the sensor arrangement, monitor geometry, materials selection and any applied electromagnetic fields, together with the analysis algorithm used to convert sensor data into an interpretation of the radiation environment.

The use of a radiation monitor in space imposes additional constraints on the design due to the practical limitations on mass, volume, power and data rate. These considerations are particularly severe for monitors intended as support instruments (rather than as a scientific payload) since their peripheral mission role is reflected in smaller shares of the subsystems budgets. Naturally, further constraints apply to the financial cost implied in the procurement, installation and exploitation of these devices.

As well as measuring particle ionising energy losses, the sensors themselves are subject to the ionising and non-ionising damage mechanisms described in Section 2.3. The monitor must therefore be sufficiently robust to survive in the space environment. Additional concerns are reliability, automated operation, thermal regulation in vacuum, and survival of launch vibrations.

A wide variety of radiation sensors are suitable for use in space, including semiconductor devices, scintillators and gas-filled detectors. These may be used, with certain restrictions, in many identification schemes involving: selective material shielding, sensor coincidence, electromagnetic fields, time-of-flight measurements, Cherenkov emission and transition radiation. Several effective sensors and identification methods are discussed in the following sections through a review of devices in past and present use. These include both small-scale monitors for satellite support functions and larger scientific instruments. This list is not exhaustive but attempts to cover a wide range of device sizes and technologies.

2.4.1 The Geiger-Mueller tube

The Geiger-Mueller (GM) tube is a gas-filled detector which allows the counting of incident ionising radiation. The Geiger avalanche discharge, triggered by the initial ionisation event, causes a saturated response — no properties of the particle or interaction may be deduced since all responses are equal.

A GM tube was the first radiation detector used in space, on the Explorer-1 satellite in 1958, and provided the first direct evidence of the trapped radiation belts [54, 55]. This instrument had an insufficient dynamic range so that the measured count rate changed from zero to maximum (128 s^{-1}) almost instantly after entering the belts. GM and other gas-filled counters have since been superseded by more compact sensors in most space applications.

2.4.2 Scintillators

A number of organic and inorganic materials are known to fluoresce in response to ionising radiation. The light yield (typically at wavelengths of 300–500 nm) is dependent on the particle’s specific energy loss and thus also on the particle species and kinetic energy.

Although the primary sensitive volume is considered to be the scintillator itself, the light produced must in turn be detected by a sensor optically coupled to the scintillator. This must be sufficiently sensitive to detect low light levels (typically of order 10^3 photons per MeV deposited), generally indicating the use of photomultiplier tubes (PMTs) in ground-based applications. PMTs are, however, bulky and fragile and require high voltage supplies, making them less suitable in space. Modern semiconductor devices (e.g. avalanche photodiodes or silicon photomultipliers) may be substituted although the very high gain of traditional PMTs has not yet been equalled in this way.

The Scintillating Fibre Detector (SFD)

The SFD, launched on the EQUATOR-S magnetospheric science satellite in 1997 [56], comprises two C_8H_8 -doped organic scintillator fibres of 1 mm diameter coupled to GaAsP photodiodes. The current outputs (non-pulsed detection) are detected with a logarithmic current amplifier and analogue to digital converter (ADC), with a current noise of less than 300 e s^{-1} . The two digitised currents are converted to an equivalent dose rate or mean particle flux in software.

Aluminium shields on the two fibres of 0.2 mm and 4 mm thickness give respective low energy thresholds of 0.4 MeV and 2 MeV for electrons and 8.6 MeV and 30 MeV for protons. Linear dose rate response has been demonstrated to greater than 17 mGy s⁻¹ and for protons at energies in excess of 100 MeV. The total dose limit before scintillator damage is estimated to be greater than 10 kGy. The instrument's mass, volume and power are relatively small, at 397 g, 332 cm³ and 105 mW, respectively. Comparison of the two channel rates allows qualitative assessment of the particle environment, while multiple scintillators may be monitored remotely via optical fibres. However, the lack of single-particle pulse measurement is a severe limitation.

2.4.3 Electrostatic analysers

In the simplest form, electrostatic analysers (EAs) such as the top-hat or cylindrical varieties comprise concentric, circularly curved surfaces held at a potential difference of order 10³ V. Incident charged particles follow the curvature and reach the sensor only if the energy-to-charge ratio (E/q) lies in a small numerical range. This may be augmented by a second stage containing a time-of-flight (TOF) detector, allowing measurement of the particle velocity and thus calculation of the mass-to-charge ratio (m/q). The selected particle E/q is approximately dependent on the applied voltage, V , according to [57, 58]:

$$\frac{E}{q} = \frac{V}{2 \ln(r_{\text{out}}/r_{\text{in}})}, \quad (2.3)$$

where r_{out} and r_{in} are the radii of curvature of the outer and inner concentric shells. For example, a typical analyser with $r_{\text{out}}/r_{\text{in}} = 1.11$ implies a detectable kinetic energy for protons of 4.7 keV per kilovolt applied. This sets a practical energy limit of order 10 keV due to the excessively high voltages required.

Miniature Ion Precipitation Analyser (MIPA)

Despite the complexity of EA instruments, substantial miniaturisation is possible. For example, the MIPA [58] is capable of resolving $m/q = 1, 2, 4, 8, 16, >30$ for ions in the 10 eV–15 keV range. It employs a 1.9 cm radius cylindrical EA at potentials of up to 3.5 kV, with a TOF detector and microchannel plate (MCP) sensor, achieving an energy resolution of 7%. This is accomplished in an overall volume of 214 cm³, 350 g mass and 1.5 W power consumption.

ASPERA-4

A second example is the ASPERA-4 [59] which contains separate EA detector heads for electrons and ions, together with neutral atom detectors. The 1.5 cm radius hemispherical top-hat EA for electrons uses an MCP sensor and covers the energy range 0.01–20 keV with 7% resolution, using potentials up to 2.8 kV. The 360° azimuthal field of view is divided into 16 detection sectors, allowing measurement of flux anisotropy. This instrument is 300 g in mass and has ~111 cm³ volume.

The ion mass analyser, a separate detector in the same monitor, is a spherical top-hat design covering $E/q = 0.01\text{--}36$ keV/e for $m/q = 1, 2, 4, 8, 16, >40$. This instrument also uses an MCP sensor but employs a magnetic separator, rather than a TOF detector, for velocity measurement. The mass and volume are 2.4 kg and 8855 cm³, respectively, of which the magnet constitutes $\sim 25\%$.

Although these EAs achieve excellent mass and energy resolution, their low energy range limits them to space plasma and solar wind measurements. Furthermore, the considerable miniaturisation that has been achieved is unlikely to continue below ~ 100 cm³, while the complexity is acceptable only for a specialised scientific payload.

2.4.4 Cherenkov emission and transition radiation detectors

Optical Cherenkov radiation is produced when a charged particle exceeds the phase velocity of light in a medium. Photons are emitted at an angle θ_c to the projectile direction which depends upon its velocity, βc , and the refractive index of the medium, n , according to [25]:

$$\tan \theta_c = \sqrt{\beta^2 n^2 - 1}. \quad (2.4)$$

This effect, combined with the dispersive properties of the medium, results in a conical shell of emitted light, with the projectile at the vertex. This cone appears as a ring of light when projected onto a surface; a ring-imaging Cherenkov (RICH) detector measures the ring size using an array of light sensors, allowing the cone angle (and hence particle velocity) to be determined.

Transition radiation occurs when an energetic charged particle crosses between two media of different optical properties. When one medium is a vacuum, the total energy radiated by a particle of charge ze is [25]:

$$I = (9.60 \text{ eV}) \times \alpha z^2 \gamma \sqrt{\rho \langle Z/A \rangle}, \quad (2.5)$$

where α is the fine structure constant, ρ is the material density (in g cm⁻³) and $\langle Z/A \rangle$ is the mean ratio of its atomic number and mass. Due to this small radiation yield, transition radiation detectors (TRDs) use multiple material layers to provide many interfaces.

Cherenkov and transition radiation are relevant only for highly relativistic particles, meaning that their use in space is limited to GCR detection.

The Alpha Magnetic Spectrometer (AMS)

The AMS, installed on the ISS, uses a variety of detection methods for high energy particle identification [60]. In addition to a RICH detector, AMS contains a TOF detector, a silicon tracker, veto counters, a TRD and a sampling electromagnetic calorimeter. It can provide accurate measurement of electrons, positrons, protons and ions in the ~ 100 MeV to $\sim$ TeV energy range and ion charge measurements up to $Z = 26$. A rigidity resolution of 2% up to 20 GV and 6% charge resolution at unit charge are achieved.

The RICH uses two radiator media: a low-refractive index aerogel and sodium fluoride, giving radiator thresholds of 3 GeV and 0.5 GeV, and velocity measurement resolutions (for unit charged particles) of $\sim 0.1\%$ and $\sim 1\%$, respectively. A plane of multipixelised PMTs provides 10,880 separate sensors on an effective 8.5 mm pitch grid, surrounded by a conical reflector to reduce lateral losses. The TOF detector is used for velocity measurement below the radiator thresholds.

The TRD comprises 20 layers of LRP375 fibre fleece radiator interleaved with 5248 gas proportional tubes filled with a 90:10 Xe:CO₂ mixture. The output signals are used in a log-likelihood estimate to distinguish electrons/positrons from protons. Particle charge measurement is performed via multiple dE/dx measurements in the nine tracker layers (2500 double-sided silicon sensors), which also provide $\sim 10\text{ }\mu\text{m}$ position resolution, over a distance of 1 m, in the 0.14 T field of a permanent magnet. On-board analysis using 650 microprocessors reduces the data rate by a factor of ~ 1000 , resulting in a mean telemetry rate of 10 Mbit s⁻¹.

AMS provides a comprehensive and accurate measurement of the high energy particle flux in the orbit of the ISS; this includes a precise extraction of the mass, charge, energy and incidence direction of single particles. Its large acceptance ($\sim 0.5\text{ m}^2\text{ sr}$) has allowed the rapid collection of a significant quantity of cosmic ray data [61]. However, these advantages come necessarily at the expense of substantial complexity, physical size and financial cost.

2.4.5 Semiconductor sensors

While AMS employs silicon sensors (in the tracker) together with other detectors, there exist many radiation monitors which rely solely on semiconductor sensor devices. Of particular interest are those where the semiconductor material itself (most commonly silicon) constitutes the sensitive volume. This contrasts with indirect situations where, for example, a semiconductor photodiode detects light from interactions in a scintillator.

Ionising energy losses in the semiconductor result in the excitation of electrons from the valence band to the conduction band, forming electron-hole (e-h) pairs. Collection of these charge carriers creates an electrical signal from which the deposited energy may be deduced. For an energy deposit E_{dep} significantly greater than 10 eV, the mean number of e-h pairs generated, N , is given by:

$$N = \frac{E_{dep}}{W}, \quad (2.6)$$

where W is the mean energy required for the creation of one e-h pair which, for silicon, is approximately $W = 3.65\text{ eV}$ at room temperature [62]. Due to correlations in the charge generation process, the variance in the number of pairs produced is less than that predicted by a Poisson model. Instead, the variance σ_F^2 is usually characterised with the following empirical formula:

$$\sigma_F^2 = FN \quad (2.7)$$

$$= F \frac{E_{dep}}{W}, \quad (2.8)$$

where F is the Fano factor [63], originally formulated to describe ionisation in gases.

The number of charge (or information) carriers generated per unit energy deposited, together with the Fano factor, provides a fundamental energy resolution limit for the sensor. The number generated in silicon is more than 10^5 e-h pairs per MeV at room temperature, significantly greater than the values for scintillators ($\sim 10^3$ – 10^4 per MeV) and gas proportional ionisation ($\sim 10^4$ e-ion pairs per MeV). Measured Fano factors are in the range 0.05–0.2 for gases, approximately 0.1 for silicon and ~ 1 for scintillators [39].

Practical device resolutions are, however, generally limited by other noise sources, such as thermal or leakage current shot noise. Further reasons for the popularity of silicon sensors are their typically compact size, robustness and low power consumption. The following sections discuss various silicon sensor technologies and their use in space radiation monitors.

RadFET sensors

Section 2.3 introduced the phenomenon of threshold voltage modification in MOSFETs due to exposure to ionising radiation. This is exploited in the RadFET component which is commonly used for dose rate and total dose monitoring (for example, see [64, 65]). Although single particle detection is not possible, multiple RadFETs behind different depths of shielding may be used as an approximate indication of radiation penetration. While the small size and simplicity of the component and readout circuit are attractive, care must be taken to account for temperature variations which may also affect V_{th} [66].

PIN diodes

The ability to resolve individual particles is a desirable capability for any radiation monitor and is essential for particle identification. This is most easily achieved in silicon sensors with a simple p-n junction under reverse bias. An ionising energy loss within the depletion region generates e-h pairs which are separated under the influence of the electric field and collected at the electrodes. The resulting current pulse is usually measured via a combination of charge-sensitive preamplifier, pulse shaper and digitiser. This ‘pulse height analysis’ allows the total ionising energy deposit of the event to be recorded.

A variation of the standard p-n junction diode may be created by the inclusion of a central compensated region, where n and p-type impurities are balanced. Due to the similarities between compensated and intrinsic (i) silicon, this configuration is known as a p-i-n (or PIN) diode. The central compensation (with higher resistivity) allows a larger electric field to be applied without causing an excessive leakage current or device breakdown. This leads to a large depletion region (sensitive volume) upon application of a reverse bias of tens to hundreds of volts. PIN diodes are typically 100–500 μm thick and have surface areas of $\sim 1\text{ cm}^2$. The sensor capacitance determines the magnitude of the equivalent noise charge observed at the output; in analogy to a parallel-plate capacitor, this is proportional to the surface area and inversely proportional to the depletion region thickness (charge separation distance). The readout circuit introduces additional noise, particularly due to thermal noise in the amplifier’s input transistor.

PIN diodes are one of the most popular sensors for space radiation environment monitoring; the following sections discuss several instruments which use these devices.

The Standard Radiation Environment Monitor (SREM)

The SREM [67] uses pulse height analysis from three PIN diodes to count and classify incident particles and is intended to alert the spacecraft and payload to space radiation hazards. One diode of 25 mm² area is mounted behind a 0.7 mm aluminium window, giving particle energy thresholds of 0.5 MeV and 10 MeV for electrons and protons, respectively. Four pulse height channels are allocated in the range 0.085–2 MeV, with two identified as protons and two as electrons.

Separately, two additional diodes are arranged in a telescope configuration, the front diode having a 25 mm² area and the rear diode a 50 mm² area. A front window of 2 mm aluminium introduces energy thresholds of 1.5 MeV and 20 MeV for electrons and protons, respectively. An energy ‘degrader’ of 1.7 mm aluminium and 0.7 mm tantalum between the diodes imposes a ~ 39 MeV proton coincidence energy threshold, while electron coincident hits are assumed to be negligible. One ion and six proton pulse height channels are assigned to the front diode alone, with a further four coincident proton channels. All pulse height channel discriminator levels are in the range 0.085–9 MeV.

The analysis electronics can handle event rates up to 100 kHz, while an internal RadFET is also supplied for integrated dose measurements. The overall monitor has a volume of 2400 cm³, mass of 2.6 kg and consumes ~ 2.5 W.

Four SREM units operating in space have demonstrated reliable operation, with good spatial resolution of the radiation belts [68] and detection of flux anisotropies [69]. However, the assertion that pulse-height channels identify single particle species is probabilistic only and significant species ‘contamination’ has been observed. Methods for deconvolution of species counts and unfolding of the incident energy spectra have also been investigated. The small number (15) and large width of the pulse height bins, however, mean that these operations must have limited spectral resolution.

The LISA Pathfinder Radiation Monitor (LPRM)

The planned LISA Pathfinder technology demonstration mission [70] will be sensitive to the electrostatic charging of its isolated test masses. This is due to the interactions of energetic charged particles and has the effect of introducing noise into the frequency band used for gravitational wave interferometry. The LPRM was designed specifically to detect protons with energies greater than ~ 100 MeV which have sufficient penetration to contribute to this charging [71].

A pair of 147 mm² area, 320 μ m thick silicon PIN diodes are used in a telescope configuration with 2 cm separation. This is surrounded by a 6.4 mm thick copper shield to produce a particle energy threshold of approximately 70 MeV/n. The diodes are operated in coincidence with 1024 pulse height bins between 50 keV and 5 MeV and have an equivalent noise charge

of $\sim 5600\text{ e}^-$. This diode thickness produces a most probable signal for a MIP at normal incidence of approximately 25000 e^- .

Coincidence operation constrains the range of valid particle incidence angles and thus reduces the energy deposit variance that may result from different path lengths in the diodes. The chosen diode size and separation still allow variations of order 10% ($\sim 2500\text{ e}^-$ for a MIP); greater separation would constrain this further, at the expense of a lower count rate and a larger instrument. However, more important than the path length variation, and of a similar magnitude to the readout noise, is the fundamental energy loss fluctuation within the sensor. For a normal incidence MIP, this is $\sim 8200\text{ e}^-$ (FWHM) and is thus likely to become the main limitation for particle identification in sensors with improved noise performance.

This monitor is an efficient design optimised for a specific purpose; material shielding is used to define low energy acceptance limits, while coincidence requirements limit the path length fluctuation and reduce the probability of false noise counts. Dose to silicon can be measured accurately from the high resolution pulse height histogramming.

Gravity Probe B (GPB) High Energy Proton Monitor (EPM)

The EPM [72] was designed to detect $\sim 35\text{--}500\text{ MeV}$ protons which may cause charging of the precision gyroscopes used in the GPB mission. This monitor comprises a stack of four silicon PIN diodes within a 2 mm thick tantalum casing, while the 45° aperture is covered by a 3 mm thick titanium hemispherical dome. The simplified monitor geometry is shown in Fig. 2.11, illustrating three possible particle paths in the sensors.

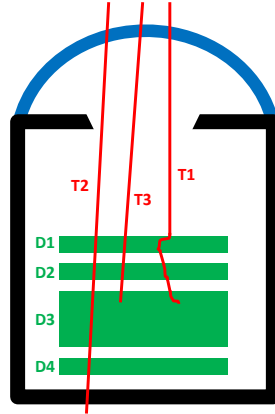


Figure 2.11: Simplified diagram of the particle detection region of the EPM. The tantalum casing, titanium dome, and four pin diodes (D1–D4) are included. Also shown are the tracks of three particles, illustrating expected behaviour of an electron (T1), high energy proton (T2) and low energy proton (T3).

The dome shield and outer casing provide proton penetration energy thresholds of approximately 35 MeV and 400 MeV, respectively. Diodes D1, D2 and D4 are of $150\text{ }\mu\text{m}$ thickness, while D3 is $700\text{ }\mu\text{m}$ thick. All the diodes are read out by charge sensitive amplifiers and pulses from D3 are histogrammed into 256 channels, of which 16 may be selected for data acquisition.

Differential amplification of coincident pulses from D1 and D2 allows the discrimination, and rejection, of events with significantly *differing* deposits in these two diodes. This behaviour is associated with electrons, which show greater multiple scattering and thus greater path length deviation: this is illustrated by track T1 in Fig. 2.11. Events with D1 and D2 deposit difference below the threshold are interpreted as protons and the deposit in D3 is measured. In this case, a coincident hit is also demanded in D4, indicating that the particle left D3: an example is shown as track T2 in Fig. 2.11. Without this constraint, the deposit due to a lower energy proton stopping at shallow depth in D3 (track T3) could equal that of a higher energy particle that passes through the sensor.

The overall instrument has mass, volume and power consumption of 3.5 kg, 2363 cm³ and 3.5 W, respectively.

Energy loss sampling

The SREM, LPRM and EPM use similar methods to detect and identify particles. These include:

1. Metallic aperture shielding to define particle detection energy thresholds and exclude lower energies.
2. PIN diode pulse height analysis for energy deposit measurement and estimation of the particle specific energy loss in silicon.
3. Coincidence measurements, using a sensor telescope, to constrain the angular acceptance of particle trajectories and to measure the particle penetration depth.

As described in Section 2.2.3, different particle species with equal velocities may have similar specific energy loss and thus leave equal energy deposits in a single thin sensor (where the change in velocity of the particle is negligible). However, an equal energy loss rate, $-dE/dx$, implies an unequal velocity loss rate, $-dv/dx$, for particles of different mass. After passing through further material of significant stopping power, these particles will therefore have different velocities and thus different specific energy loss in a second sensor. Multiple sampling of a single particle's dE/dx should therefore allow its identity to be deduced.

The resolution of this measurement method is limited by multiple scattering and energy loss fluctuations. Additional problems include the difficulty of slowing MIPs (requiring thick degraders) and the loss of particles from the telescope between sensors due to large-angle deflection or oblique incidence angles. These loss modes are generally indistinguishable from the case where a lower energy particle stops within the telescope. Identification of these confounding effects is only possible with a detailed tracker detector, such as that implemented in AMS, with multiple sensor layers and transverse position reconstruction (which are further enhanced by a magnetic field).

Particle identifications based on dE/dx measurements in small monitors are therefore necessarily limited to probabilistic interpretations. Despite this, the method remains popular as it may be implemented with a small mass and volume appropriate for an engineering support instrument. By suitable design of the sensor telescope (number of sensors, thicknesses, spac-

ing, inter-sensor degraders, etc.), coincident pulse height channels may be optimised for high purity of certain species/energies, although deterministic interpretations remain generally inaccurate.

Furthermore, a statistical deconvolution of the mean incident particle spectra may be possible if sufficiently detailed information is collected. Meanwhile, reliable dose and count rate measurements may still be obtained with each sensor, independently of any identification attempts.

Pixel sensors

One of the disadvantages of the PIN diode is the generally large equivalent noise charge (often $\sim 1000\text{ e}^-$) obtained in the energy deposit measurement. This is due to the relatively large diode capacitance, large leakage current and the use of discrete readout circuit components with limited noise performance.

A reduction in noise may be obtained by decreasing the size (and hence capacitance) of the sensor. To maintain a useful detection volume this implies the use of pixel arrays, where each pixel is a separate charge collection node on a single die. This configuration has other advantages: for example, allowing imaging (position reconstruction) and windowing (variation of the total sensitive area).

A number of pixel sensor architectures are available; three common varieties are CCDs, hybrids and CMOS monolithic active pixel sensors (MAPS).

1. **CCDs** are known to be effective at collecting ionisation charge from particle interactions [73] and have been investigated for use in space radiation monitors [74]. However, such monitors have not been adopted in practice; this may be due to CCD radiation damage sensitivity.

A further problem is that most devices available commercially have a large number of pixels ($>100,000$) making them unsuitable for applications with low bandwidth or limited processing ability. This problem is compounded by the relative complexity of using CCDs, which must be supplied with several voltages and clock signals. Moreover, small volume production of application specific CCDs is not generally economic due to the specialised process requirements.

2. **Hybrid pixel sensors** consist of a sensor chip bump-bonded onto the surface of a separate ASIC containing the readout electronics. This hybrid construction allows the two chips to be constructed of different materials, each optimised for their different tasks. It also allows a high pixel density over large areas with significant analysis circuitry per pixel.

The choice of sensor chip material may be tailored to the intended radiation environment (e.g. Cd(Zn)Te for X-ray detection), while the readout chip may be manufactured in processes intended for very large scale integrated electronics. Bonding of the two chips, using a flip-chip technique with solder or indium bump bonds, is a complex and labour-intensive process [42].

The Medipix2 [75] and Timepix [76] hybrid sensors were developed primarily for medical imaging but are also under investigation for other uses. Several Timepix sensors were deployed in October 2012 on the ISS to explore their use in single-chip personal space-flight dosimeters. These use a 300 μm thick silicon sensor chip with 2 cm^2 area, bump bonded to a readout chip containing a Wilkinson ADC for each of the 65536 pixels. Particle identification is achieved through pixel cluster shape analysis, although this requires the use of a personal computer [77]. Five Timepix sensors are also employed in the LUCID space radiation detector which is planned for launch on the TechDemoSat-1 spacecraft in 2013 [78].

3. **CMOS MAPS** are in development as an alternative to hybrid sensors in several applications, having already found widespread use for visible light imaging [79]. These comprise a low resistivity p-doped silicon substrate, with n-wells implanted in a thin (8–20 μm) p-type epitaxial layer to form collection diodes. Charge carriers are collected via diffusion from interactions in the epitaxial layer, rather than in the small depletion regions surrounding the n-wells. The use of a CMOS fabrication process allows the signal analysis circuitry to be included on the same chip, and within each pixel, resulting in a compact, rugged device with potentially much lower noise. This also means that application specific chips may be manufactured at lower cost.

Radiation-hardened CMOS MAPS have been developed for use in space as imagers [80] and star trackers [81]. These use enclosed geometry transistors and guard rings to reduce radiation-induced leakage currents, giving operational doses up to ~ 300 kGy. Unlike CCDs, MAPS do not require such complex input waveforms to drive the readout process. They are also potentially less sensitive to radiation damage as they do not rely on a high CTE (each pixel is read independently).

In addition to substantial development efforts for use in particle physics (e.g. EM calorimetry and tracking [82]), MAPS are being investigated for space radiation monitoring [83] although this remains a highly novel application.

2.4.6 Implications for the design of the HMRM

The reviews of the Earth radiation environment and of existing monitors have provided guidance for the design of the HMRM. The following features, in addition to dosimetry measurements, are identified as being especially desirable:

1. Sufficient miniaturisation to allow integration of one or more units onto a wide range of satellites with minimal impact on systems budgets.
2. A low energy proton detection threshold of less than ~ 2 MeV to allow coverage of the important energy range up to 10 MeV, which is neglected by many other monitors.
3. Some degree of single particle identification should be made. This is not expected to give the performance of a scientific instrument but should be sufficient to allow a probabilistic assessment of the environment and its potential for spacecraft damage.

The selection of the detection scheme and sensor technology, together with a detailed explanation of the monitor geometry design, is described fully in the next chapter.

2.5 Summary

The radiation environment in Earth orbit comprises electrons, protons and ions with energies covering many orders of magnitude. Total omnidirectional fluxes in commercially important orbits may exceed $10^7 \text{ cm}^{-2} \text{ s}^{-1}$, with energies up to $\sim 6 \text{ MeV}$ and $\sim 500 \text{ MeV}$ for electrons and protons respectively. These fluxes are dominated by particles trapped in the Earth's magnetosphere, with transient increases at higher energies due to solar events. The particles lose energy in matter primarily by ionisation and excitation of atoms; rarer non-ionising energy losses can cause damaging atomic displacement, which is more probable for proton and ion primaries. Particularly at risk are spacecraft semiconductor electronics, including solar cells, image sensors, processors and memory.

Earth orbit particle distribution models should be interpreted as approximate and indicative only, since the environment is continually evolving and because they are based upon limited experimental measurements. The widespread use of radiation monitors on spacecraft is therefore desirable for improved mission planning, evaluation of spacecraft shielding effectiveness and (in some applications) real-time alerting.

Device budgets for support instruments are usually restricted and measurement accuracy is generally compromised to some extent when compared to scientific instruments. Simple monitors often offer only dosimetry measurements: this is, however, only a crude indication of the damage capability of the radiation, which depends greatly on particle species and energy. Some degree of single particle identification, or longer-term energy spectrum reconstruction, is therefore desirable.

Finally, due to the variation of particle fluxes around large spacecraft, and flux anisotropy, the measured flux may be highly dependent on monitor positioning. This can only be addressed fully by providing multiple monitors at important locations on a single spacecraft. In this application a miniaturised monitor is clearly required to avoid imposing an excessive burden on the spacecraft.

Chapter 3

Monitor geometry design and optimisation

In this chapter the formal HMRM device requirements are introduced and the design methodology for this instrument is explained. The optimisation of geometry elements, such as sensors, aperture and casing, resulted in a set of recommendations which guided the prototype design. This is presented together with simulation results intended to evaluate its basic properties. Analysis algorithms, for execution within the monitor and ground segment, are addressed separately in Chapters 4 and 6.

3.1 Monitor design methodology

The most important decisions to be made within the design process were the choices of radiation sensor and particle identification scheme, in which one or more sensors should be employed — several examples relevant for space applications were reviewed in Chapter 2. The identification scheme was chosen with reference to the device requirements and by considering the particle energies found in Earth orbit. Following this a simple conceptual design was created, defining the general architecture of the monitor and allowing a suitable sensor technology to be selected.

The monitor development was divided subsequently between two main aspects: the detector geometry (comprising the sensing elements, aperture and surrounding structures) and the support geometry (comprising the ancillary electronics, interfaces, etc.). Various parameters of the detector geometry were identified which, together with a selection of figures of merit and design principles, were used to guide the geometry optimisation.

A series of incremental changes to the baseline conceptual design were made; each of these was evaluated by simulating simplified models with the Geant4 Monte Carlo toolkit [84, 85]. This allowed a number of design recommendations to be issued [86] for possible inclusion in the final monitor design, subject to further practical constraints. Most of these were successfully incorporated into the monitor prototype, which was then simulated in full to assess its suitability in the reference orbits identified in Chapter 2.

3.1.1 Device requirements

The ESA statement of work [87] for the HMRM project specifies the following functional and physical requirements for the monitoring device:

1. The volume shall be of order 1 cm^3 .
2. The mass shall be approximately 20 g.
3. The device shall measure the accumulated dose due to charged particles.
4. The device shall discriminate different parts of the LET spectrum, thereby giving an estimate of the primary particle energy. The dynamic range should be chosen to maximise discrimination over the primary particle energy ranges found in typical satellite Earth orbits.
5. The device shall provide real-time monitoring.
6. The particle count rate shall be measurable in environments with an omnidirectional integral flux up to $10^7 \text{ cm}^{-2} \text{ s}^{-1}$.
7. Detected events shall be classified in the device with respect to particle species (electron, proton or ion).
8. The device shall be radiation tolerant of a total ionising dose (TID) of at least 100 kRad (1 kGy).

The specification also states that the monitor may be either a single ASIC, or several such devices in a single package, provided that this results in a space-qualifiable, miniature monitor. It is also recognised in the statement of work that such severe limits on volume and mass mean that the monitor may not have the full particle identification capabilities of a scientific instrument.

Although many commercially important orbits have peak fluxes below $10^7 \text{ cm}^{-2} \text{ s}^{-1}$, the review in Chapter 2 showed that this value may be exceeded on some occasions. It was therefore decided to target a maximum omnidirectional flux of $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ to allow for wider usage or a greater operational margin.

An additional consideration is the device's output data rate. This should be extremely small, allowing convenient integration of the monitor into a range of satellites, possibly including multiple units on a single spacecraft. A mean rate of order 100 bit s^{-1} was targeted, providing a severe constraint on the monitor analysis algorithm and data products which are discussed in the next chapter.

3.2 Monitor architecture

This section considers the selection and modelling of the radiation sensors and particle identification scheme for the HMRM. These are largely constrained by the device requirements, and the particle species and energies present in Earth orbit.

3.2.1 Detectable particle energy ranges

As described in Chapter 2, medium Earth orbits experience intense fluxes of electrons and protons due to their passage through the radiation belts of the magnetosphere. Added to this are transient increases in the flux of higher energy particles, especially protons of order 10 MeV, due to solar events such as SEPs and CMEs. In addition, there is a far smaller but deeply penetrating flux of very high energy GCRs. Ion species are also present in all three of these components but generally in smaller numbers. The initial design therefore considered electron and proton detection only, with ions assessed at a later stage.

The monitor design was optimised for detection and discrimination of particles below energies of 6 MeV for electrons and 500 MeV for protons. These limits were chosen to include all trapped fluxes in the reference orbits, as modelled by AE-8/AP-8, and are also suitable for detection of steady state solar particles and transient events such as SEPs and CMEs. Notably, however, most of the GCR flux (peaking at approximately 1 GeV/n) exceeds these limits. It was decided that although these particles will be detectable their very low flux makes quantitative study with such a small monitor highly inefficient. For example, given a sensitive area of 1 mm², a mean of just three hits per hour due to GCRs would be expected, compared to more than 10⁵ s⁻¹ from magnetospheric particles. GCRs may be considered in future simulations but were not a primary driver of the monitor design.

3.2.2 Particle identification scheme

Many of the detection schemes and technologies reviewed in the previous chapter are unsuitable for use in a device with a volume of order 1 cm³. Further important constraints include tolerance of mechanical shock, reliability, low power consumption and low data rate.

It is clear that *stopping* the highest energy particles through continuous energy loss interactions in matter is not possible in a ~1 cm³ volume. This excludes any identification method based on ‘calorimetry’ principles, where the particle’s entire kinetic energy is converted to a detectable signal. Electrostatic analyser techniques are not practical for particle energies above ~10 keV, due to the very large voltages required. Meanwhile, achievable magnetic fields cannot create a sufficiently large trajectory deflection over ~1 cm distances to be useful in high energy particle discrimination.

The energy loss sampling method was instead selected as the most suitable approach for particle detection and identification in the HMRM. This involves measuring particle energy deposits in a telescope of two or more solid state sensors. The mean energy loss rate in each sensor, dE/dx, can be estimated from these deposit measurements, allowing the particle identity to be deduced. Although different particles with similar velocities may have equal energy loss rates, sampling over multiple sensors reveals differences in behaviour. This is achieved by slowing the particle so that each dE/dx value is sampled at a different velocity. At this stage several potential problems become apparent:

1. Particles may miss subsequent sensors due to scattering or oblique incidence angles. These cases are indistinguishable from those where a lower energy particle stops in the

previous sensor. These losses must be mitigated by bringing the sensors close together and the probability estimated through simulation.

2. Higher energy particles will require greater amounts of inter-sensor ‘energy degrader’ material to obtain a significant change in velocity. Enlargement of this material, however, may increase the monitor size and promote scattering loss. The choice of degrader composition and thickness must account for these opposing effects.
3. Energy deposit fluctuations will be the dominant limitation of particle identification resolution. This has two major causes: path length variation in the sensors and dE/dx variation due to the stochastic energy loss process. Little can be done to remedy the latter as it is a fundamental characteristic of the particle’s behaviour. Path length variation due to varying incidence angles may be constrained by requiring coincident sensor hits, limiting the incidence angle to be within a region of solid angle defined by the sensor telescope configuration.

3.2.3 Sensor technology selection: CMOS MAPS

Path length deviation within a sensor due to multiple scattering may be reduced by using a thinner sensor, which is also desirable as it allows coincident hits (multiple dE/dx samples) at lower particle energies. A thin sensor does, however, imply a smaller signal from a transmitted particle: a correspondingly lower sensor noise is therefore required to retain a sufficient signal-to-noise ratio. These considerations imply that pixel sensors are most appropriate.

An ASIC incorporating a CMOS APS array was chosen as having several practical advantages over competing semiconductor pixel sensor technologies, the most important of which are:

1. A wide range of readout and digitisation electronics may be integrated on the same chip, since these are easily fabricated in the same CMOS technology as the pixel array. This results in size and power reductions which are important for a miniature instrument.
2. Closely related sensor pixel designs have been proposed for use in electromagnetic calorimeters at collider experiments [88] and have achieved low equivalent noise charge values of less than ten electrons per pixel [89]. This excellent noise performance is due to the low leakage current and low capacitance of the pixel diode, which result primarily from its small size.
3. The overall ASIC thickness may be varied in the range 50–250 μm by back-thinning of the substrate, allowing a variable amount of material to be present between the sensors in the telescope.
4. The thickness of the sensitive epitaxial layer is 12 μm . This gives an expected signal due to a minimum-ionising particle (MIP) at normal incidence of approximately 1000 electrons and thus a signal-to-noise ratio close to 100.
5. Design techniques have been developed to increase the radiation hardness of CMOS APS sensors [90, 91]. This is an important consideration for a device which must be tolerant of a total irradiation of at least 1 kGy.

6. Charge collection within the continuous epitaxial layer of the device results in a large collection efficiency and a fill factor of nearly 100% [92].

The HMRM sensor ASIC

Each HMRM ASIC, developed by RAL Technology Department, incorporates a 50×50 monolithic active pixel sensor array and is fabricated in a $0.18 \mu\text{m}$ radiation hardened process technology. The pixel pitch of $20 \mu\text{m}$ provides a total sensitive area of 1 mm^2 . Each pixel is defined by a single charge collecting diode formed by the junction of an n-well with the $12 \mu\text{m}$ thick p-epitaxial layer. The pixels are expected to show approximately 10 electrons RMS equivalent noise charge and a full-well saturation of 19,000 electrons.

Each ASIC die measures 10.3 mm by 2.4 mm and has a thickness of $250 \mu\text{m}$ which may be reduced by back-thinning to a minimum of $50 \mu\text{m}$.

The pixels are read out simultaneously ('snapshot mode') and digitised on chip in a 3-bit analogue-to-digital conversion process involving seven comparator levels. Although not common in this sensor design, snapshot readout is required in this application to ensure that particle coincidences in the sensor stack are detectable. Each comparator level is programmed as a 7-bit threshold setting, allowing customisable, non-linear pixel digitisation schemes. A diagram showing the high-level ASIC design is provided in Appendix B and the ASICs are discussed further in Chapter 7.

Unlike a typical pulse-height analysis arrangement, digitisation of the collected charge is not triggered by a particle hit. Instead, pixel exposure and readout take place on a continuous cycle at a rate of 10 kHz . The pixels may integrate charge for a variable period within each readout cycle, ranging from $3.125 \mu\text{s}$ up to the maximum of $100 \mu\text{s}$; control of this variation, or 'electronic shuttering', allows the monitor to adapt to changes in flux magnitude. A field-programmable gate array (FPGA) was selected to provide computational, control and readout functions within the HMRM. These tasks include the generation of the waveforms required to drive the ASICs and reception of the output pixel data. The FPGA and its algorithms are described in further detail in the next chapter.

3.2.4 Conceptual design

The HMRM concept, shown in Fig. 3.1, comprises a telescopic stack of multiple sensors (number initially unspecified) through which particles may pass. This, together with the additional control electronics, is to be surrounded by a casing which provides structural support and shielding of low energy particles. An aperture allows selective admission of these lower energy particles, while a thin aperture cover blocks visible and ultraviolet photons (to which the sensors may also be sensitive). Particles are identified, where possible, from the characteristic combination of energy deposits in the sensors.

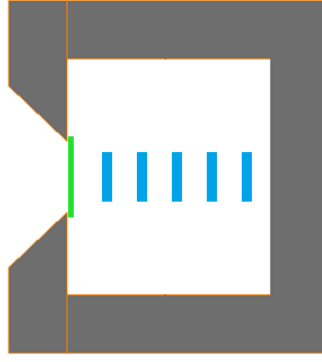


Figure 3.1: Conceptual diagram for the HMRM geometry; the cross-sectional view shows the essential features including the central sensor telescope, outer casing and shaped aperture with cover.

3.2.5 Telescope detection modes

The principal detection modes are the ways in which a single particle may hit the sensor telescope and give meaningful information about its characteristics. These modes, labelled C1, C2, etc., may be logically defined as:

$$C1 = S_1 \times \overline{S_2} \times \overline{S_3} \times \overline{S_4} \times \dots \quad (3.1)$$

$$C2 = S_1 \times S_2 \times \overline{S_3} \times \overline{S_4} \times \dots \quad (3.2)$$

$$Cn = \prod_{i=1}^n S_i \times \prod_{i=n+1}^N \overline{S_i}, \quad (3.3)$$

where S_n indicates a detected hit to the n^{th} closest sensor to the aperture and $\overline{S_n}$ indicates no hit, within a single sensor integration period. This labelling system can be extended to a telescope comprising any number of sensors, resulting in N overall modes for N sensors.

Despite not involving a coincident hit, C1 is important as it includes the lowest energy particles which cannot reach the second sensor. The remaining modes require successively higher minimum particle energies to occur, due to the progressively greater thicknesses of material the particle must penetrate. This allows each event to be assigned a *minimum* mean particle energy (dependent on its species) from allocation to one of these modes.

It is appealing to suppose that the mode also indicates a *maximum* mean particle energy, since a higher energy would result in penetration to the next sensor in the stack. However, particle losses between the sensors of the telescope must be negligible for this to be true.

Sensor hit combinations which do not correspond to a mode (e.g. $S_1 \times \overline{S_2} \times S_3$ for a three-sensor monitor) are less meaningful as the particle is likely to have undergone large angle scattering. This leads to a significantly greater, unconstrained path length and therefore uncertain energy loss.

The order of sensor hits over such short distances cannot be resolved and so ‘reverse’ coincidences, where a particle enters through the back of the monitor, cannot be distinguished. The monitor is, however, assumed to be mounted with the aperture facing away from the

spacecraft; the presence of this large effective shield means that energies of order 100 MeV for protons and 10 MeV for electrons are necessary for this to occur.

3.2.6 Event pile-up

Two or more particles may hit a sensor within a single integration period; this, known as *pile-up*, will result in under-counting of particles. Detection of this eventuality, via a pixel image recognition algorithm, is beyond the processing ability of the HMRM. Pile-up will therefore also result in misidentification as this process must assume that each event is caused by a single particle. This is not a concern for dose measurements, as the energy deposited in a pile-up event is simply the sum of the equivalent deposits expected from separate events (assuming sensor saturation is not reached). The count rate will be affected, however, with a simplistic measurement of non-zero sensor signals per unit time always underestimating the true particle incidence rate.

During a single sensor integration time period of t , pile-up occurs when there are two or more particle hits ($k \geq 2$) with probability given by:

$$P(k \geq 2) = 1 - P(k < 2) \quad (3.4)$$

$$= 1 - (1 + \lambda t) \exp(-\lambda t), \quad (3.5)$$

where λ is the mean particle arrival rate.

The relationship between the mean rate and the pile-up probability is shown in Fig. 3.2. A small ($\sim 1\%$) pile-up rate corresponds to a mean particle incidence rate at the sensor of less than $\sim 10^5 \text{ s}^{-1}$, provided that the exposure time may be reduced to $3.125 \text{ }\mu\text{s}$. A method for monitoring pile-up within the device is presented in Chapter 4, together with an algorithm to adjust the electronic shutter according to the particle environment.

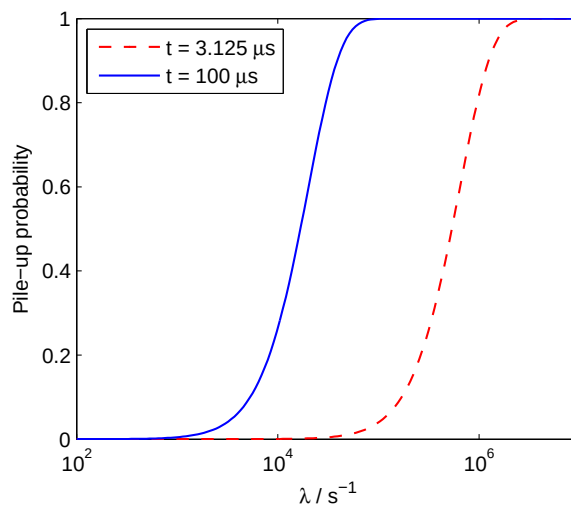


Figure 3.2: Pile-up probability as a function of mean particle incidence rate, for the maximum and minimum available exposure times, t .

3.3 Geometry optimisation methodology

From the conceptual design (Fig. 3.1), the following elements of the geometry were identified as impacting particle detection:

1. Aperture size and shape.
2. Casing material and thickness.
3. Aperture covering material and thickness.
4. Number of sensors.
5. Thickness of sensor ASICs.
6. Material and thickness of inter-sensor degraders, if present.
7. Positioning, spacing and alignment of sensors.

3.3.1 Design principles

Subject to practical constraints, mainly imposed by the small size of the monitor, the geometry elements were optimised according to the following design principles:

1. **The low energy particle detection threshold should be minimised.** This implies minimisation of the stopping power of the aperture cover, while still retaining its opacity to light. There is a potential conflict between this aim and that of minimising pile-up, since a lower energy particle detection threshold will result in a greater total incidence rate.
2. **The monitor must be blind to visible and UV light.** This is defined by calculating the predicted pixel signal generated in the front (most exposed) sensor due to sunlight, when pointing the monitor directly at the sun. This signal is required to be much smaller than the pixel noise (estimated at ten electrons); this ensures that fluctuations in this signal will not have an appreciable effect on the sensor performance by producing spurious detections.
3. **The probability of pile-up should be small in typical MEO omnidirectional particle fluxes (of order 1% in fluxes up to $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ and of order 10% in fluxes up to $10^8 \text{ cm}^{-2} \text{ s}^{-1}$).** Single particle identification will be incorrect in pile-up events so that the monitor's misidentification rate will be at least equal to the pile-up rate. However, the acceptance of the device cannot simply be made arbitrarily small to achieve low pile-up rates, since sufficient particles must be detected to make reliable, timely measurements of the particle environment.

In general, a balance must be sought between count rate and pile-up probability through suitable detector geometry design. Some flexibility is obtained by varying the sensor integration time electronically via a shutter control algorithm. However, operation in higher flux environments should be prioritised since these are most damaging and hence

of greatest interest. This should be achieved through the design of the casing and aperture.

4. **The particle energy threshold required to reach each successive sensor in the telescope should be an approximate geometric progression.** This energy spacing should be achieved through the choice of sensor substrate thickness where possible, or otherwise by including additional ‘degrading’ material between sensors. These choices should optimise the response for protons of energies up to 500 MeV and electrons up to 6 MeV, as identified in Section 3.2.1.
5. **The spacing between sensors should be minimised to reduce the probability of scattering losses.** As described earlier, these losses increase particle identification ambiguity. However, this aim is in potential conflict with point 4, since introducing material between sensors increases their separation and promotes large angle scattering as well as the desired energy loss. The relative merits of these two aspects must be balanced in the final design.
6. **The particle energy necessary for a coincident hit between two or more sensors should be minimised.** The advantages of coincidence detection mean that the monitor should be designed for this to occur over a wide particle energy range. This may be achieved by making the first ASIC as thin as possible.

3.3.2 Figures of merit

In some cases, the optimal geometry parameter could be selected from simple physical arguments. However, in general, geometry changes have multiple, often conflicting, impacts on performance. This was addressed by introducing three figures of merit: particle identification efficiency, identification purity and monitor effective area. Proposed geometry changes were applied to a simplified optimisation model and simulated within Geant4. Calculation of the figures of merit then allowed these changes to be evaluated.

Effective area

The effective area, A_e , of a sensor or sensor combination is defined as the cross sectional area of the sphere which would experience the same incidence rate when exposed to the same isotropic flux. The advantage of this quantity is that the mean incidence rate, λ , can be estimated easily from an omnidirectional integral flux value, F_o , via $\lambda = A_e F_o$. Note that A_e is a purely geometric quantity, depending only on size and positioning of the sensors, casing and aperture.

A_e is calculated by finding the solid angle averaged projected area of the sensor over all angles subtended by the aperture and/or coincidence telescope. The casing is assumed to be an absolute shield whereas the aperture cover is assumed totally transmissive. The formula may be expressed as:

$$A_e = \frac{1}{4\pi} \int \sin \theta \cos \theta \, d\theta \, d\phi \, ds, \quad (3.6)$$

where θ is the angle from the sensor surface normal to the observation direction, ϕ is the azimuth angle and ds is a sensor surface element. Analytical evaluation of this formula is rather cumbersome and it was instead evaluated computationally using a Monte Carlo ray-tracing method, where particle scattering is neglected.

An isotropic ‘fluence’ of N vectors was randomly selected from a sphere of radius R . The coordinates of intersection of these vectors with the planes corresponding to each of the sensors and the aperture were calculated, allowing identification of valid sensor and coincidence hits. The fraction of vectors (c/N) producing the required hit pattern (single sensor or detection mode) is recorded and the effective area is calculated as:

$$A_e = \frac{c}{N} \pi R^2. \quad (3.7)$$

The simulation proceeds until the statistical uncertainty in this quantity reaches a negligible magnitude; A_e values found in this manner may be used to calculate the expected mean hit rate (and hence determine the pile-up probability) in a specified flux.

Particle identification

Quantification of the monitor’s ability to identify single particles is a more difficult task than the evaluation of effective area. This ability depends partly on the details of the identification method which is employed. The suitability of this method in turn depends partly on the definition of the geometry. This interdependence is removed by use of a simple, provisional identification method (the operational identification algorithm for use within the HMRM is described in Chapter 4).

Each event is identified according to the simultaneous combination of energy deposits on the four sensors. Rather than attempt to allocate certain particle identities to these deposit combinations *a priori*, the deposit range selections are optimised to improve efficiency and purity for a given particle species and energy. These quantities are defined as:

$$\text{Efficiency} = \frac{\text{Number of particles correctly identified as type X}}{\text{Number of type X particles incident}}, \quad (3.8)$$

$$\text{Purity} = \frac{\text{Number of particles correctly identified as type X}}{\text{Total number of particles identified}}, \quad (3.9)$$

where ‘type X’ denotes a particle species and primary energy range. The geometry evaluation process is as follows:

1. The particle primary energy ranges are divided into log-spaced bins (for electrons in the range 0.04–6 MeV and protons in the range 1–500 MeV).
2. The candidate monitor geometry is exposed to a reference simulation within Geant4. This comprises an isotropic flux of equal numbers of electrons and protons distributed according to E^{-1} differential energy spectra. This provides efficient sampling of particle energies across the ranges chosen for nominal HMRM sensitivity. In general, 10^8 – 10^9 primaries of each species were used.
3. For each coincident detection mode, a characteristic energy deposit combination is selected for each primary energy bin. The simulated events are allocated to the nominal

primary energy bins according to their sensor energy deposit values. Purity and efficiency are then calculated for each bin.

The deposit range is adjusted using a heuristic optimisation algorithm in an attempt to maximise the selection purity (the fraction of identifications that are correct).

4. Together, the final purity and efficiency obtained for each primary energy bin constitute a measure of the monitor's ability to discriminate different particle types.

This process is repeated for different geometry configurations and the resulting purity and efficiency values are compared. Note that the values of these figures of merit are dependent on the simulated energy spectra and on the total electron/proton ratio. These remain constant for geometry comparison purposes; in general, however, these values will vary for different radiation environments.

3.3.3 Geometry optimisation model

The simplified Geant4 geometry model shown in Fig. 3.3 is an extension of the basic conceptual HMRM design, considering only the sensor telescope region. This includes the sensor ASICs, printed circuit board (PCB), outer casing and aperture. The thicknesses, spacings and materials of these geometry elements may be controlled via user command. Additional monitor features, such as electronic components, are neglected during this initial design phase.

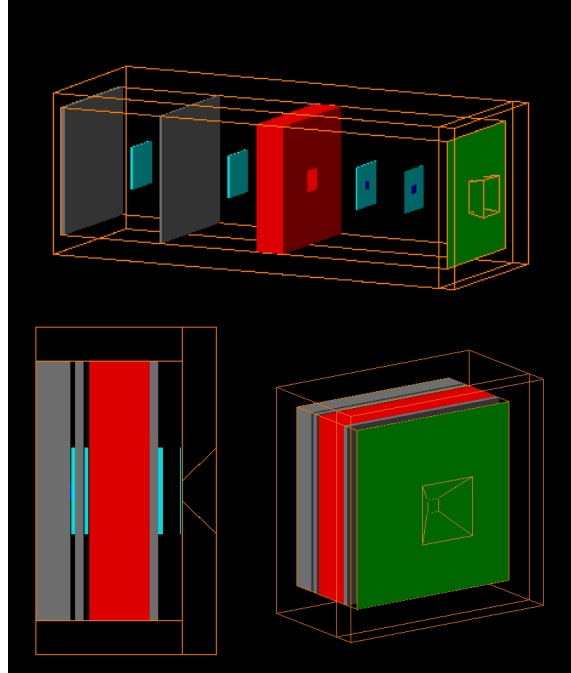


Figure 3.3: HMRM Geant4 optimisation geometry (with transparent outer casing): the top view is expanded to show the four ASICs, PCB (between second and third ASICs), energy degrader (between third and fourth ASICs) and the rear shield. The lower views show the layout with reduced spacings; the aperture is shown in a square-pyramidal configuration with an internal window material layer.

At this stage, the HMRM design was limited to a four-sensor telescope. This was judged to be the maximum number of ASICs which could be included, based on an assessment of PCB layout/interconnection and power consumption.

A baseline choice of geometry element sizes and compositions was simulated (optimisation run 1) and the particle identification efficiency and purity calculated for each primary energy bin. Each geometry modification considered was simulated as a single change to this baseline design and evaluated in a similar way. The results for these simulation runs are presented in Appendix C and summarised throughout this chapter.

ASIC simulation model

For the purposes of the detector geometry optimisation, the ASICs are modelled as a simple silicon block; the pixel array, hereafter referred to as the *sensor*, is included as a sensitive sub-region with $1 \times 1 \text{ mm}^2$ area and $12 \text{ }\mu\text{m}$ thickness.

Circuit tracks and passivation layers on the ASIC surface will influence the low energy particle detection thresholds and so must be included in the ASIC simulation model. Figure 3.4 shows a diagram of the circuitry of a single pixel; aluminium track coverage ranges from zero to three layers over small areas of the pixel, each with a thickness of $\sim 0.5 \text{ }\mu\text{m}$. Averaged thicknesses of these materials, listed in Table 3.1, are implemented as plane parallel layers over the simulated ASIC surface.

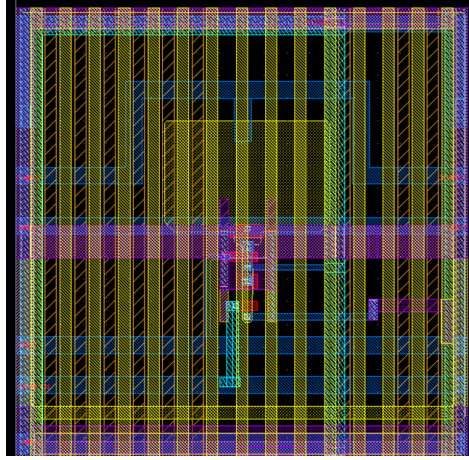


Figure 3.4: Diagram of the circuit track layout for one HMRM ASIC sensor pixel (courtesy N. Guerrini, RAL Technology).

During the geometry optimisation, the pixellation, noise and limited resolution of the sensor data are neglected. These aspects are introduced into the full sensor response model described in Chapter 4.

Table 3.1: Composition of the simulated HMRM ASIC sensor. The materials were implemented as parallel layers in the order listed; the substrate thickness is prior to back-thinning. The epitaxial layer is treated as the sensitive layer for particle detection.

Material	Layer thickness / μm
Silicon nitride (Si_3N_4)	0.3
Aluminium	0.6
Silicon dioxide (SiO_2)	1.7
Silicon epitaxial layer	12
Silicon substrate	238

3.4 Aperture window material selection

The CMOS APS used for the HMRM is a sensor technology originally developed for visible light imaging [89] and despite a surface covering of passivation layers and circuit tracks, the pixels are sensitive to light. Photon sensitivity is expected to be limited to energies greater than the silicon band gap of 1.14 eV (1087 nm wavelength), with quantum efficiencies of order 10% in the visible and near ultraviolet.

An opaque aperture cover (or *window*) is required to reduce the resulting signal due to sunlight to a magnitude smaller than the pixel noise. In so doing, material is placed in front of the first sensor, raising the low energy particle detection threshold. This should be minimised by careful material selection, while still ensuring that reliable light attenuation is obtained.

In this section the spectral transmissivity and expected pixel signal due to sunlight are calculated for candidate window materials. Geant4 simulations are also used to find the particle energy detection thresholds, resulting in the selection of a suitable window material. Two candidate materials were chosen:

1. **Aluminised polyimide film:** this comprises a 7.6 μm layer of polyimide ($\text{C}_{22}\text{H}_{10}\text{O}_5\text{N}_2$, often referred to by the brand name *Kapton*) with a coating of aluminium on both sides, each of 100 nm nominal thickness, manufactured by Sheldahl [93]. If necessary, multiple layers of this film could be overlaid to achieve the desired thickness. Although this material is used extensively to reflect sunlight for spacecraft thermal control [94], it has also been reported to degrade chemically after exposure to high proton fluences [95], resulting in decreased reflectivity at wavelengths greater than 500 nm.
2. **Aluminium foil:** pure aluminium with a naturally occurring oxide surface layer (typically of 2–10 nm thickness [96]). This foil is available in a wide range of thicknesses, although a minimum thickness limit of 8 μm was chosen due to the risk of handling and launch vibration damage.

3.4.1 Window light transmission

The spectral transmissivity of a stratified medium may be calculated using classical electromagnetic theory; the formulae presented here are those derived in [97]. Each constituent layer is modelled as a plane, parallel, homogeneous and isotropic material, with the whole assembly surrounded by a vacuum. A further assumption is that the electromagnetic wave is at normal incidence to the surface, making the polarisation irrelevant. In the general case, the transmissivity is given by:

$$T = \frac{p_b}{p_a} \left| \frac{2p_a}{(M_{11} + M_{12}p_b)p_a + (M_{21} + M_{22}p_b)} \right|^2, \quad (3.10)$$

while the reflectivity is:

$$R = \left| \frac{(M_{11} + M_{12}p_b)p_a - (M_{21} + M_{22}p_b)}{(M_{11} + M_{12}p_b)p_a + (M_{21} + M_{22}p_b)} \right|^2, \quad (3.11)$$

where the characteristic matrix, M , for the medium is the product of the matrices of the N constituent layers:

$$M = \prod_{j=1}^N m_j. \quad (3.12)$$

In the special case of normal incidence, these matrices are given by:

$$m = \begin{pmatrix} \cos \beta & -i\sqrt{\frac{\mu}{\epsilon}} \sin \beta \\ -i\sqrt{\frac{\epsilon}{\mu}} \sin \beta & \cos \beta \end{pmatrix}, \quad (3.13)$$

with:

$$\beta = \frac{2\pi}{\lambda} \tilde{n}z \quad (3.14)$$

$$p = \sqrt{\frac{\tilde{\epsilon}}{\mu}}, \quad (3.15)$$

where $\tilde{\epsilon}$ and $\tilde{n}$ are the complex spectral permittivity and index of refraction of the layer material, z is the layer thickness, and p_a and p_b are evaluated at the initial and final vacuum interfaces, respectively.

Refractive index data for aluminium were obtained from experimental [98] and theoretical [99, 100] investigations. These were used to calculate the spectral reflectivity of 100 nm aluminium foil, which is compared with the experimentally measured reflectivity of aluminised polyimide film in Fig. 3.5. This validates the derivation of Equations 3.10 and 3.11 and demonstrates that the film's optical properties are dominated by the aluminium layer.

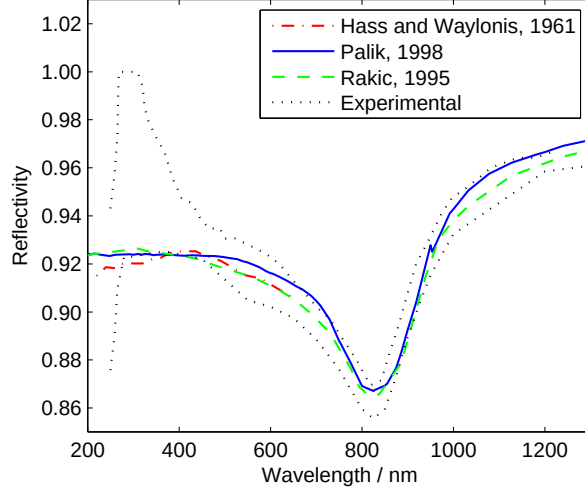


Figure 3.5: The spectral reflectivity of pure aluminium (100 nm thickness), calculated using different refractive index datasets, compared with experimental data (68% confidence interval indicated by the dotted lines) for aluminised polyimide film [101].

The calculated transmissivity of three thicknesses of free-standing aluminium foil are shown in Fig. 3.6, using the different refractive index datasets. These results show a rapid increase in attenuation as the layer thickness is increased.

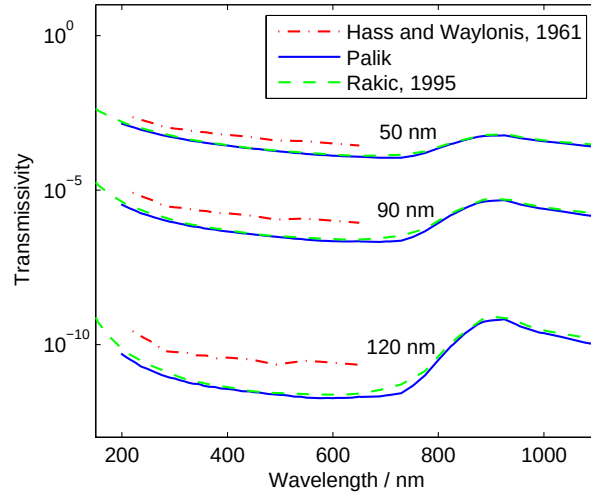


Figure 3.6: Calculated spectral transmissivity of three thicknesses of aluminium foil, using three refractive index datasets.

Sensor light response

The worst-case situation will be assumed, where the front sensor is fully illuminated with the window material as the only intervening object. The sun is modelled as a black body radiator with a temperature of $T = 5778$ K (the effective temperature of the photosphere [102]) and the incident spectrum, above the absorption of the Earth's atmosphere, is modelled by a Planck distribution.

The main source of uncertainty is the light sensitivity of the sensors, as this analysis was performed before a functioning device was available. The sensors are assumed to be unresponsive to photons with energy less than the silicon band gap (1.14 eV, corresponding to a wavelength of 1087 nm). At greater energies, up to approximately 3.65 eV (340 nm), each interacting photon will yield one electron-hole pair [39]. Above this energy a mean of one electron-hole pair is assumed per 3.65 eV of photon energy.

The interactive quantum efficiency is defined as the ratio of the number of photons detected to the number incident on the sensor outer surface, at a given frequency (it also accounts for interaction probability, sensor detection efficiency and fill-factor). This quantity is unknown for the HMRM sensor and instead data were obtained from published values of comparable technology (CMOS APS with a similar pixel design, size and epitaxial layer thickness) and are shown in Fig. 3.7. These two datasets are approximately indicative of the range over which the majority of devices (including the HMRM sensor) are expected to lie. The data have been extrapolated down to the silicon band gap energy (1.14 eV) and up to 8.27 eV, which is the photon energy at which the solar spectral intensity is $\sim 0.4\%$ of its peak intensity. These therefore become the photon energy limits between which the analysis is carried out. This extrapolation is a worst-case estimate, since the true response is expected to decrease at these two extremes.

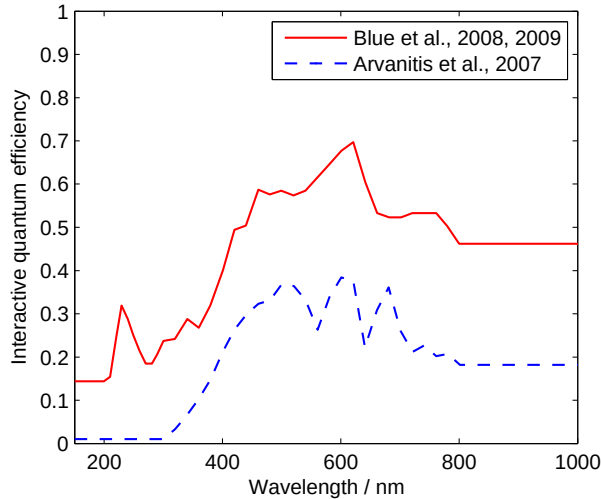


Figure 3.7: Estimates of the sensor interactive quantum efficiency [103, 104].

Expected pixel signal

The expected mean pixel charge signal is calculated assuming a pixel area and maximum integration time of $400\text{ }\mu\text{m}^2$ and $100\text{ }\mu\text{s}$, respectively. This was evaluated for windows of various thicknesses of aluminium, using all combinations of the input spectral data (quantum efficiency and $\tilde{n}$). This gives an approximate indication of systematic uncertainty in the output results, represented in Fig. 3.8 by a maximum and minimum estimate range.

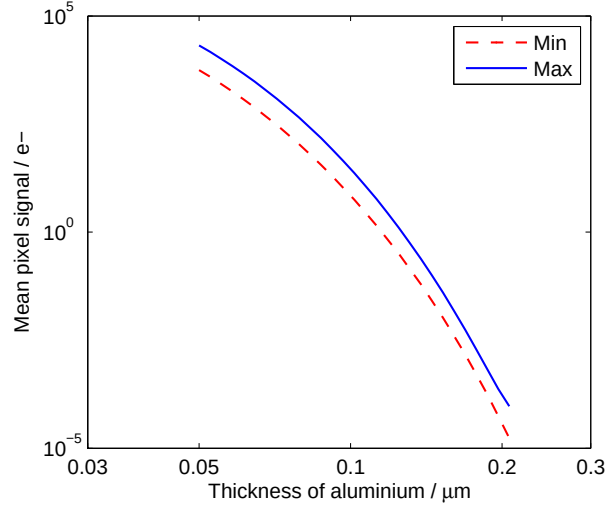


Figure 3.8: Estimated mean pixel signal for a $100\text{ }\mu\text{s}$ exposure (minimum and maximum estimates) caused by solar illumination of a silicon CMOS APS sensor covered by aluminium foil, as a function of foil thickness.

3.4.2 Charged particle attenuation

A further consideration is the effect of the window material on incident charged particles, where stopping or scattering may occur. The candidate materials were included in Geant4 simulations with the ASIC model; isotropic fluences of electrons and protons were generated and used to find the fraction detected as a function of particle primary energy. These data are shown in Figs. 3.9 and 3.10 for various numbers of layers of the aluminised polyimide film and various thicknesses of pure aluminium foil.

The particle energy at which 50% of particles are detected (depositing $\geq 365\text{ eV}$) may be read from these transmission plots — these extracted data are given in Fig. 3.11. Also indicated on these figures is the detection threshold due to the ASIC surface dead layer alone, with no window material present.

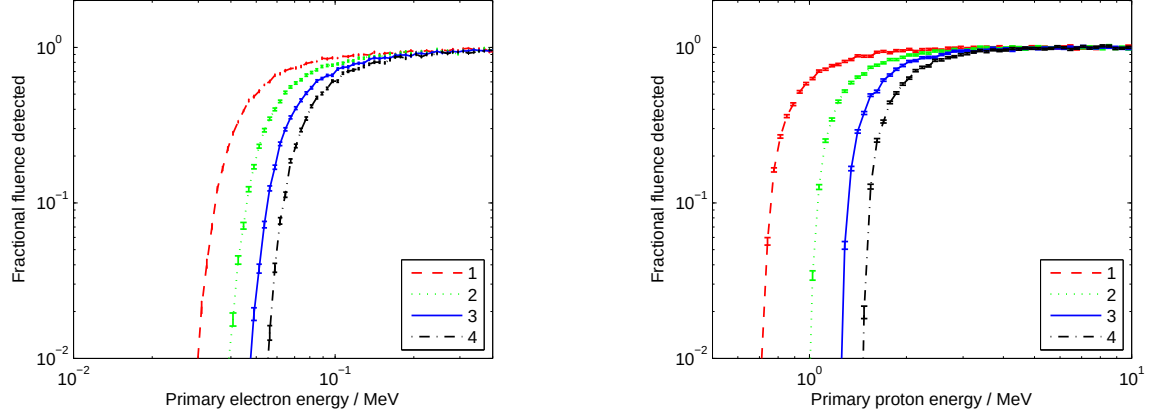


Figure 3.9: Fractional isotropic fluences detected by the first sensor, as a function of primary particle energy, when using 1–4 layers of the aluminised $7.6\text{ }\mu\text{m}$ polyimide film as an aperture window. This is shown for electrons (left) and protons (right).

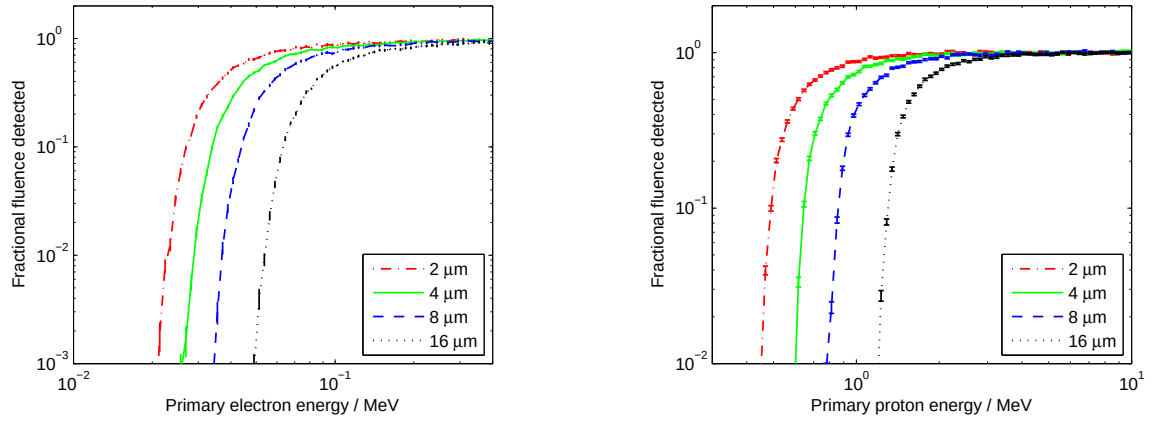


Figure 3.10: Fractional isotropic fluences detected by the first sensor, as a function of primary particle energy, for aperture windows of various thicknesses of aluminium. This is shown for electrons (left) and protons (right).

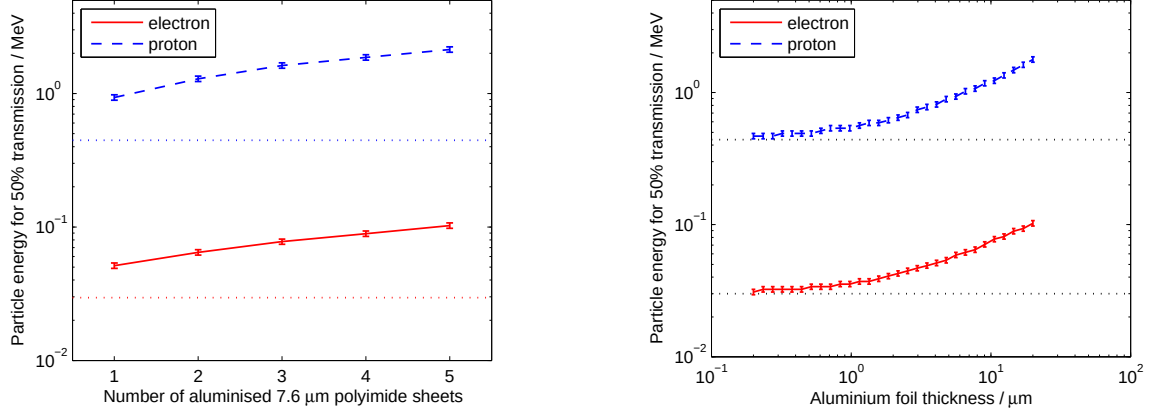


Figure 3.11: 50% transmission energies for electron and proton fluences passing through aperture windows consisting of multiple layers of aluminised polyimide film (left) and single layers of aluminium foil of various thicknesses (right). The horizontal dashed lines indicate the 50% energy when no window material is present.

3.4.3 Aperture window selection

The expected pixel signal due to sunlight is found to change rapidly with small changes in the thickness of aluminium, especially for thicknesses of order 100 nm. This is significant for the polyimide film which, since it is not designed for such a critical application, is likely to have an aluminium thickness tolerance of at least 10% (200 ± 20 nm in total). Also note that because the optical attenuation is provided by the aluminium layer (polyimide acting only as structural support), most of the undesirable particle stopping power of the film is only indirectly necessary. For these reasons a single free-standing aluminium foil is preferred.

The results suggest that an aluminium foil with thickness greater than 200 nm produces a negligibly small optical pixel signal. The choice of thickness thus becomes a compromise between the greater strength and increased particle attenuation that greater thicknesses provide. A foil of 8 µm was chosen, having been identified as safe to handle and bond to the casing without risk of damage. Moreover, the particle attenuation suffered with this foil is comparable to the alternative of a double layer of polyimide film. This choice provides low-energy 50% detection thresholds of (0.063 ± 0.001) MeV for electrons and (1.05 ± 0.01) MeV for protons.

3.5 Sensor telescope optimisation

The telescopic arrangement of the four sensor ASICs must be optimised according to the design principles. This involves the choice of die thicknesses and spacings, and the design of the inter-sensor energy degrader. In the following discussion, each ASIC die is denoted D1, D2, etc. in order of proximity to the aperture, while the small pixel sensor region of the ASIC is denoted S1, S2, etc. with the corresponding numbering system.

The ASICs (initially 250 μm thick) may be back-thinned to achieve coincident hits at lower particle energies; this has an estimated safe minimum of 50 μm (allowing for handling and wire-bonding stresses). Sufficient material should be retained such that particles are slowed significantly between sensors, allowing measurement of their dE/dx at different velocities.

Nevertheless, it was anticipated that additional material would be needed to slow higher energy particles. The geometry concept therefore allows the inclusion of a material layer between D3 and D4, the composition and thickness of which must be optimised.

Finally, the spacings of the ASICs are significant parameters as they determine the solid angle acceptance of each coincident detection mode. Greater separations also constrain the selection of a smaller incidence angle range, potentially reducing the variance in path length (and thus in energy deposit). The disadvantages of increased separation include a larger monitor and a greater probability that particles will exit the telescope between sensors. A compromise which balances these conflicting effects was sought.

3.5.1 ASICs D1 and D2

The first two sensor ASICs may be considered in isolation; the most important aim is to achieve a sensor coincidence at the lowest possible particle energy (this will be the C2 mode). This maximises the proportion of events for which at least two dE/dx estimates are obtained, improving particle identification ability.

ASIC D1 is thus thinned to the minimum limit (50 μm) and bonded directly to the front (sensor-containing) surface of D2. This close positioning of D1 and D2 is possible because the wirebond contacts are along two edges only, allowing the ASICs to be offset and bonded as shown in Fig. 3.12. Although the small sensor separation enlarges the accepted incidence angle range, scattering losses are minimised with this approach. Such losses leave electron identification highly ambiguous and so this is an acceptable compromise.

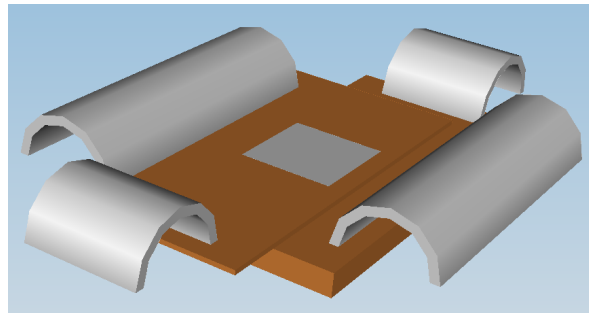


Figure 3.12: Diagram showing the arrangement of D1 and D2 — the ASICs are offset, allowing clearance for the semicircular wirebonds along two sides of each chip (represented in this CAD by gray hemi-cylindrical volumes. The pixel sensors (central square region) on each die remain vertically aligned (courtesy O. Poyntz-Wright, RAL Space).

The D1/D2 arrangement was simulated in Geant4 to find the particle detection thresholds of the second sensor; the fractions detected, as a function of particle energy, are shown in Fig. 3.13 and the extracted 50% detection thresholds are given in Table 3.2. S1 thresholds are defined by the 8 μm aluminium window and ASIC surface passivation, while S2 thresholds are defined by these layers and by the thickness of D1.

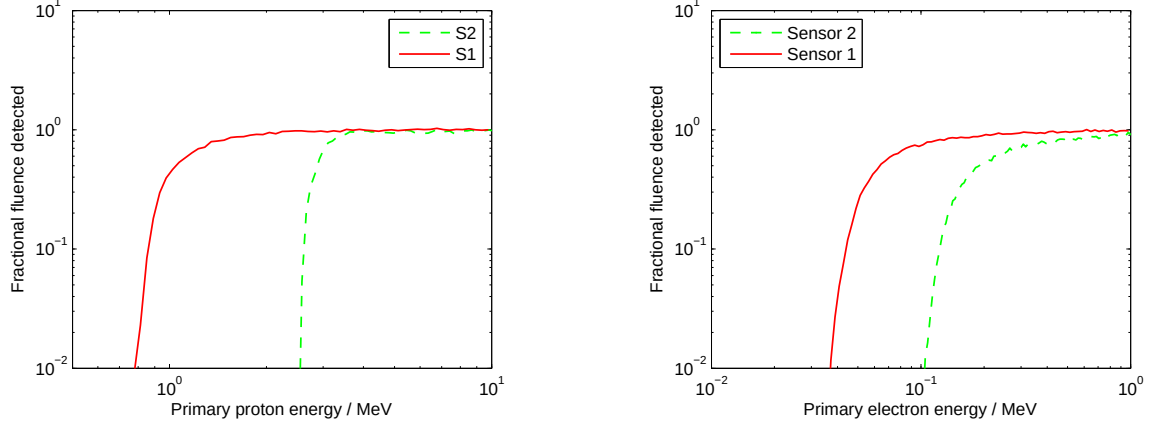


Figure 3.13: Fractional isotropic fluences detected by S1 and S2, as a function of primary particle energy, for protons (left) and electrons (right).

3.5.2 ASICs D3 and D4

Following the design principle of geometrically increasing threshold energies for each sensor, the target thresholds for the remaining sensors (S3 and S4) were calculated as a geometric progression between the thresholds for S2 and the chosen maximum energy limits (500 MeV protons and 6 MeV electrons). These targeted thresholds are also listed in Table 3.2.

Table 3.2: Detection thresholds (50% transmission) for particle flux reaching the four sensors. The values for S1 and S2 were determined via Geant4 simulation (statistical uncertainties were less than 1%), while the values for S3 and S4 are extrapolated targets based on a geometric progression.

Sensor number	Particle energy for 50% detection / MeV	
	Proton	Electron
1	1.05	0.063
2	2.88	0.191
3	16.07	0.600
4	89.62	1.900

A constraint on the telescope layout was imposed due to the requirement of a PCB between D2 and D3 to carry the electrical connections to these ASICs. Furthermore, the upper surface of each ASIC (containing the sensor and wirebond contacts) must face away from the PCB in order to accommodate the wirebonds.

A central hole is provided in this PCB to prevent additional particle attenuation between S2 and S3: this is visible in the optimisation geometry shown in Fig. 3.3. An equal die thickness for ASICs D2, D3 and D4 was considered desirable to simplify the device selection process and avoid multiple back-thinning requirements. Simulation with the optimisation geometry showed that the use of 250 μm ASICs provided S3 with the correct approximate thresholds, while the S4 threshold remained too low. This indicated that an energy degrading layer was required between S3 and S4.

Inter-sensor energy degrader material

Both the material and thickness of this layer must be chosen to achieve the required primary energy reduction but minimise lateral scattering. A denser material can achieve the required stopping power in a thinner layer and hence less lateral spreading for a given angular deflection. However, higher density typically implies higher atomic number and thus a greater probability of large angle electron scattering. To investigate the relative importance of these two effects, and to seek a compromise, Geant4 simulations were run using several candidate elements chosen for their availability and range of atomic numbers (beryllium, aluminium, titanium and tantalum).

Beams of 2 MeV electrons were fired at normal incidence into blocks of the test materials, all of which had a thickness chosen to give an areal density of 0.5 g cm^{-2} (and hence similar stopping powers). Particles exiting the block were histogrammed according to their lateral distance from the beam axis — this is shown in Fig. 3.14. This measures beam dispersion, taking into account both angular scattering and the thickness of the sample. A separate set of histograms (Fig. 3.15) shows the energy distribution of the exiting particles, confirming that equal areal densities give approximately equal energy reductions ($\sim 0.6 \text{ MeV}$). Both plots are normalised relative to the number of *incident* particles, hence backscattering (which gives no forward exit particle) results in a decrease in the total probability of any parameter value. The backscatter fraction, and the probability that the particle exits within a 0.5 mm radius (the sensor size), are given in Table 3.3 for each material.

Table 3.3: Particle scattering and transmission data for 2 MeV electrons incident on layers of four materials. The probability that an incident particle exits within a circle of 0.5 mm radius about the beam axis is given, together with the backscattering fractions. These data were obtained through Geant4 Monte Carlo simulation (statistical uncertainties were less than 1%).

Element	Layer thickness / mm	Backscattering fraction	Probability of $r < 0.5 \text{ mm}$
Tantalum	0.44	0.77	0.23
Titanium	1.10	0.41	0.30
Aluminium	1.85	0.26	0.22
Beryllium	2.70	0.03	0.35

Beryllium, the lightest material tested, gives the lowest backscattering rate and the greatest probability of detection with a sensor situated behind the degrader. However, beryllium

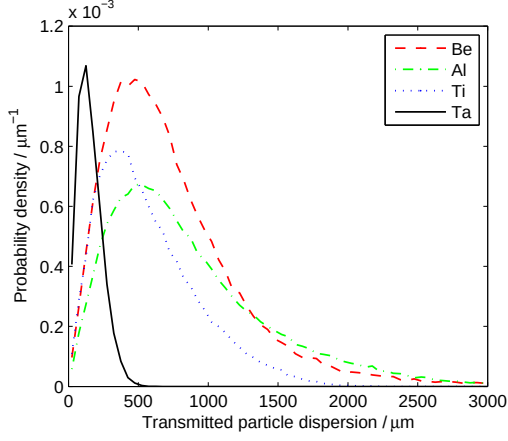


Figure 3.14: Distribution of the electron exit distance from the beam axis.

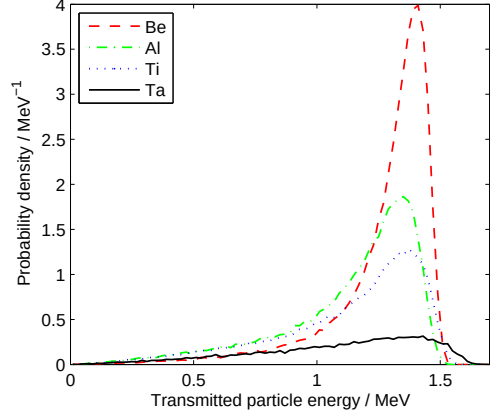


Figure 3.15: Distribution of the kinetic energy of electrons which exit the shield sample.

would increase the length of the monitor by $\sim 10\%$ and is a safety risk due to its toxicity. Molybdenum, showing similar properties to tantalum, was also considered and rejected. Instead, the second best material tested, titanium, was chosen since this could be incorporated into the supporting structure and results in a more compact design.

Degrader thickness

A series of monitor geometry simulations were performed using different thicknesses of titanium degrader between 0.5 mm and 5 mm. A thickness of 2 mm (areal density of 0.9 g cm^{-2}) was selected as providing appropriate energy thresholds for S4. The relative effect of material choice on particle identification was tested by also simulating equal areal densities of aluminium (3.36 mm) and molybdenum (0.89 mm), with the optimisation geometry (optimisation runs 2 and 3, respectively). The aluminium and titanium layers produced similar detection efficiencies for high energy electrons (3.3–6 MeV), while the molybdenum layer led to a significant decrease consistent with the predicted increased scattering losses.

3.6 Outer casing and aperture

Introduced in Section 3.2.6, the probability of event pile-up must be minimised to ensure accurate radiation monitoring. Figure 3.2 showed that a 5% pile-up probability implied a mean of 0.36 hits per readout frame. The hit rate to a planar area A , from one side only, due to an isotropic omnidirectional flux, F_o , is $F_o A/4$. Thus, neglecting particle detection thresholds, a 1 mm^2 sensor in a $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ isotropic flux is expected to receive a maximum mean hit rate of $2.5 \times 10^5 \text{ s}^{-1}$. This indicates a mean of 0.78 hits in the shortest possible charge integration time ($3.125 \text{ } \mu\text{s}$), and a resulting pile-up probability of 18%. The exposure of the HMRM sensors must be reduced by a factor of ~ 2 to ensure an acceptable pile-up rate (of order 1%).

Enclosing the sensor telescope in a material casing structure provides attenuation of lower energy particles. This structure also gives mechanical support and enables attachment to the spacecraft.

3.6.1 Aperture design

An appropriately shaped aperture in the casing allows particles which are within the sensor telescope acceptance to enter with negligible loss of kinetic energy. This means that the aperture should match, or be greater than, the solid angle subtended by the coincident detection modes (introduced in Section 3.2.5). Note that mode C1 (no coincidence) has an unrestricted angular acceptance and hence must always be partially obscured by the casing.

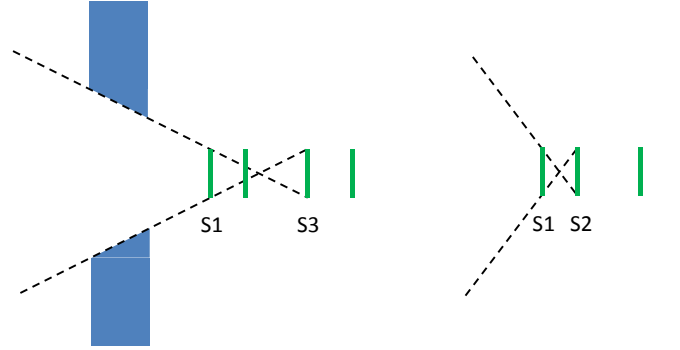


Figure 3.16: Illustration of the entrance aperture when defined by the C3 mode (left), with the C2 mode solid angle also shown (right).

Trial aperture shapes were matched to the solid angle of the C2 and C3 modes, as illustrated in Fig. 3.16. The preferred choice of C2 has the advantage of also encompassing the solid angles defined by the C3 and C4 modes, providing uninterrupted trajectories for these particles. However, C2 produces a larger aperture than C3 and thus risks a higher pile-up rate.

For both definitions, the exposure (and resulting pile-up rate) of the individual sensors must be assessed; of these, S1 is most exposed and hence S1 pile-up limits monitor operation in high flux environments.

The aperture shape, in addition to its angular size, must be considered. This may be designed to match the square-pyramidal region of solid angle defined by the square sensors in coincidence. An alternative is to use a conical aperture, sized to contain this solid angle. A geometry simulation (optimisation run 4) tested the use of a C2 conical aperture for comparison with the baseline C2 pyramidal aperture (shown in Fig. 3.3), from which a negligible difference in performance was obtained. The conical shape was finally adopted since its pitch angle may be manufactured with greater accuracy.

The sensor effective areas were calculated for geometries with C2 and C3 conical apertures via a ray-tracing simulation. The estimated S1 pile-up probabilities for these configurations in high-flux environments are shown in Table 3.4.

Table 3.4: Calculated effective areas, A_e , for the four individual sensors (S1–4) and the four detection modes (C1–4) using the two proposed circular aperture definitions. The expected pile-up probabilities resulting from isotropic fluxes of $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ and $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ are also given. Statistical uncertainties are 0.001 mm^2 for effective area values and 0.01% for pile-up percentages.

Single sensor or detection mode	Using C2 aperture definition			Using C3 aperture definition		
	A_e / mm^2	Pile-up probability in a $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ isotropic flux	Pile-up probability in a $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ isotropic flux	A_e / mm^2	Pile-up probability in a $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ isotropic flux	Pile-up probability in a $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ isotropic flux
S1	0.245	0.28%	17.90%	0.150	0.11%	8.08%
S2	0.236	0.26%	16.91%	0.132	0.08%	6.51%
S3	0.154	0.11%	8.42%	0.057	0.02%	1.41%
S4	0.035	0.01%	0.56%	0.008	0.00%	0.03%
C1	0.245	0.28%	17.90%	0.150	0.11%	8.08%
C2	0.183	0.16%	11.23%	0.126	0.08%	5.95%
C3	0.052	0.01%	1.19%	0.052	0.01%	1.19%
C4	0.007	0.00%	0.02%	0.007	0.00%	0.02%

The C3 aperture definition reduces pile-up to an acceptable level ($<10\%$) in a $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ flux. Note that pile-up in a real radiation environment (rather than the ray-tracing approximation) is expected to be higher due to casing penetration; this is assessed in Section 3.8.1.

Adoption of the C3 aperture design means that the C2 detection mode (with a larger solid angle) is partially obstructed by the casing surrounding the aperture, potentially affecting particle identification. Comparison of the simulated monitor with C3 aperture (optimisation run 5) and the baseline geometry (with C2 aperture) revealed increased purity and efficiency for $\sim 5\text{--}30 \text{ MeV}$ protons when using the C3 aperture. However, purity and efficiencies were decreased for detection of higher energy protons ($>30 \text{ MeV}$), while electrons were not affected significantly.

ASIC radiation exposure

A further consideration for aperture design is the radiation exposure of the HMRM electronics, particularly the non-sensor areas of the first ASIC which is closest to the aperture. Although radiation tolerant ASIC design techniques are used, limiting this exposure should result in improved performance and lifetime.

The spatial extent of the most exposed region depends on the diameter of the aperture and its proximity to the ASIC surface. A minimum is placed on this latter parameter since sufficient distance (estimated at 1.2 mm) is required to accommodate the ASIC wirebonds. The directly exposed spot diameter at the location of D1 (neglecting particle transmission through the casing) is given in Table 3.5 for both aperture designs, together with their principal dimensions. Note that the ASIC has a surface measuring $10.3 \times 2.4 \text{ mm}^2$. Approximately half of the chip is directly exposed when using the C2 aperture, while less than 15% is exposed with a C3 aperture.

Table 3.5: Parameter values for the two proposed aperture definitions: these assume a front casing thickness of 2 mm. Also given is the radius of the circular region on the first ASIC die (D1) which is directly exposed to radiation through the aperture.

Parameter	Parameter value	
	C2 definition	C3 definition
Shape	Conical	Conical
Inner radius	1.5 mm	0.7 mm
Outer radius	13.3 mm	2.8 mm
Pitch angle	80°	46°
D1 exposure radius	2.7 mm	1.0 mm

Aperture selection

The C3 aperture was found to reduce pile-up on the first sensor by more than 50%, in a $10^8 \text{ cm}^{-2} \text{ s}^{-1}$ flux, compared to the C2 definition. Although both apertures give a negligible pile-up rate in a $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ flux, the additional dynamic range is desirable to offset the expected increase due to casing penetration. This advantage, together with the reduction in ASIC radiation exposure, led to the adoption of the C3 conical aperture.

3.6.2 Casing design

The aperture design method assumed that complete attenuation occurs for particles incident upon the casing. Construction of such a casing is not feasible since energetic particles in the near-Earth environment have mean ranges exceeding ~ 1 cm in common metals (and hence similar to, or larger than, the planned overall monitor size). Instead, the casing should shield sufficient low energy particles (which constitute the overwhelming majority) such that their contribution to pile-up is negligible compared with the flux of particles entering through the aperture.

The AP-8 and AE-8 modelled proton and electron spectra for the reference orbits, presented in the previous chapter, show that the different orbits and species considered result in different requirements. Approximately 99% of the particle flux is due to electrons below 1.3 MeV and protons below 20 MeV (excluding orbit A, for which the proton limit rises to ~ 70 MeV).

Following the same argument as described when choosing the inter-sensor degrader material in Section 3.5.2, a casing composed of titanium was preferred. This represents a compromise between greater density (providing a smaller volume) and increased particle scattering. Using CSDA estimates, the particle energies identified above imply a mean range in titanium of approximately 2–3 mm. The baseline monitor geometry (run 1) was implemented with a 2 mm titanium shield on all surfaces.

Bremsstrahlung in the casing

In addition to increased particle scattering, a high-Z material such as titanium is expected to cause greater electron bremsstrahlung losses. This may have two adverse effects on monitor performance: (i) creation of false hits due to photon absorption in the sensors and (ii) increase of particle energy loss fluctuations in the casing, reducing primary energy resolution. The former is predicted to have reduced importance due to the thin sensors used in the HMRM, which have low efficiency for photon absorption.

Nevertheless, these effects may be reduced by using a lower Z material (e.g. aluminium) while retaining equal areal density to provide approximately equal stopping power. However, this would result in a less compact monitor (e.g. a 3.36 mm aluminium casing rather than 2 mm of titanium).

Electron bremsstrahlung mean energy losses are approximately constant as a fraction of the electron energy:

$$\frac{dE}{dx} = -\frac{E}{x_0}, \quad (3.16)$$

where x_0 is the radiation length in the material; $x_0 = 3.6$ cm for titanium and $x_0 = 8.9$ cm for aluminium. However, accounting for the different material thicknesses, the fractional energy loss for crossing an areal density of a is:

$$\frac{E_0 - E}{E_0} = 1 - \exp\left(-\frac{a}{\rho x_0}\right), \quad (3.17)$$

where ρ is the material density. For the thicknesses considered above, these losses are 3.7% and 5.4% for aluminium and titanium, respectively. Compared to the 40% reduction in casing volume (itself the largest single monitor component) this is a relatively small difference

A further consideration is the electron energy at which bremsstrahlung becomes an important energy loss mechanism. The critical energy in a material, where ionisation losses equal radiation losses, is given by the approximate parameterisation [25]:

$$E_c = \frac{610 \text{ MeV}}{Z + 1.24}, \quad (3.18)$$

giving values of $E_c = 26$ MeV for titanium and $E_c = 43$ MeV for aluminium. These energies are greater than the HMRM discrimination limits, so that significant modification of the primary energy is not a concern. Furthermore, the flux of these energies in Earth orbit is negligibly small. A final test was to simulate a 3.36 mm aluminium casing (run 2), verifying that identification purity and efficiency were not significantly improved over the titanium baseline choice.

Casing thickness

A 2 mm titanium HMRM casing is estimated to have a mass of approximately 14 g; reduction of the lateral shielding to 1 mm thickness (while retaining 2 mm on the front face surrounding the aperture) gives a casing mass of 8.2 g. This reduction is estimated to be of order 10% of the total monitor mass and so would be a desirable modification. Simulation of this change

to the optimisation geometry (run 6) resulted in negligible impact on identification purity and efficiency. Most impurity particles are expected to enter through the aperture itself, or through the front casing face, and thus the lateral shielding is of secondary importance.

An additional concern is the backscattering of particles from the rear casing wall, inside of the monitor, allowing them to re-enter the telescope. Table 3.3 showed that aluminium has a 37% smaller backscattering rate for 2 MeV electrons than titanium. An internal 1 mm layer of aluminium was therefore added inside the casing, on this surface only, to reduce the probability of these events.

3.7 Monitor design implementation: HMRM version 1.0

The preferred characteristics of the HMRM detector geometry, as presented in this chapter, were summarised as a set of design recommendations [86] provided in Table 3.6. Approximate practical constraints (where known) had been applied during the optimisation process, including mass and volume limits as well as consideration of difficulties in the construction process.

Further limitations were recognised during the development of the first version of the full electromechanical design. These concerned the spaces between PCBs and ASICs, some of which had to be increased to accommodate the ASIC wirebond connections. These changes are also summarised in Table 3.6.

Table 3.6: HMRM detector geometry design recommendations.

Design recommendation	Adopted in HMRM v1.0?
Front die (D1) to be back-thinned to 50 μm	Yes
D2, D3 and D4 to be 250 μm thick	Yes
No additional shielding between D1 and D2	Yes
No additional shielding between D2 and D3	Yes
Degrader (between D3 and D4) to be 2 mm Ti	Yes
Back shield (behind D4) to be 1 mm Al + 1 mm Ti	Yes
Window to D1 distance close to 0.5 mm	No: 0.98 mm used
S2 to S3 distance smaller than 0.25 mm if possible	No: 0.54 mm used
S3 to S4 distance to be 30 μm (in addition to degrader thickness)	No: 0.15 mm used
Entrance window to be 8 μm aluminium	Yes
Entrance aperture solid angle defined by C3 mode	Yes
Conical or pyramidal aperture cross section acceptable	Yes (conical)
Titanium casing	Yes
Casing to be 2 mm thick on the front surface, 1 mm on sides	Yes

A computer aided design (CAD) for the first implementation (HMRM version 1.0) was developed by RAL Space. This provides dimensions and materials for all structural components, together with positions of the ASICs and all detector geometry elements. The majority of

the mass (estimated at 52 g) of the monitor is included in this model with, most significantly, an accurate representation of the detector geometry region.

To assess the realistic particle detection properties of this design, each individual volume of the CAD model was implemented as a solid volume in the Geant4 geometry description shown in Fig. 3.17. This resulted in a highly faithful representation, as illustrated by the comparison in Fig. 3.18, and has been used for all subsequent monitor simulations. Figure 3.19 shows an annotated cross-section view of the Geant4 geometry identifying the principal components. Finally, the first prototype HMRM unit is shown in Fig. 3.20.

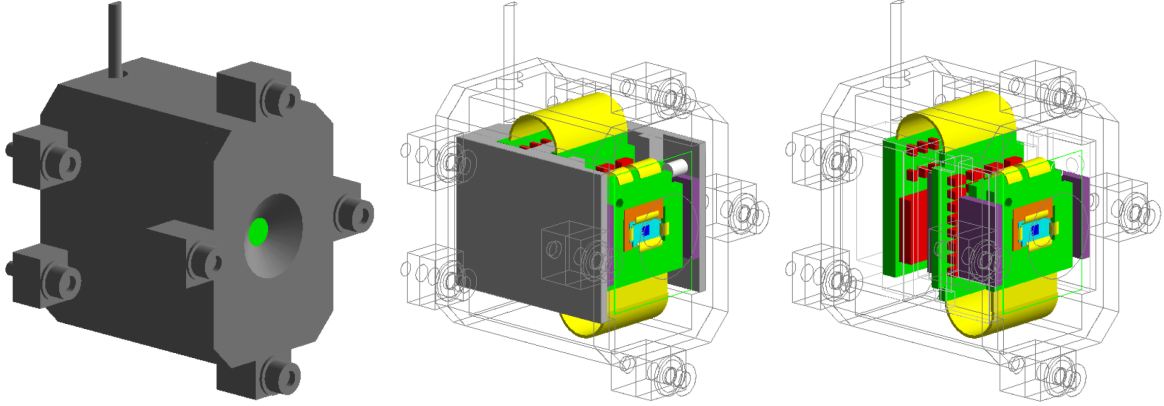


Figure 3.17: The HMRM version 1.0 Geant4 geometry model, shown in three views revealing the internal structures. Left to right: (1) the titanium casing, aperture and aluminium window; (2) aluminium support, front PCB, first sensor; (3) four PCBs (electronic components are approximate only), flex harnesses. The outer dimensions of the casing, excluding bolts/lugs, measure $1.7 \text{ cm} \times 2.4 \text{ cm} \times 2.2 \text{ cm}$.

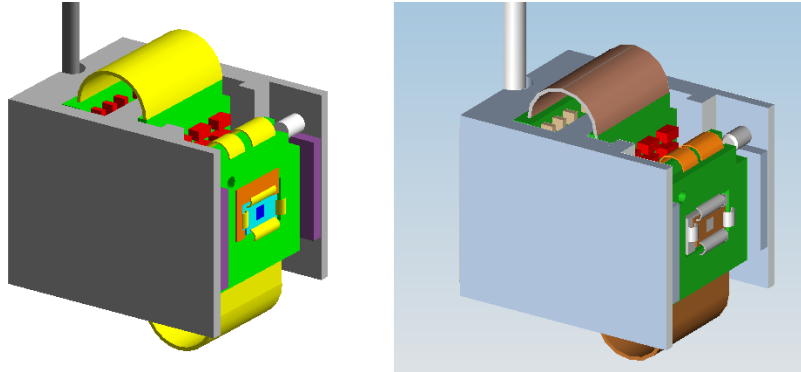


Figure 3.18: Comparison of the Geant4 geometry model (left) and electromechanical CAD (right; courtesy O. Poyntz-Wright, RAL Space), with the outer casing removed.

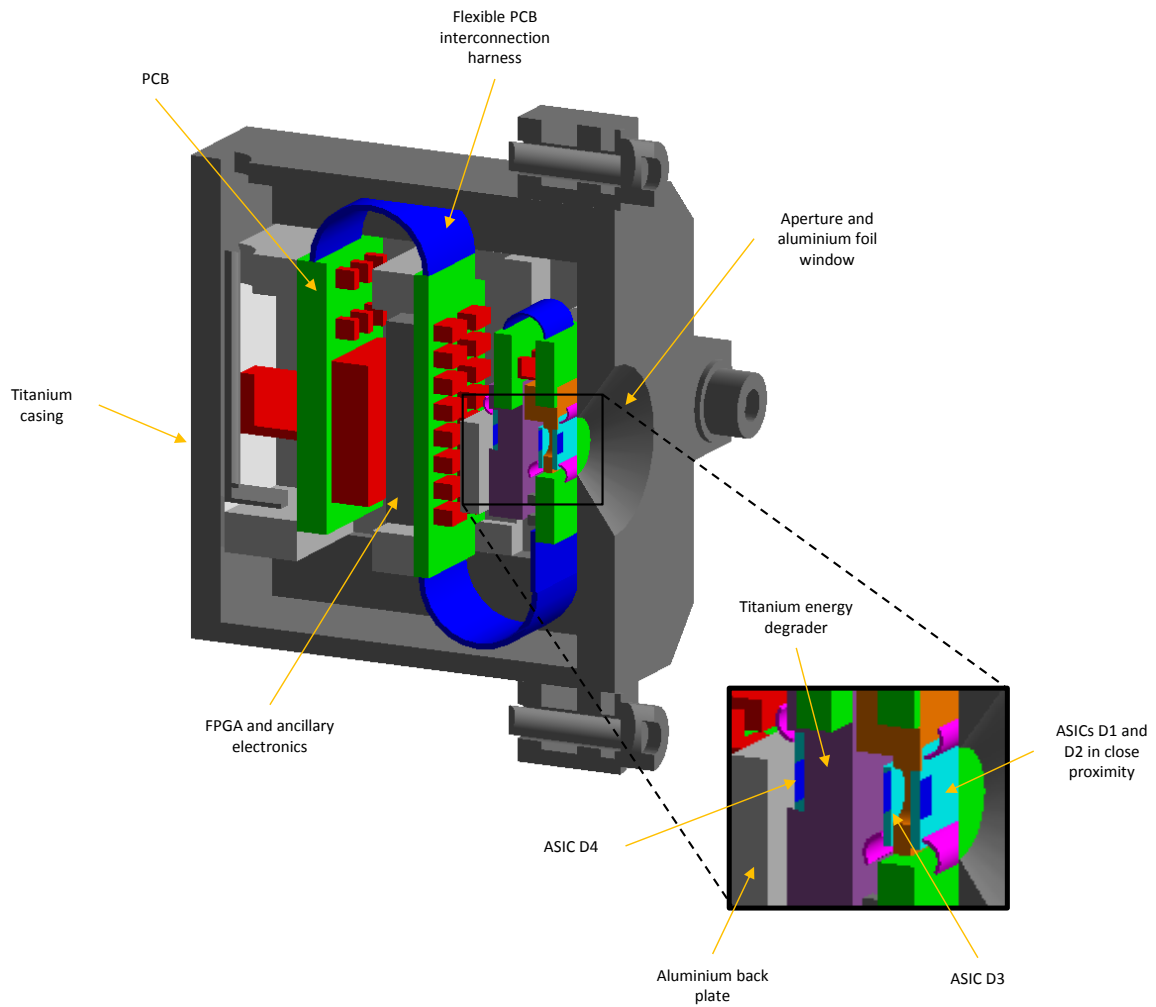


Figure 3.19: Cross-section view of the HMRM Geant4 geometry, showing the principal elements of the detector geometry (the positions and sizes of some electronic components are approximate only). The sensor telescope region is shown in greater detail.

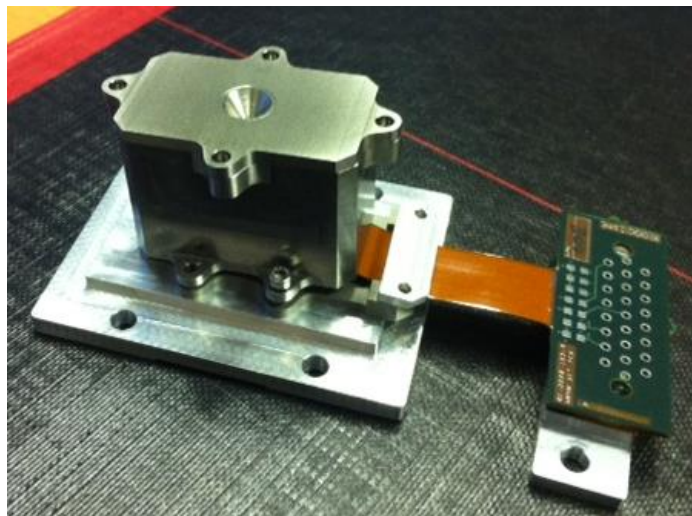


Figure 3.20: Photograph of the first HMRM prototype unit.

3.8 HMRM geometric response functions

Monitor characteristics which were estimated during the design phase may now be evaluated accurately using the detailed Geant4 simulation geometry. This includes calculation of the geometric response, quantified by the effective area parameter defined in Section 3.3.2. Whereas this was introduced previously as an idealised measure, this quantity can now be evaluated with realistic particles as a function of their primary kinetic energy, $a(E)$.

3.8.1 Energy-dependent effective area

Isotropic fluences of electrons and protons, simulated in Geant4 using the HMRM v1.0 geometry, were used to determine the fraction of hits as a function of particle energy and thereby calculate $a(E)$. This, for electron and proton primaries incident on sensor 1, is shown in Fig. 3.21. Several characteristics may be observed from these plots: the graphs confirm the low-energy detection limits for both species, below which particles fail to penetrate the aperture window and ASIC passive layers (as found earlier in Section 3.4).

Secondly, a plateau region is visible at intermediate energies for both particles. This is the energy range over which the aperture covering material is approximately transparent but where the casing still stops most incident particles. The $a(E)$ for both species here is close to the design value of 0.15 mm^2 (calculated using the ray-tracing model). The electron plateau is less well defined than for protons due to their increased scattering and straggling, which tend to smear out the transmission/attenuation behaviour.

At higher energies, the effective area increases again; this occurs due to particles penetrating the casing (at energies above 30 MeV and 2.5 MeV for protons and electrons, respectively). Note that the effective area is expected to reach a second plateau at 0.5 mm^2 (the value for a bare 1 mm^2 plane area) at sufficiently high energies. However, the second proton plateau is visible at $\sim 0.33 \text{ mm}^2$: this is because flux originating at the rear of the monitor was neglected in order to increase simulation efficiency. In actual operation, the host spacecraft would provide a large amount of shielding in this region, making this an acceptable approximation.

Figure 3.22 shows the effective areas for each of the four detection modes as a function of particle energy for both species. Note that the C1 function differs from the S1 function (the former imposes an anticoincidence condition on S2–4). The design aim of achieving approximately geometrically increasing thresholds for modes C2, C3 and C4 has been achieved satisfactorily, as indicated by the reasonably linear onset spacing on the log-scale plot.

Although each successive detection mode has a minimum (threshold) energy, there is no maximum energy. As discussed in Section 3.2.5, such a maximum would only be seen if inter-sensor losses were negligible. These losses may only be avoided under carefully controlled conditions, where the sensors are closely spaced or have increasing width towards the rear of the stack. This could not be achieved fully in the HMRM due to ASIC spacing requirements and the need to use a single ASIC design.

As predicted, the choice of the C3-defined aperture has provided an uninterrupted acceptance

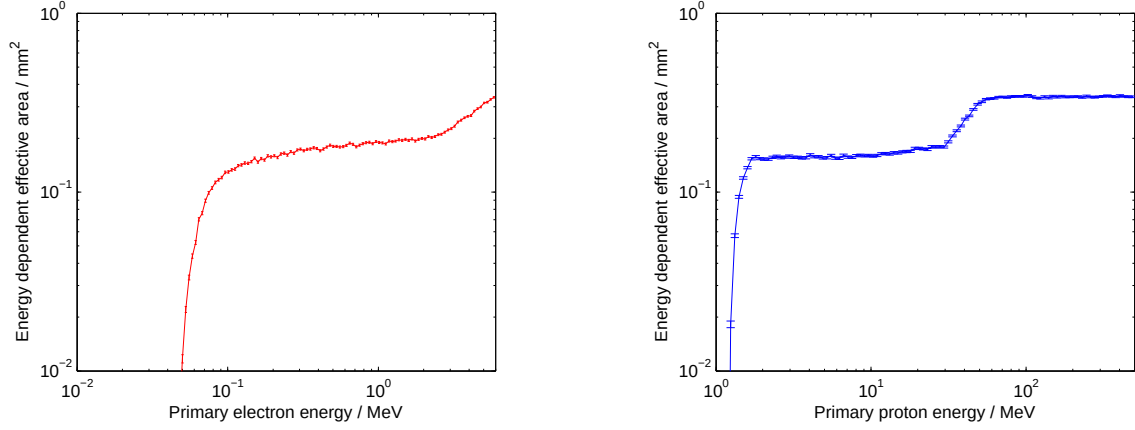


Figure 3.21: Energy dependent effective area for S1, evaluated via Geant4 simulation, for electrons (left) and protons (right).

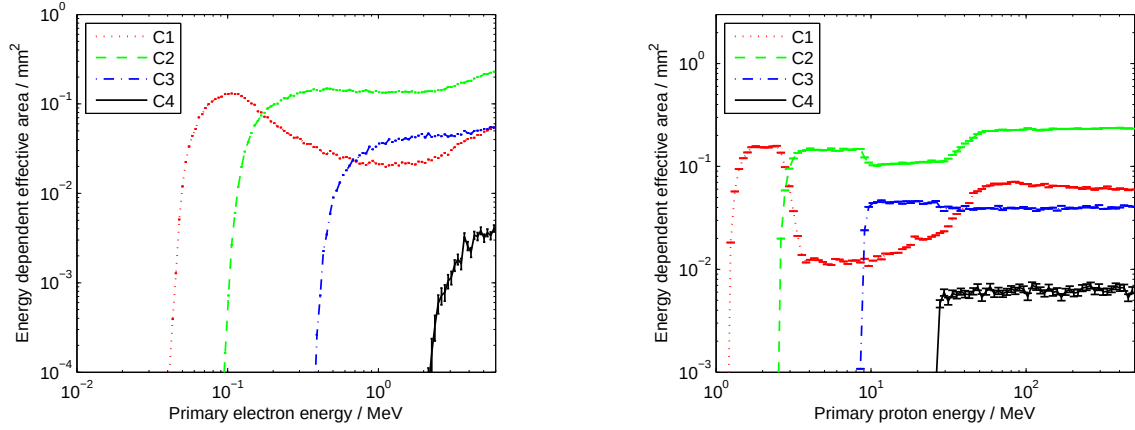


Figure 3.22: Energy dependent effective area for the four detection modes for electrons (left) and protons (right).

region for C3 and C4 events; this is seen in the flat effective area functions for protons for these two modes. In contrast, the C1 and C2 functions show increases around 30 MeV, when protons may penetrate the front casing to the sides of the aperture and still create a valid coincidence event.

The energy-dependent effective area for sensor 1 has been used to obtain more accurate estimates of the pile-up rates for the HMRM in the five reference orbits. This requires folding of the orbit energy spectrum with $a(E)$, integration over all energies, then summation over both species. This determines the total hit rate on S1 as a function of orbital time (fluxes are assumed to be isotropic at this preliminary stage) and is shown in Fig. 3.23.

The pile-up probabilities resulting from these hit rates are shown in Fig. 3.24, assuming an exposure time of 3.125 μ s (the minimum) per sensor readout. Satisfactory performance (pile-up less than 5%) is predicted in the commercially important orbits A, D and E. Orbits B and C (which are, respectively, a 400–4,000 km elliptical orbit and a 10,000 km MEO orbit) both

include periods where the pile-up probability, and hence particle misidentification probability, is in excess of 10%. Nevertheless, satisfactory performance is still expected.

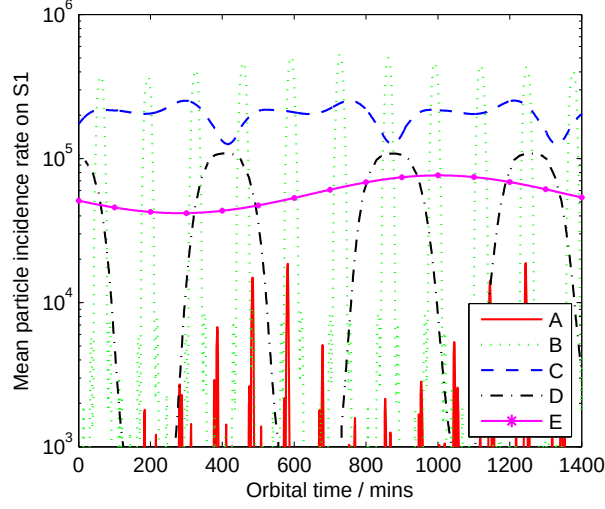


Figure 3.23: Predicted mean particle incidence rate on S1 in the five reference orbits as a function of orbital time. The calculation used the AP-8 and AE-8 flux models, together with the energy-dependent effective area for the monitor.

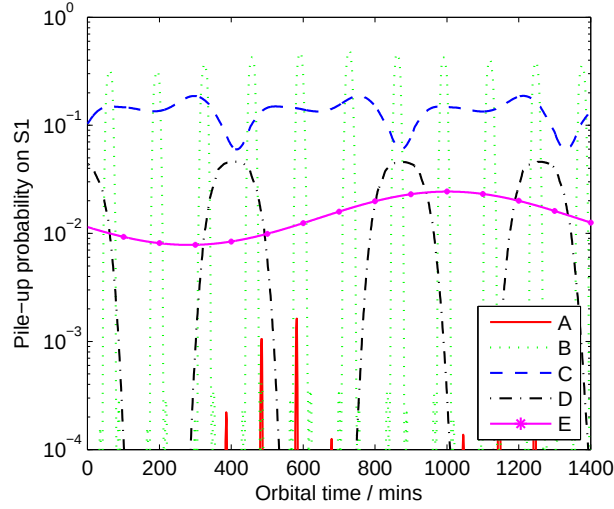


Figure 3.24: Predicted pile-up probability on S1 in the five reference orbits as a function of orbital time. The mean incidence rates shown in Fig. 3.23 were assumed.

3.8.2 Unidirectional effective area

The omnidirectional effective area, A_e , was introduced as an idealised quantity, calculated by assuming that particles travel as linear rays which are stopped completely by the casing and transmitted completely by the aperture and sensors. This concept was extended to the

use of simulated charged particles, providing analogous quantities which also include the effects of particle scattering and attenuation. This updated quantity, the energy-dependent omnidirectional effective area, $a(E)$, is thus naturally a function of species and energy.

Both effective areas (idealised and energy-dependent) have been calculated in terms of isotropic fluences. The general response will, however, show a strong dependence on fluence anisotropy, since the telescopic alignment of sensors favours detection of particles travelling close to the central axis. A second extension, therefore, is to express the effective area as a unidirectional quantity — this is now a function of direction, particle species and particle energy.

As before, the quantity may be calculated for hits to a single sensor or for a combination of sensors. This unidirectional energy-dependent effective area, $\alpha(E, \Omega)$, can be used to predict the mean incidence rate λ (to a specified sensor or mode) due to species s as:

$$\lambda = \int f_u^s(E, \Omega) \alpha^s(E, \Omega) d\Omega dE, \quad (3.19)$$

with f_u^s the unidirectional differential flux for species s . The total mean incidence rate is then the sum over all species present. Note also that the omnidirectional value is an angular average of the unidirectional function:

$$a(E) = \frac{1}{4\pi} \int \alpha^s(E, \Omega) d\Omega. \quad (3.20)$$

The effective area is crucial to particle identification and spectrum reconstruction, while the use of the unidirectional quantity allows treatment of anisotropic fluxes. These applications will be covered in detail in Chapters 4 and 6.

Selected parts of the $\alpha(E, \Omega)$ function for the HMRM are shown in Figs. 3.25 and 3.26 for protons, and Figs. 3.27 and 3.28 for electrons. In these plots, the elevation angle, θ , is measured from the monitor's principal axis, which runs perpendicularly through the centres of the sensors. The azimuthal angle, φ , is measured in the plane perpendicular to this axis; all modes show little variation with φ (see, for example, Fig. 3.25).

Particles were incident on the HMRM from all directions in these simulations, revealing the characteristics of ‘reverse’ coincidences. These occur when particles pass through the sensor telescope from angles near $\theta = 180^\circ$. Figure 3.28 shows that their probability is negligible for electrons up to 6 MeV. However, Fig. 3.26 shows that protons of energy ~ 100 MeV are equally likely to produce a C3 mode from the forward or reverse direction. Presence of the spacecraft behind the monitor (not included in these simulations) will raise the energy threshold for these events further still, while decreasing their flux. Future simulations will assess the frequency of these events with the addition of a spacecraft model.

3.9 Reduction of the number of sensors

Difficulties due to PCB design, ASIC power consumption and project time constraints were encountered during construction of the first HMRM prototype device. These resulted in the omission of ASICs D1 and D4 from the first prototype. The remainder of the monitor simulation and response modelling will address the original design, although much of the monitor algorithm and data analysis method extends naturally to any number of sensors.

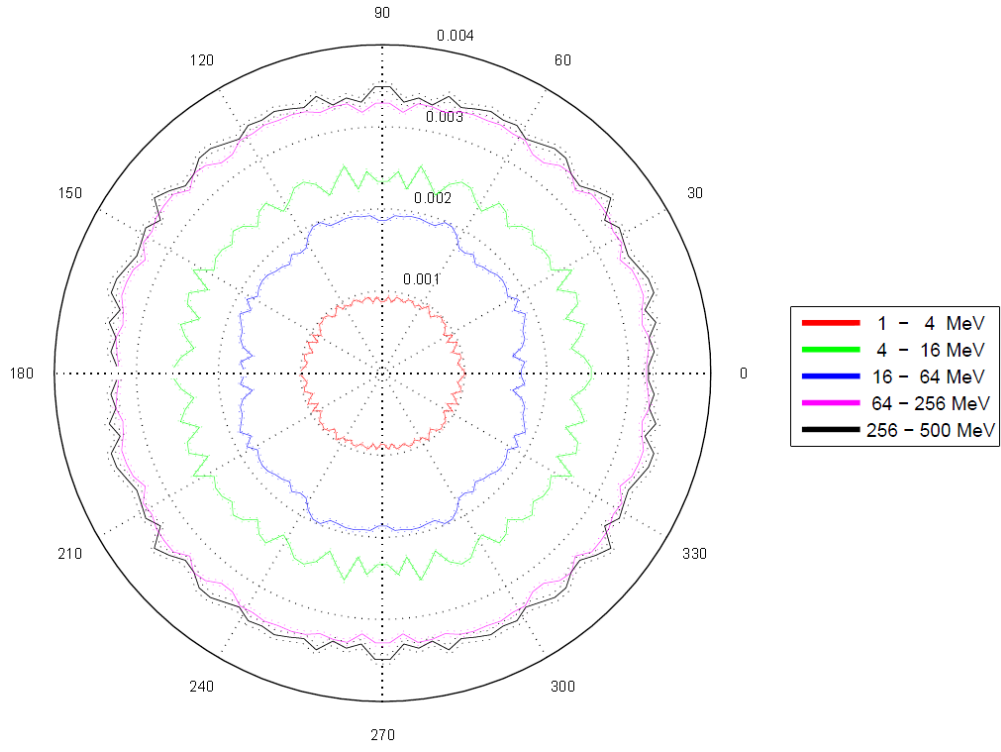


Figure 3.25: Azimuthal dependence (φ in degrees) of the HMRM unidirectional effective area for C2 mode protons of selected energy ranges. The effective area is given in units of cm^2 .

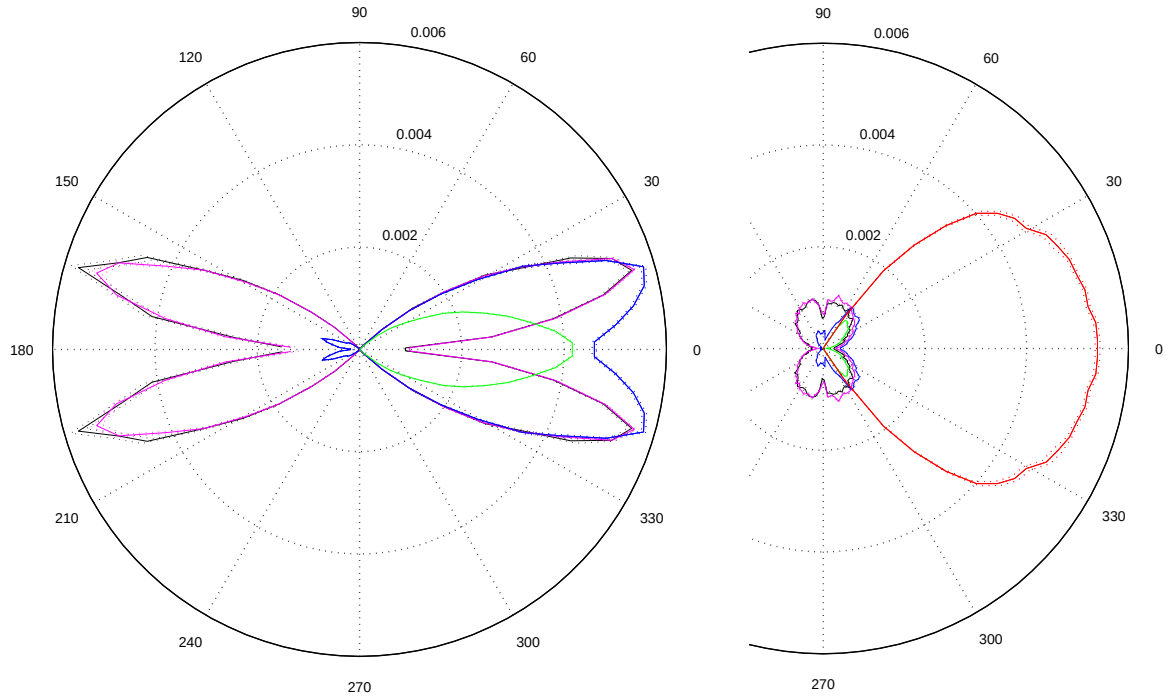


Figure 3.26: Elevation angle dependence (θ in degrees) of the HMRM unidirectional effective area for C3 mode protons (left) and C1 mode protons (right) of selected energies (see Fig. 3.25). The effective area is given in units of cm^2 .

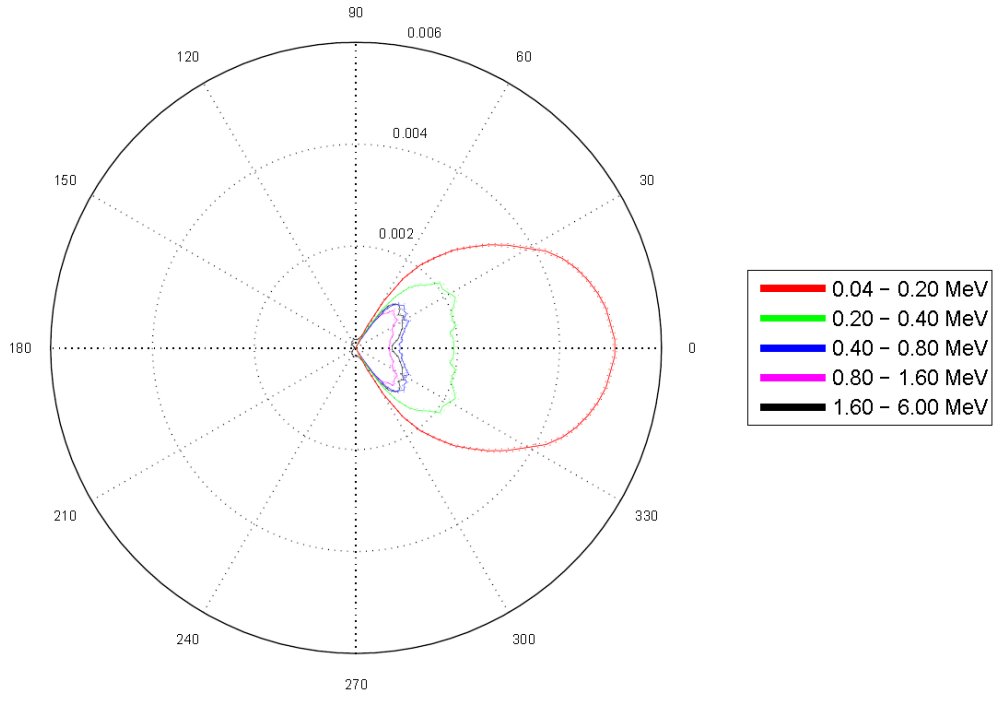


Figure 3.27: Elevation angle dependence (θ in degrees) of the HMRM unidirectional effective area for C1 mode electrons of selected energy ranges. The effective area is given in units of cm^2 .

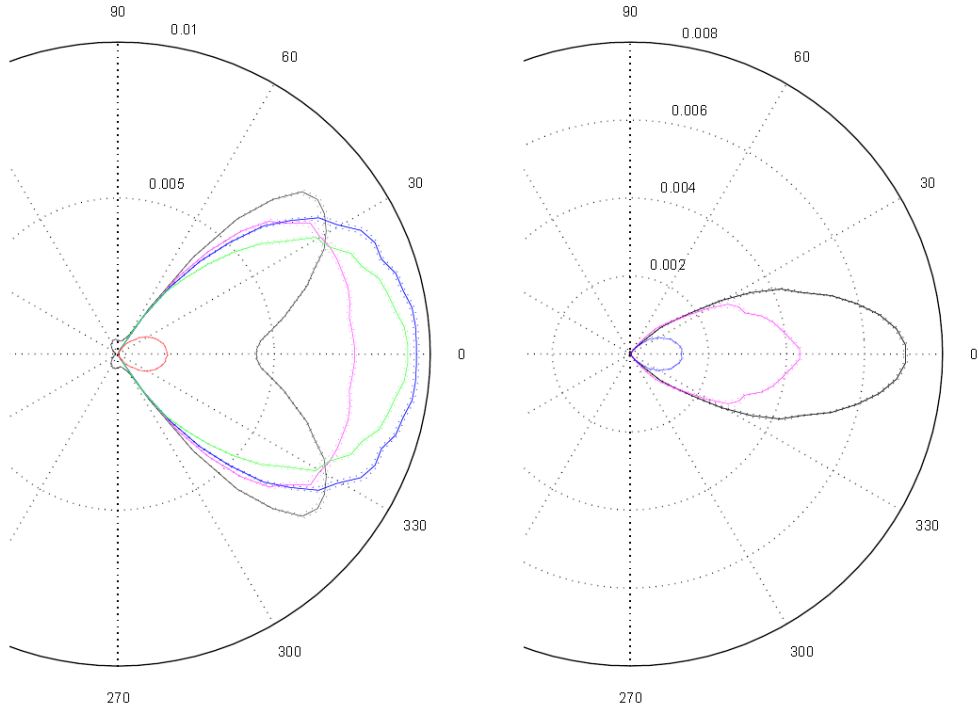


Figure 3.28: Elevation angle dependence (θ in degrees) of the HMRM unidirectional effective area for C2 mode electrons (left) and C3 mode electrons (right) of selected energies (see Fig. 3.27). The effective area is given in units of cm^2 .

Chapter 4

Radiation monitoring and event identification algorithms

This chapter describes the algorithms used by the HMRM to detect and classify particles. These algorithms are executed in real-time by an FPGA within the HMRM which analyses the sensor pixel data to produce useful measurements. These include the ionising dose rate and particle count rate, as well as outputs from an event identification algorithm. An initial assessment of the algorithms is made using Monte Carlo simulations, while full simulations of monitor operation in Earth orbits are given in Chapter 5. The reconstruction of particle energy spectra, performed as offline analysis in the ground segment, is considered in Chapter 6.

4.1 Introduction to HMRM operation

The purpose of the HMRM is to provide real-time measurements of the local particle flux. Information generated by the monitor is transmitted as data packets, which may be used by the host spacecraft to react to changes in the particle environment. Monitor output data may also be used, over longer time periods, in more detailed analyses. These are not intended to be executed within the monitor or spacecraft but rather on the ground, as part of the control and support system (referred to as the mission *ground segment*). These may include statistical methods to reconstruct the incident particle energy spectra, an application which is investigated in Chapter 6.

4.1.1 Monitor data products

The output measurements and calculation results (referred to as *data products*) should be of two types: (i) simple quantities, allowing immediate appraisal and potential incorporation into autonomous operations of the spacecraft; (ii) larger but more complete datasets which maximise information content (for ground segment use). The former requirement implies that some analysis of the raw sensor data must be done within the monitor to create useful data products. Furthermore, these quantities must give timely information about the potential for

radiation damage in the current environment. According to these requirements, the following quantities were chosen:

1. Particle count rate on the first sensor (chosen because it is most exposed and thus receives the highest rate).
2. The rate of ionising dose absorbed in each sensor.
3. Lifetime cumulative ionising dose absorbed in each sensor.
4. Identified particle count rates (classified by most probable species and kinetic energy).
5. Alert flags based on the rate measurements listed above.

Particle count and dose rates give simple indications of the intensity of the external particle flux. Threshold values may be chosen so that, if exceeded, the host spacecraft is alerted via a simple bit flag. Furthermore, by finding the dose rate for all four sensors (located at progressively greater depths of shielding), an indication of the penetrative ability of the radiation is obtained. The lifetime cumulative dose is a simple continuous integral of this dose rate and may assist in long-term mission planning. This may also help assess eventual changes in the sensor characteristics caused by radiation damage.

Where possible, events should be identified as being due to a specific particle (species and energy) and count rates determined based on these identifications. However, the necessity of real-time provision of these data, and the limited computational ability of the HMRM, means that particle identification must be attempted on an event-by-event basis, rather than a longer-term (and more complex) statistical analysis of accumulated data. Indeed, this was acknowledged in the project statement of work [87]. Any single event identification will have an associated uncertainty, or misidentification probability, which must be negligible for the identification to have a deterministic interpretation: this is a very challenging requirement for such a small monitor.

The quantities listed above are all suitable for immediate use, with minimal complexity of interpretation. In addition, a general means of characterising each sensor readout frame is required for later ground segment analysis.

4.1.2 ASIC readout cycle

The HMRM detects and identifies particles by measurement of their ionising energy losses in CMOS active pixel sensors. This sensor architecture involves the collection of ionisation charge in the pixel diodes during a period of charge integration. Each diode voltage must then be ‘read out’ (selected and amplified) and digitised within the ASIC. Continuous detection thus requires a continuously repeating cycle of exposure and readout; this contrasts with other detector architectures where digitisation and analysis are triggered by individual radiation events.

The HMRM ASICs are to be operated on a readout cycle with a fixed period of 100 μs and thus with an output frame rate of 10 kHz. Within this fixed-period readout cycle, the actual charge integration time may be altered in response to the radiation environment (e.g. shortened in

higher fluxes). This is achieved via the *electronic shutter* algorithm described in Section 4.2.3, which varies the exposure time by factors of two between 3.125 μs and 100 μs .

All 2550 3-bit pixel values must be transferred from each ASIC to the FPGA within the 100 μs cycle time, implying a total mean data transfer rate of 0.3 Gbit s^{-1} (for four sensors). Furthermore, analysis of each sensor readout must be completed within the 100 μs period to avoid data processing pile-up. Both of these constraints place critical requirements on the FPGA performance.

4.1.3 The HMRM FPGA

The computational device selected for use in the HMRM is an Actel Igloo AGL1000-CSG281, a flash-based FPGA with 10^6 system gates, over 24,000 D-type flip-flops and 18 kB of RAM [105]. A photograph of this component, after attachment to the HMRM central PCB, is shown in Fig. 4.1.



Figure 4.1: Photograph of the HMRM FPGA mounted on the monitor central PCB section. The FPGA package measures 10 mm by 10 mm by 1.05 mm.

This device is not a radiation-hard component, a compromise that was considered acceptable for use in the prototype monitor. Although radiation hardened FPGAs are available (for example, devices with 3 kGy (Si) total dose hardness are described in [106]), the range of commercially available radiation hard devices have higher power consumption and larger package sizes than the Igloo. Future versions of the HMRM may instead use a radiation-tolerant ASIC for control and processing functions.

The principal advantages of an FPGA are its potential for high speed combined with an ability to guarantee the order and duration of computational processes. This is especially important in the HMRM since the sensors are to be read out at a fixed frequency, with all analysis and control processes synchronised with the frame readout cycle. This cannot be so easily ensured with devices such as microprocessors, which are interrupt driven.

The small size and power of the chosen model are further advantages when used in a miniature monitor. However, these also result in computational constraints: floating point arithmetic and advanced mathematical functions are impractical, and the monitor algorithm must therefore be designed according to these limitations. The ability to reprogramme the FPGA using a personal computer is a further valuable advantage during the development phase.

The FPGA has multiple roles within the HMRM. These include:

1. **Programming and control of the sensor ASICs.** The FPGA generates the various clock and signal waveforms required to drive the ASIC operation. It also stores data tables which must be programmed into the ASICs. These include the trim and threshold registers, discussed further in Chapter 7.
2. **ASIC readout.** At the end of each sensor exposure period, the sensor pixel signals are digitised in the ASIC and transferred via a 9-bit bus (three pixels in parallel) to the FPGA.
3. **Data analysis.** The input digital pixel values are analysed and reduced to a set of calculated data products which are the monitor output data.
4. **Communications.** The FPGA must handle all communications between the HMRM and the spacecraft or spacecraft operator (mediated by the spacecraft computer). This includes the creation and transmission of telemetry data packets and the reception and execution of telecommands. All communications are carried out according to the Controller Area Network (CAN) 2.0 A/B protocol [107].

The data analysis functions of the FPGA are collectively referred to as the *physics algorithm*. These are the steps needed to convert the pixel values from each sensor into a set of meaningful data products. The remainder of the FPGA's functions, which do not directly influence the particle detection process, are not considered further here.

The physics algorithm [108] was developed as a flowchart of logical steps (presented in full in Appendix D) and was therefore not implementation-specific, although care was taken to ensure the level of complexity was suitable for the selected FPGA device. This algorithm flowchart was then transcribed into very high speed hardware description language (VHDL) by RAL Space for programming into the FPGA.

4.1.4 Data integration cycle

Although analysis of the sensor data must be completed within a single readout cycle, the data products must characterise the radiation environment over longer timescales for convenient interpretation. This involves the aggregation of data, followed by the calculation of a useful quantity, in a repeating cycle.

Distinct from the charge integration (pixel exposure) period, the *data integration period* is an integer number of HMRM readout cycles over which a specific data product is measured. For integral quantities (e.g. lifetime dose or orbital dose), the length of this period is well defined. However, differential quantities, such as the count rate, must be approximated as an average over a finite period. The choice of period is a compromise between time-resolution (frequency of calculation), the telemetry budget available for downlink and the counting statistics uncertainty, determined by the magnitude of the incident flux and the effective area of the monitor. A small monitor is necessarily at some disadvantage in this respect.

The requirement of measurement timeliness implies that data integration periods should be shorter than the characteristic time over which the environment changes. Proton flux at

the Earth after an SEP event, as described in Chapter 2, may rise rapidly, peaking within 30 minutes [15]. The flux then decreases more slowly, over many hours.

Furthermore, in LEO and MEO, fast changes in particle flux are caused by movement of the spacecraft with respect to the magnetic field (and therefore the particles trapped by the field). For example, Fig. 4.2 shows the omnidirectional integral proton flux, according to the AP-8 model [6], as a function of orbital time for 700 km and 23,000 km altitude orbits (reference orbits A and D). Orbit A shows a peak in flux once every 99 minute orbit. This occurs during the spacecraft passage through the SAA and the flux remains above 50% of the peak value for a period of approximately 5 minutes per orbit. Meanwhile, the orbit D flux exceeds 50% of its peak for periods of 40 minutes as the spacecraft moves through the wider proton belt near 0° latitude. Note that the monitor's effective area is approximately $2 \times 10^{-3} \text{ cm}^2$ for low energy particles (increasing for the more penetrating higher energies). An integration period of approximately 1–10 s is required to ensure that tens or hundreds of measurements are produced during these transient events.

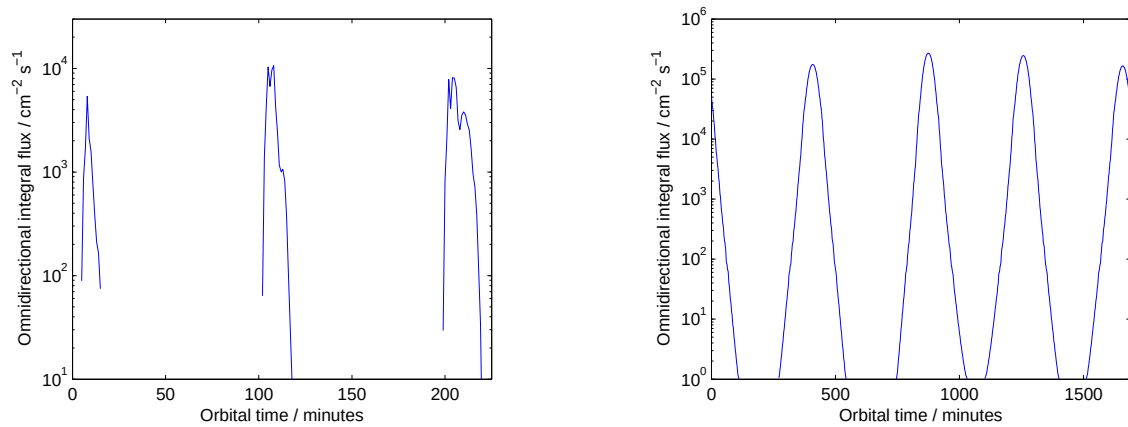


Figure 4.2: The omnidirectional integral proton flux exceeding 1 MeV as a function of time in orbits of 700 km altitude (left) and 23,000 km altitude (right). These are reference orbits A and D, respectively. The flux data are obtained from the AP-8 model [6].

4.1.5 Output data packets

A further consideration is the allowable, and desirable, frequency of data packet transmission. The HMRM output data have been divided between two data packet types, named *standard* and *extended*, both of which are transmitted to the spacecraft computer and may then be downlinked to Earth. Standard packets (48 bytes) are specifically intended for immediate use, containing rate and alert flag information, and are created at user-specified intervals of order 1 s. Extended packets are larger (128 bytes) and contain the remaining detailed monitor data. Requiring further analysis, they are not suitable for immediate interpretation and are to be transmitted at intervals of order 100 seconds. Both packet types follow the Consultative Committee for Space Data Systems (CCSDS) telemetry recommendations [109]. In addition, there is a third packet type (the *transparent frame* packet) which is only to be used for sensor

test and calibration (discussed in Chapter 7).

The provisional choices for the prototype monitor are for standard packet transmission every 2^{14} cycles (1.64 s) and extended packet transmission every 2^{20} cycles (104.9 s). These planned packet transmission periods are compatible with the predicted count rates and integration periods. For example, a data integration period of 1.64 s in a $10^6 \text{ cm}^{-2} \text{ s}^{-1}$ omnidirectional flux is expected to give a count rate measurement with less than 2% uncertainty (assuming a constant effective area of $2 \times 10^{-3} \text{ cm}^2$ for sensor 1 and neglecting pile-up). The suitability of these time periods depends on the characteristics of the particle environment and is further complicated by the dynamic alteration of the charge integration period (via the shutter control algorithm). A test of the time-dependent aspects of the monitor operation is provided by the orbit simulations presented in Chapter 5.

4.2 Physics algorithm

This section describes the algorithm used to analyse the pixel data and calculate the data product quantities. Discussion of the event characterisation and identification method, however, is postponed to the next section. Appendix D presents the entire HMRM physics algorithm in flowchart form, illustrating how the various subroutines are coordinated.

Due to the computational limitations of the FPGA, approximations are required in some of the data product calculations. Furthermore, scaled units are chosen to make efficient use of memory while avoiding floating point arithmetic and maintaining numerical precision: these must be taken into account when interpreting the data.

4.2.1 Digital pixel signal processing

At the end of each readout cycle, 2500 digital pixel values are obtained from each sensor ASIC by the FPGA (an additional 50 test pixel values are ignored). Each pixel signal is a 3-bit value, defined in the ASIC by a series of seven analogue comparator levels.

A single particle hit is, in general, expected to cause a non-zero signal in multiple pixels. This is due to the particle's path in the sensor and also due to charge diffusion or sharing between adjacent pixels. Thus despite the low energy resolution per pixel, many unique event patterns are possible across the pixel array.

For estimation of the deposited dose, the pixel values must be summed after applying an appropriate weighting for each of the eight possible values (a weighting is necessary because the comparator levels are not linearly spaced). The weighted sum provides a single reconstructed energy deposit value for each sensor (which may be zero) from which all of the data products are derived. Comparator level and pixel weighting selection are discussed in Section 4.4.2.

To prevent numerical overflow within the FPGA in the event of sensor saturation, only the first twenty-five non-zero pixel signals read from each sensor are summed and subsequent pixels are discarded. Charge sharing is expected to produce a cluster of up to nine non-zero

pixels per hit, in normal circumstances, although this number may be revised after sensor testing.

Cluster imaging

Some event information is lost in the summation process, primarily the shape of the pixel cluster and indeed whether there was more than one such cluster on a single sensor (indicating particle pile-up). Pixel cluster shape analysis has been investigated as a means of particle identification [110]. However, the image reconstruction algorithms used by this method require significant computational ability, which is not achievable with the HMRM FPGA or other similar devices. Furthermore, the cluster shape is dependent on the hit location with respect to the pixel diode locations, complicating the analysis. Modelling of the cluster shape also requires detailed knowledge of the charge diffusion characteristics of the ASIC, which were not fully understood during the design phase. The inability to detect the number of clusters also determines the need for statistical pile-up inference, rather than direct measurement.

Pixel mask

Before summing the pixel values, each individual pixel may be rejected according to the status of a mask bit held in the FPGA memory. This mask is used to ignore pixels which are identified as abnormal during the sensor calibration. Clearly, if a large proportion of pixels are masked, monitor performance will suffer: this effect is analysed in Section 4.4.1.

4.2.2 Sensor 1 particle count rate

The number of ‘empty frames’, E , where the pixel sum for sensor 1 was zero (all pixels below lowest threshold) may be counted over a series of m readout frames. The fraction of non-empty frames, f , is then given by:

$$f = \frac{m - E}{m}. \quad (4.1)$$

However, in general, this is not proportional to the particle incidence rate. This is due to the occurrence of pile-up: two or more particles hitting any combination of sensors in one exposure period cannot be distinguished from a single particle. This means that the interpretation of Equation 4.1 as indicating a particle incidence rate will always result in an underestimate of the true rate.

A more accurate estimate is obtained from Poisson statistics by relating the mean incidence rate, λ , to the probability of an empty frame, P_0 , by:

$$\lambda = \frac{-\ln(P_0)}{t} \quad (4.2)$$

$$\approx \frac{-\ln\left(\frac{E}{m}\right)}{t}, \quad (4.3)$$

where E/m is the observed fraction of empty frames and t is the exposure time of one frame. Note that the probability of a false hit detection has been assumed to be negligible. Equation 4.3 makes a statistical correction for the appearance of pile-up frames and provides an

unbiased estimate of the particle incidence rate. However, calculation of the logarithm is not easily achieved by the FPGA and Equation 4.1 must in fact be used; the accuracy of this approximation must therefore be quantified.

Figure 4.3 compares the result of Equation 4.3 with the simplistic approximation over a λ range of 10^3 s^{-1} to $2 \times 10^5 \text{ s}^{-1}$, using both the maximum and minimum available charge integration periods. For an approximate sensor 1 effective area of $2 \times 10^{-3} \text{ cm}^2$ this corresponds to an omnidirectional flux range of $5 \times 10^5 \text{ cm}^{-2} \text{ s}^{-1}$ to $10^8 \text{ cm}^{-2} \text{ s}^{-1}$, a relevant choice for the HMRM operation.

When a charge integration time of $3.125 \text{ }\mu\text{s}$ is used, the approximation error is just 3% at $10^7 \text{ cm}^{-2} \text{ s}^{-1}$, increasing to 25% at $10^8 \text{ cm}^{-2} \text{ s}^{-1}$. This is judged to be an acceptable approximation for use in the HMRM. Note that use of the approximation with the longer integration time results in a significantly less accurate result (57% error at $10^7 \text{ cm}^{-2} \text{ s}^{-1}$ flux). It is thus important that the charge integration (pixel exposure) time can vary in response to changes in the flux magnitude.

Since m is chosen to be a power of two ($m = 16384 = 2^{14}$), multiplication/division in the FPGA are achieved quickly via simple bit shift operations. In the final implementation, the calculated mean rate value is also scaled to avoid the need for a floating-point number representation.

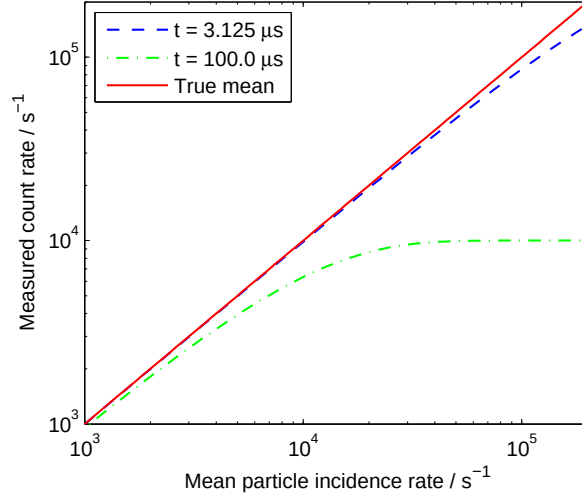


Figure 4.3: The particle count rate obtained by counting empty readout frames. This is shown compared to the true mean incidence rate for two values of t , the pixel charge integration time per readout. These are the maximum and minimum integration times available.

4.2.3 Electronic shutter control

The time period, $t(n)$, during each $100 \text{ }\mu\text{s}$ readout cycle over which charge is integrated in the sensor pixels, is given by:

$$t(n) = \frac{T}{2^{x-n}}, \quad (4.4)$$

where the integer shutter index, n , varies over the range $0 - x$ and T is the maximum integration time (equalling the readout frame period). The nominal values are $x = 5$ and $T = 100 \mu\text{s}$, so that $t(n)$ ranges from $3.125 \mu\text{s}$ to $100 \mu\text{s}$ in six steps.

The measured frequency of empty frames can be used to determine the pile-up probability and thus indicate whether to decrease the shutter index in response to a large flux. Conversely, in low particle fluxes, the period should be lengthened to maximise the fraction of the time that the monitor is sensitive to particles. Although the pile-up probability is given by:

$$P(k \geq 2) = 1 - P(k \leq 1) \quad (4.5)$$

$$= 1 - (1 + \lambda t) \exp(-\lambda t), \quad (4.6)$$

this is unnecessarily complex for the FPGA to calculate efficiently. Instead, the chosen pile-up *probability* thresholds (both a maximum and minimum, the minimum determining when to increase the shutter index) may be used to derive empty frame *number* thresholds. For a given data integration period, m , this is:

$$P(k \geq 2) = 1 - P(k = 0) + P(k = 0) \ln[P(k = 0)] \quad (4.7)$$

$$\approx 1 - \left(\frac{E}{m}\right) \left(1 - \ln \left[\frac{E}{m}\right]\right). \quad (4.8)$$

This is shown in Fig. 4.4, where E is plotted as a function of pile-up probability, having chosen $m = 16384$. For example, selecting pile-up thresholds of 5% and 0.5% implies empty frame thresholds of 11485 and 14773; these values are calculated during calibration and loaded via telecommand into the FPGA. In operation, the measured E is simply compared with these two specified limits, indicating whether to increase or decrease the shutter. The shutter index, n , may therefore change by ± 1 step after every m frames. Note, however, that n remains constant during the same m frames over which the count rate is calculated.

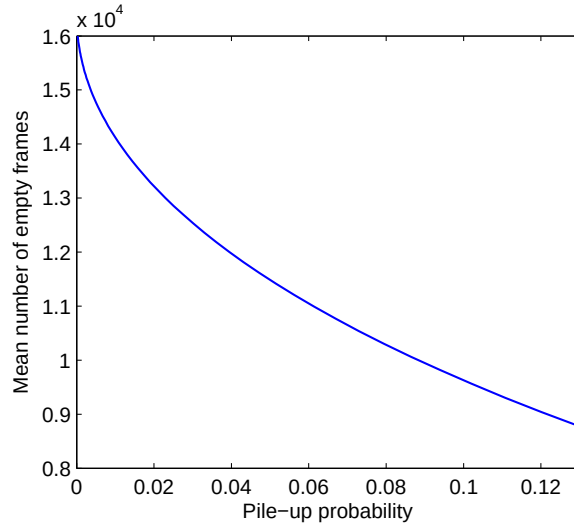


Figure 4.4: The mean expected number of empty frames (in a series of $m = 16384$ readouts) as a function of the pile-up probability.

4.2.4 Dose rate and cumulative dose

At the end of each readout cycle the energy deposited in each sensor is reconstructed via a weighted sum of the digital pixel values. This sum is then scaled to account, on average, for variation of the integration time caused by the electronic shutter. This must be done on an event-by-event basis since the dose data integration period, a series of z readout frames, may span multiple shutter index settings ($z > m$ is possible). The scaled and normalised dose rate for sensor i , R_i , is given by:

$$R_i = \frac{1}{z} \sum_{j=1}^z \frac{S_{ij}}{2^{(n_j-x)}}, \quad (4.9)$$

where S_{ij} is the weighted pixel sum for sensor i on readout j and n_j is the shutter index in use during that readout. This definition means that the dose rates R_i are interpreted as the mean dose per 100 μs of exposure. The dose units depend on the pixel value weightings, which are scaled to make the most efficient use of memory; further details are given in Section 4.4.2.

The dose rate integration period (a series of z readout frames) may vary depending on the flux environment. This is achieved by specifying both a minimum desired number of hits and a maximum allowable time period. Integration proceeds until the minimum count is reached (implying an approximate maximum statistical uncertainty in the result). However, this could result in unacceptably long integration times during periods of low flux, so the process also ends after the maximum time elapses. When either limit is reached, the dose rate is calculated and the integration restarted.

The lifetime cumulative dose, C_i , for each sensor is given by:

$$C_i = \sum_j S_{ij} 2^{x-n_j}, \quad (4.10)$$

a simple continuous sum of the scaled sensor dose.

4.2.5 Particle identification count rates

The information obtained in each readout cycle, comprising the four recovered energy deposit values, must be used to make a particle identification or otherwise characterise the event. This is to be divided into thirty-two categories, or *channels*, each with a separate 23-bit event counter. Particle energy spectrum reconstruction in the ground segment benefits from a large number of channels; however, all of these 23-bit channel counts must be downloaded in each extended data packet. The requirement for the monitor to have low data bandwidth led to the choice of 32 channels, although this may be later revised without affecting the general algorithm method.

Of the 32 channels, a small number are chosen which reliably identify the incident particle type (species and energy), each of which is used to generate a count rate measurement. In the present implementation, three channels are selected for this process (these count rates are included in the standard packet and thus the telemetry rate is again the limiting concern).

The count rate, U_k , is given by:

$$U_k = \frac{J_k}{\sum_{j=1}^p 2^{n_j}}, \quad (4.11)$$

where J_k is the event count for channel k over a total of p cycles (thus $0 \leq J_k \leq p$) and n_j is the shutter index during the j^{th} readout. This approximate calculation neglects the effects of pile-up and misidentification. The result should be multiplied by 3.2×10^5 to obtain the mean counts per second.

The remainder of the thirty-two channel counts are not normalised in this way and are instead intended for statistical analysis in the ground segment. The method by which events are allocated to the thirty-two channels is referred to as the *event identification algorithm*.

4.3 Event identification algorithm

The information obtained in each readout frame (four energy deposit measurements) must be recorded in a way which allows later analysis of the data. This should preserve as much information as possible without exceeding the practical limits imposed by the FPGA memory and spacecraft telemetry budget. Storage and transmission of the four values from each frame is certainly not possible: for example, representation of each value by a 17-bit integer (as used by the FPGA) would necessitate a mean data transmission rate of 0.68 Mbit s^{-1} , whereas the HMRM should average a rate of order 100 bit s^{-1} . A method of data reduction is therefore required.

4.3.1 Pulse height analysis

Pulse height analysis is a well-established method which involves the histogramming of the amplitude of analogue pulses from a sensor. The allocation of an event to a specific pulse-height ‘bin’ categorises the energy deposit and may strongly suggest a particle identity for that event. Moreover, creation of a pulse height spectrum over many events allows a statistical interpretation of the particle environment.

The pixel sum obtained for each HMRM sensor is analogous to a pulse height measurement. However, the pulse height histogramming of each sensor deposit in isolation does not take into account the extra information provided by simultaneous deposits in the other sensors. This coincidence information must be used to assist with particle identification, since a deposit in a single sensor is highly ambiguous.

4.3.2 4-D histogram

The simplest extension to the standard analysis method is to simultaneously histogram the four energy deposit values. This is a four-dimensional (4-D) histogram, with the criterion that each variable must lie within a range that forms part of a linked set of four bins. In this scheme, a separate event count is kept for each 4-D bin, which is equivalent to a traditional pulse height channel. This method is robust and fast enough to be executed within the $100 \text{ } \mu\text{s}$

limiting time for each event analysis in the HMRM FPGA. Crucially, it also preserves the coincidence information of each event. As a general method it is also scalable to any number of sensors and any number of histogram channels, although telemetry and processing budgets have provided the limit of thirty-two channels for the HMRM prototype. The majority of the channels will only be used in ground segment analysis, allowing construction of the energy deposit spectrum and providing input data for particle spectrum deconvolution methods (discussed further in Chapter 6).

The requirement of real-time particle identification, however, means that some channels should be interpreted as belonging to specific particle types (species and/or kinetic energy). The success of this identification method depends on the selection of the 4-D energy deposit histogram bin edges, collectively referred to as the identification look-up table, or *ID table*. Accurate Monte Carlo simulations are invaluable in this process as they model the full energy deposit variance and give the true identity of each event. However, the table may also be chosen to incorporate results from experimental tests with, for example, monoenergetic particle beams.

4.3.3 Energy deposit fluctuation

Figure 4.5 shows the Geant4 simulated spectrum of energy deposits in a 12 μm silicon layer resulting from a selection of monoenergetic particles at normal incidence.

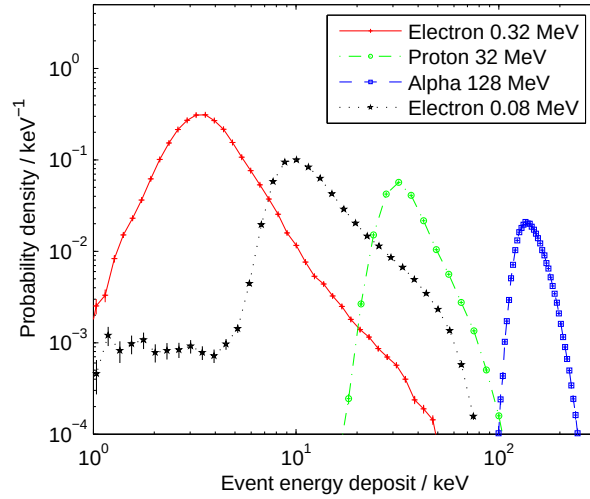


Figure 4.5: Energy deposit probability distributions in a 12 μm silicon layer due to various particle types (species and energy combinations) at normal incidence.

This is a highly idealised situation: no other material volumes were included in the simulation and the detector response is assumed to be perfect. However, these results illustrate a fundamental limitation inherent in the identification of particles by sampling of their energy deposits in thin layers. Even in the absence of fluctuations caused by variations in primary energy or incidence angle, a large range of deposits is possible from each particle type. As

introduced in Chapter 2, this is a result of the stochastic nature of the energy loss processes and path length variations due to scattering in the detector.

Energy loss fluctuation alone is sufficient to limit the resolution of particle kinetic energy measurements, since similar energies of the same species give overlapping deposit distributions. This is demonstrated in Fig. 4.5 where, for example, an energy deposit of 7 keV has approximately equal probability of being due to an electron of 0.08 MeV or 0.32 MeV.

4.3.4 Particle species ambiguity

In addition to ambiguities between particles of different energies, different particle *species* may produce the same energy deposits. This can be observed clearly by plotting mean energy deposit as a function of primary particle energy, for electrons, protons and alpha particles ($^4\text{He}^{2+}$ ions). This is shown in Fig. 4.6, where the error bars indicate the central 68% range of the energy deposit distribution at selected particle energies. Note that these data were again obtained through simulation of particles onto an isolated 12 μm silicon layer at normal incidence; the most probable MIP energy deposit for this arrangement is approximately 3.5 keV. Variation of the incidence angle and the inclusion of external materials (which modify particle energy) will generally increase the deposit variance for any given particle type.

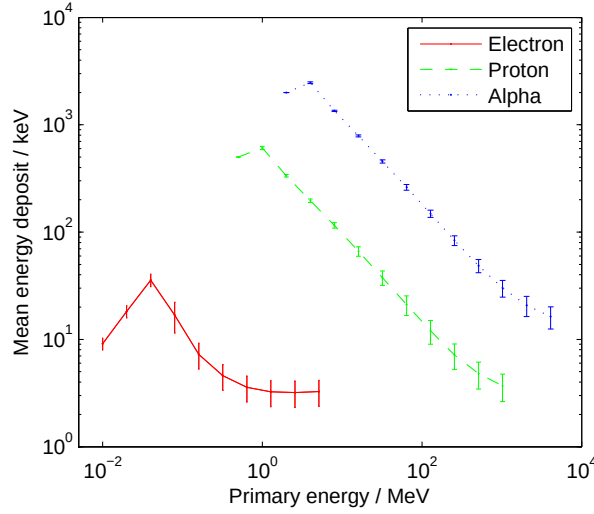


Figure 4.6: Mean energy deposits in a 12 μm silicon layer due to protons, electrons and alpha particles at normal incidence, as a function of the particle primary energy. The error bars indicate the central 68% range of the energy deposit distribution.

4.3.5 HMRM simulation

To advance to a more realistic situation, the full HMRM simulation geometry is exposed to isotropic fluences of electrons and protons using the Geant4 toolkit. Details of the generation position, direction, species and energy are stored for each primary particle, allowing

later analysis. Each primary (and any secondaries generated) are fully tracked through the geometry before the total ionising energy deposit in each sensor is recorded (the *event* energy deposit). The primary energy for each species is selected randomly from an E^{-1} differential energy spectrum, allowing efficient sampling across the chosen energy ranges (0.04–6 MeV and 1–500 MeV for electrons and protons, respectively).

The sensors now receive incident particles from all angles, limited only by the casing and aperture configuration. Figure 4.7 shows the energy deposit distributions in sensor 1 (S1) resulting from a range of particle primary energies. Note that these energy deposit data are directly generated by Geant4 and that device response models (introducing noise and limited resolution) have not been applied at this stage.

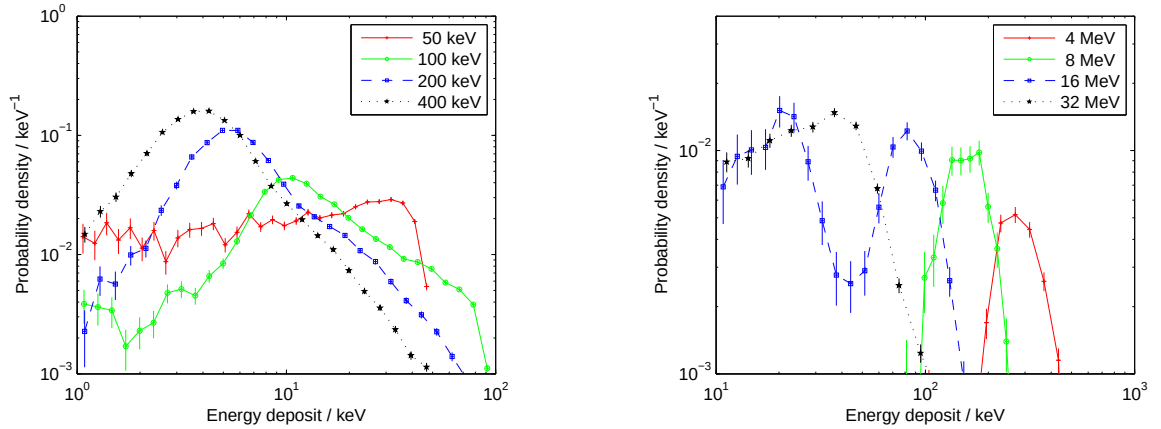


Figure 4.7: Simulated energy deposit distributions in sensor 1 of the HMRM due to isotropic monoenergetic fluences of electrons (left) and protons (right) at selected primary energies.

The result of the isotropic fluence is as expected; energy deposit variance for each particle has increased due to the greater range of possible trajectories through the sensor. In addition to this, higher energy particles penetrate the casing and thus arrive at the sensor with modified energy, causing significantly different deposits to those entering via the aperture. This may be observed, for example, with 16 MeV protons, which show a bimodal distribution of energy deposits. A further general observation to be made from Fig. 4.7 is that electrons have greater energy deposit variance than protons (due to their increased angular scattering).

The identification problem may be illustrated by plotting energy deposit as a function of particle primary energy in a mixed species environment. This is shown for energy deposits in S1 in Fig. 4.8, having simulated isotropic fluences of electrons and protons. Both species were generated according to a power-law (E^{-1}) differential energy spectrum.

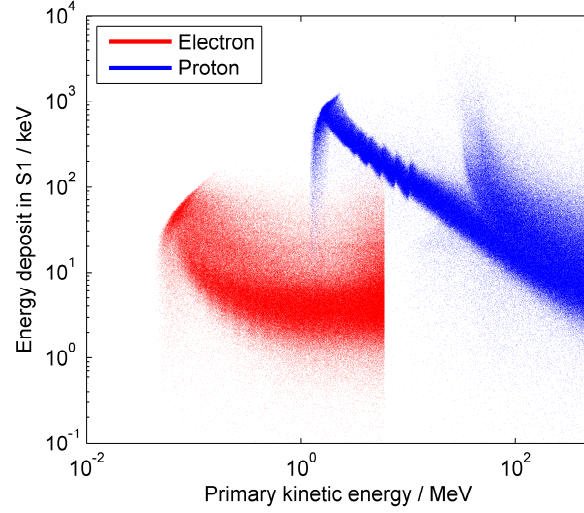


Figure 4.8: Simulated energy deposits in sensor 1 of the HMRM due to isotropic fluences of electrons and protons. Each point marks the event energy deposit at the primary particle energy responsible for that event; the colour coding indicates the particle species.

For the species and energies simulated for Fig. 4.8, events with an energy deposit of more than 200 keV are reliably identifiable as protons. In contrast, smaller deposits may be due to either species. There is a difficulty, however, in applying this observation in general. Unless all species and all energies are simulated, no such definite statements can be made; instead, an assumption is made regarding which particles are present in significant numbers in the radiation environment. Higher energy electrons are not expected to produce greater energy deposits than the MIP deposits (~ 4 keV) shown. Although radiative energy losses increase at higher energies, the photon absorption in such thin sensors is predicted to be negligible.

Protons between 2 MeV and 20 MeV show a consistent power-law relationship between primary energy and most probable energy deposit (with deposit proportional to $E^{-0.02}$). Above this energy, the proton deposit variance increases due to particles penetrating the casing to the sides of the aperture. At higher (relativistic) proton energies the deposits will reach the MIP value and thus be indistinguishable from electron MIPs.

The monitor detector modes C1–4, introduced in Chapter 3, are four general ways in which particles may interact with the four sensors. These are identified as especially significant because each mode implies a minimum particle energy (for a given species). Three of these modes (C2–4) involve sensor coincidence, which constrains the particle incidence angles and may therefore decrease energy deposit variance.

C1 mode

The data for Fig. 4.9 are produced from the same simulation as Fig. 4.8 but with the additional requirement that the event selection is limited to those depositing energy in S1 and not in the other sensors; this corresponds to the C1 detection mode.

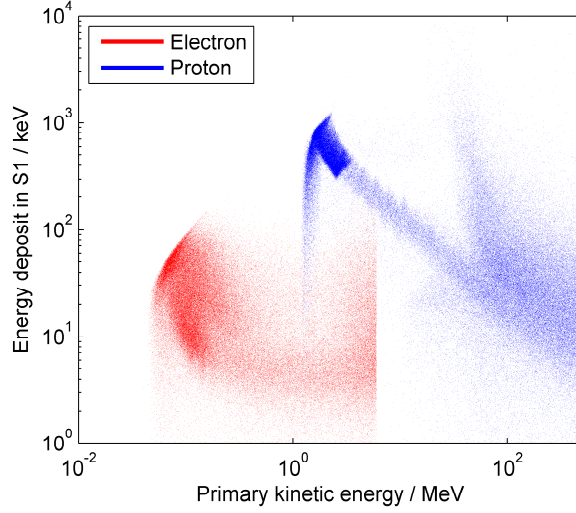


Figure 4.9: Simulated energy deposits in sensor 1 of the HMRM due to isotropic fluences of electrons and protons, when constraining event detection to the C1 mode only.

The relative identification probability for particles with low kinetic energy is now enhanced as their small penetration makes a C1 event more likely. This applies to protons below 3.5 MeV and electrons below ~ 200 keV. The association of low energy particles with the C1 mode is, however, probabilistic. A deterministic identification cannot be made because higher energy particles may still cause C1 events — this is due to the scattering loss of particles between S1 and S2, and due to range straggling.

C2 mode

Choosing events that only deposit in the first two sensors selects the C2 detector mode. This is shown in Fig. 4.10, where S1 energy deposit is plotted against S2 deposit for each event. This graph shows three main event regions, comprising pure proton events at large energy deposits, pure electrons at small deposits and a mixed intermediate region. The overlap covers a greater proportion of the electron distribution than that of the protons, making reliable electron identification less likely.

Secondly, in Fig. 4.11, the distribution due to each species is shown divided into primary energy bins. In each case, these are plotted together with all events due to the other species, indicating the extent of discrimination between species and energy. Some separation of proton energies is observed, although the electron energies are mostly indistinguishable.

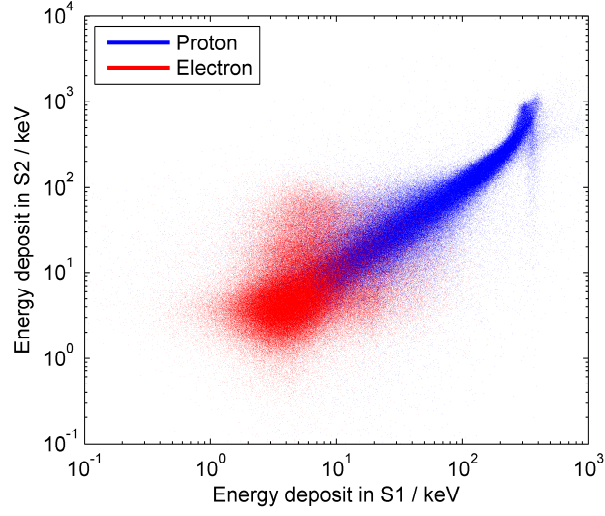


Figure 4.10: Distribution of coincident energy deposits in HMRM sensors S1 and S2 for C2 mode events arising from simulated isotropic fluences of electrons and protons.

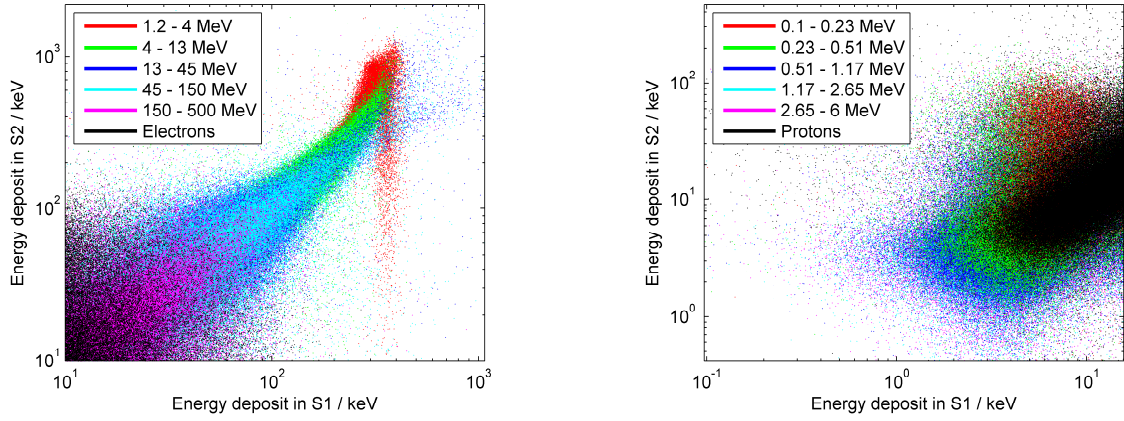


Figure 4.11: Distribution of coincident energy deposits in HMRM sensors S1 and S2 for C2 mode events due to selected proton (left) and electron (right) energy ranges. In each case, the species not divided into energy ranges is also included as a single dataset to illustrate the region of parameter space where the two species are largely indistinguishable.

C3 mode

Graphs analogous to those shown in Figs. 4.10 and 4.11 become progressively more difficult to draw, and interpret, as the number of sensors involved increases. Instead, C3 events are plotted by their energy deposits on S2 and S3 only, ignoring S1. This is shown in Fig. 4.12.

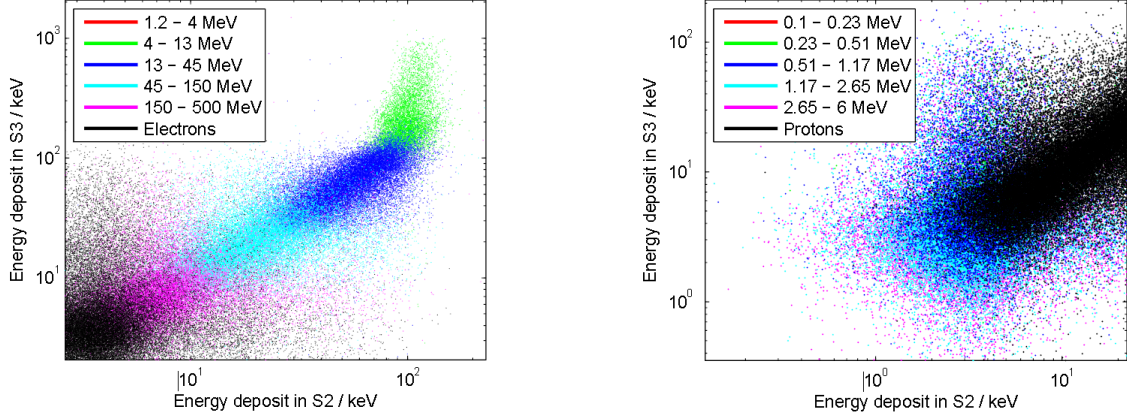


Figure 4.12: Distribution of coincident energy deposits in HMRM sensors S2 and S3 for C3 mode events due to selected proton (left) and electron (right) energy ranges. In each case, the species not divided into energy ranges is also included as a single dataset to illustrate the region of parameter space where the two species are largely indistinguishable. Energy deposits in S1 are ignored except for coincidence mode identification.

Again, there is limited separation of species, together with good discrimination of proton energies up to ~ 100 MeV. However, all electron energies result in similar deposit values, meaning that they cannot be distinguished.

C4 mode

Electron selection efficiency by the C4 mode is significantly less than that for protons. This is due to the inclusion of the inter-sensor degrader between S3 and S4, which causes most electrons to scatter out of the telescope. The proton plot for C4 is shown in Fig. 4.13, again showing energy discrimination in the 10–500 MeV range.

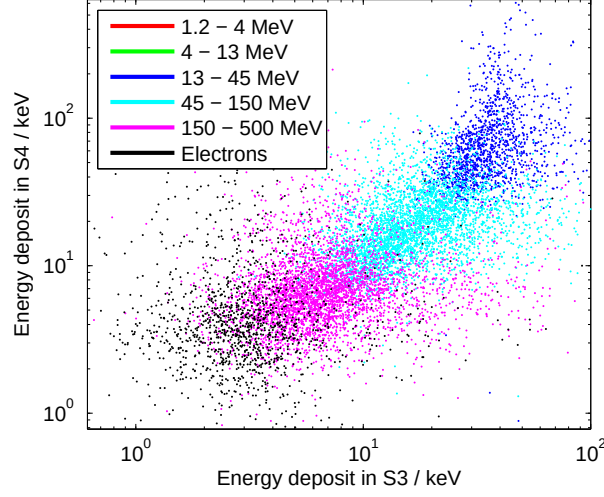


Figure 4.13: Distribution of coincident energy deposits in HMRM sensors S3 and S4 for C4 mode events due to selected proton energy ranges. Energy deposits in S1 and S2 are ignored except for mode identification. All energy deposits due to electron primaries are shown as a single dataset, illustrating their similarity to high energy proton deposits.

4.3.6 Identification efficiency and purity

As previously recognised, the notion of identifying a single event requires that the particle identification uncertainty be negligible. The plots presented above show that in general, over the wide range of electron and proton energies considered, this is not possible through energy deposit measurement. Most combinations of sensor energy deposits may result from either of the particle species studied. Single particle identification is therefore necessarily limited to the few exceptional instances where a single species is 100% probable, within the uncertainties of the simulation. Other events may be assigned a weaker, probabilistic identification; these may be further strengthened by the application of prior knowledge concerning the radiation environment.

When selecting the ID table channel definitions (the 4-D energy deposit bins) two figures of merit are introduced: identification purity (ν_k) and identification efficiency (ϵ_k). These are used to evaluate the suitability of a candidate channel, k , and are defined as follows:

$$\epsilon_k = \frac{\text{Number of particles identified}}{\text{Total number of viable particles incident}} \quad (4.12)$$

$$\nu_k = \frac{\text{Number of correct identifications}}{\text{Total number of events identified}} \quad (4.13)$$

Here, ‘identified’ denotes allocation to channel k , while ‘viable’ is defined to mean that the particle created one of the four detection mode patterns (C1–4). This definition of viability acknowledges the fact that a monitor cannot be expected to detect all incident particles.

A ‘correct identification’, used in Equation 4.13, indicates that the particle allocated to channel k was correctly described by the nominal particle specification for that channel, a concept

used only in the purity definition. The efficiency indicates the probability of particle allocation to channel k , given that the particle is incident in one of the four principal detection modes. In contrast, the purity is the probability that a particle allocated to channel k is correctly described by that channel's nominal particle identity (species and kinetic energy).

Identification efficiency

In conditions of negligible pile-up, the identification efficiency is expected to be independent of the incident particle flux (each particle interaction event is independent). It is therefore an objective quantity, independent of the external environment and of any choice of nominal identity for a channel. This means that it is characterised solely by the monitor geometry, sensor response and identification algorithm.

The unidirectional effective area introduced in Chapter 3, $\alpha_c^s(E, \vec{\Omega})$, gives the rate of viable mode c hits, r_c^s , due to species s , by:

$$r_c^s = \alpha_c^s f_u^s, \quad (4.14)$$

where f_u^s is the unidirectional flux of this species and r_c^s is differential in direction ($\vec{\Omega}$) and energy (E). This differential hit rate is related to the channel k allocation rate, r_c^s , by the identification efficiency, according to:

$$n_{k,c}^s = \epsilon_{k,c}^s r_c^s \quad (4.15)$$

$$= \epsilon_{k,c}^s \alpha_c^s f_u^s, \quad (4.16)$$

where $\epsilon_{k,c}^s$ is the unidirectional differential efficiency for species s , incident in mode c , to be allocated to channel k . The mean number of counts, μ_k , expected in channel k over a time period of T is:

$$\mu_k = \sum_s \int_0^T \int_{4\pi} \int_0^\infty \epsilon_{k,c}^s \alpha_c^s f_u^s dE d\vec{\Omega} dt, \quad (4.17)$$

where the sum is over all species present. A channel suitable for real-time particle identification is one with zero efficiency for all species and energies other than the nominal identity for that channel. Most channels, however, will have non-zero detection efficiencies for multiple particle types, preventing simplistic particle identification. In addition, there may be multiple channels that share non-zero identification efficiency for a single particle type. This introduces the possibility of deconvolving the incident flux of these particles by correlating the count rates in these shared channels.

Identification purity

The unidirectional identification purity for channel k may be written in terms of the efficiency as:

$$\nu_k = \frac{\sum_{c=1}^4 \int_{E_1}^{E_2} \epsilon_{k,c}^p \alpha_c^p f_u^p dE}{\sum_{c=1}^4 \sum_s \int_0^\infty \epsilon_{k,c}^s \alpha_c^s f_u^s dE}, \quad (4.18)$$

where channel k is said to identify particle species p , over the energy range E_1 – E_2 . Thus, as well as the fundamental monitor characteristics α and ϵ , ν depends on f_u^p and f_u^s . This means

that the identification purity of any channel depends on the external flux of the ‘impurity’ particle types. Therefore, a channel cannot be assigned a definitive purity value. Only if the detection efficiency for all impurities is negligible can the channel be interpreted consistently as indicating a single particle type.

4.3.7 Response functions

The identification efficiency and effective area, as numerical functions of species, primary energy and direction, constitute a set of response functions for the monitor. These data and Equation 4.17 may be used in two distinct ways:

1. The forward problem: evaluation of expected μ_k for given f_u^s . This is a simple computation which predicts the monitor response in a given radiation environment. This mean prediction may also be Monte Carlo sampled by assuming a Poisson distribution model. The less efficient alternative to this method would be to run a new particle transport simulation for every flux environment.
2. The inverse problem: solution of Equation 4.17 for the external flux, f_u^s , when provided with a set of measured channel counts, N_k , which are used to approximate the mean values μ_k . Methods for reconstructing the incident spectra will be explored in Chapter 6.

The identification efficiency has been calculated from Geant4 Monte Carlo simulations in a similar way to the effective area (described in Chapter 3).

4.3.8 ID table selection

The ID table is chosen to have thirty-one entries, thirty of which should correspond to one of the four detector modes. The exception is the empty frame table entry, selected when each sensor sum is less than the minimum detectable signal (chosen with reference to the sensor noise). Channel 32 (without a table entry) counts events which would otherwise go unrecorded. This means that, with suitable selection of the remaining thirty entries, the table can be constructed so that any possible readout will be allocated to one (and only one) channel. The histogram criterion for allocation of a sensor readout to a channel may be written as:

$$(L1 \leq S1 < U1) \times (L2 \leq S2 < U2) \times (L3 \leq S3 < U3) \times (L4 \leq S4 < U4), \quad (4.19)$$

where S_n is the energy deposit for sensor n and the eight values $L1-4$, $U1-4$ comprise the ID table entry for that channel.

The table must be selected to maximise the potential for describing the particle environment. A subset of channels should also be suitable for single event particle identification. It is possible for the table to be optimised according to the approximate radiation environment expected. This would then allow the calculation of a purity for each channel, using Equation 4.18. However, this should be avoided if the monitor is intended for general purpose use. A more acceptable possibility is the ‘tailoring’ of a table to a particular monitoring need,

for example, allocating more channels to C4 modes for enhanced detection of higher energy particles.

The chosen ID table (Table 4.1) was selected manually using the energy deposit plots presented in Section 4.3.5. Entries were chosen to select regions of relative particle purity in the 4-D energy deposit parameter space. These higher purity selections have generally lower detection efficiency, so mixed channels with high efficiency were also selected. While not suitable for immediate interpretation, these are important for offline ground-segment analysis. Furthermore, all channels are chosen to be mutually exclusive so that a unique allocation is possible for each event. Other tables are possible based on different selection criteria or for more specific purposes; for example, earlier work considered table selection using a heuristic search algorithm [2].

Table 4.1: Provisional ID table for the four-sensor HMRM. The entries are given in keV, rather than scaled FPGA units.

Channel	ID table entry / keV								Mode	Peak response ratio e/p	Species primary energy range / MeV			
	Sensor 1		Sensor 2		Sensor 3		Sensor 4				Electron		Proton	
											Min	Max	Min	Max
1	0	0.2	0	0.2	0	0.2	0	0.2	Empty frame					
2	0.4	1.27	0	0.2	0	0.2	0	0.2	1	388	0.05	0.11	1.50	43.7
3	1.27	4.05	0	0.2	0	0.2	0	0.2	1	461	0.05	0.16	1.41	1.68
4	4.05	12.9	0	0.2	0	0.2	0	0.2	1	1220	0.08	0.16	1.41	1.49
5	12.9	41	0	0.2	0	0.2	0	0.2	1	591	0.06	0.13	1.41	1.49
6	41	130	0	0.2	0	0.2	0	0.2	1	152	0.07	0.11	1.40	1.48
7	130	415	0	0.2	0	0.2	0	0.2	1	0.124	0.15	0.19	1.40	2.80
8	415	616	0	0.2	0	0.2	0	0.2	1	0	-	-	1.44	2.71
9	616	916	0	0.2	0	0.2	0	0.2	1	0	-	-	1.53	2.14
10	916	1360	0	0.2	0	0.2	0	0.2	1	0	-	-	1.94	2.04
11	1360	2020	0	0.2	0	0.2	0	0.2	1	0	-	-	2.62	47.4
12	2020	3000	0	0.2	0	0.2	0	0.2	1	0	-	-	23.6	47.3
13	0.4	2	0.4	200	0	0.2	0	0.2	2	3620	0.28	1.68	7.90	449
14	2	10	0.4	1.3	0	0.2	0	0.2	2	1790	0.31	1.92	151	467
15	1.3	7	40	200	0	0.2	0	0.2	2	5050	0.13	0.31	7.51	309
16	8	30	10	33	0	0.2	0	0.2	2	62.3	0.12	0.33	47.7	196
17	30	80	24	120	0	0.2	0	0.2	2	3.05	0.16	0.36	15.0	44.7
18	80	180	80	160	0	0.2	0	0.2	2	0.0136	0.38	0.50	6.62	12.8
19	80	300	160	230	0	0.2	0	0.2	2	0	-	-	5.03	7.54
20	130	300	230	360	0	0.2	0	0.2	2	0	-	-	3.90	5.54
21	200	400	360	500	0	0.2	0	0.2	2	0	-	-	3.06	4.11
22	230	450	500	800	0	0.2	0	0.2	2	0	-	-	2.93	3.74
23	230	450	800	1300	0	0.2	0	0.2	2	0	-	-	2.99	3.64
24	0.2	3000	1	10	1	100	0	0.2	3	249	0.52	2.55	115	458
25	0.2	3000	10	30	10	100	0	0.2	3	2.21	0.47	1.11	36.1	106
26	0.2	3000	30	200	10	30	0	0.2	3	1.42	0.65	1.48	33.4	64.5
27	0.2	3000	30	200	30	100	0	0.2	3	0.0807	0.79	0.84	14.2	29.6
28	0.2	3000	30	200	100	1000	0	0.2	3	0	-	-	9.47	13.2
29	0.2	3000	0.2	3000	1	10	1	40	4	48.6	3.08	5.76	108	460
30	0.2	3000	0.2	3000	10	100	8	40	4	0.704	2.69	5.84	40.8	99.8
31	0.2	3000	0.2	3000	10	100	40	1000	4	0	-	-	29.0	39.0
32	Other events													

Also shown in Table 4.1 are the approximate characteristics of each channel. For both species tested, the energy range between which the detection efficiency was greater than 50% of maximum is given. This provides an indication of the particle energy ranges selected, although *absolute* limits cannot be specified. The peak response ratio of electrons to protons is also provided for each channel. This indicates the relative magnitude of response to the two species, at their maxima of channel response. Twelve channels did not select any electrons

and were thus pure proton; there were no pure electron channels, although four have a peak selection ratio greater than 10^3 .

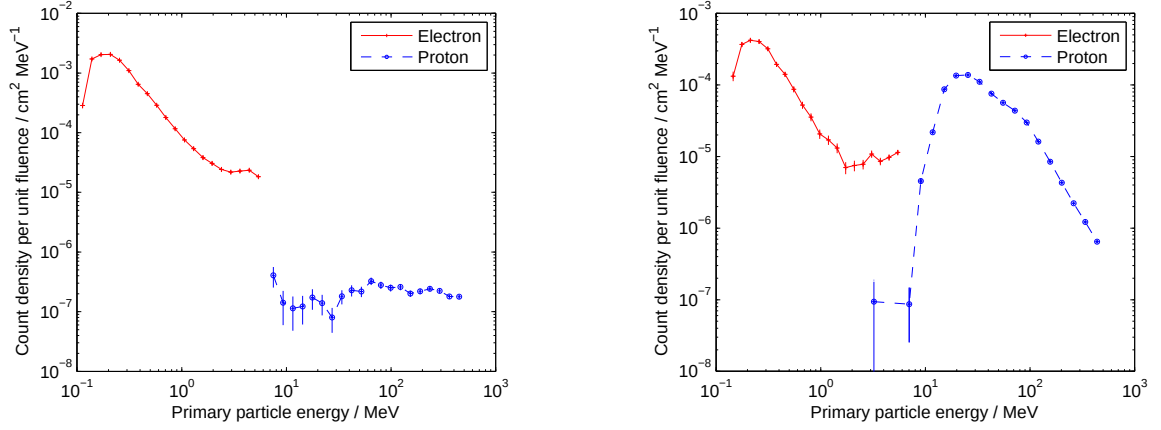


Figure 4.14: Omnidirectional response to electrons and protons for two ID table channels. Left: channel 15, showing relatively high electron detection efficiency. Right: channel 17, a mixed efficiency channel.

The channel k omnidirectional count rate response, λ_k^s , for species s , is given by:

$$\lambda_k^s = \frac{1}{4\pi} \int \epsilon_{k,c}^s \alpha^s d\vec{\Omega}. \quad (4.20)$$

This quantity gives the expected channel count per unit incident omnidirectional fluence, assuming an isotropic distribution. This is shown as a differential function of primary energy, for three channels (a high efficiency electron channel, a mixed channel and a pure proton channel), in Figs. 4.14 and 4.15.

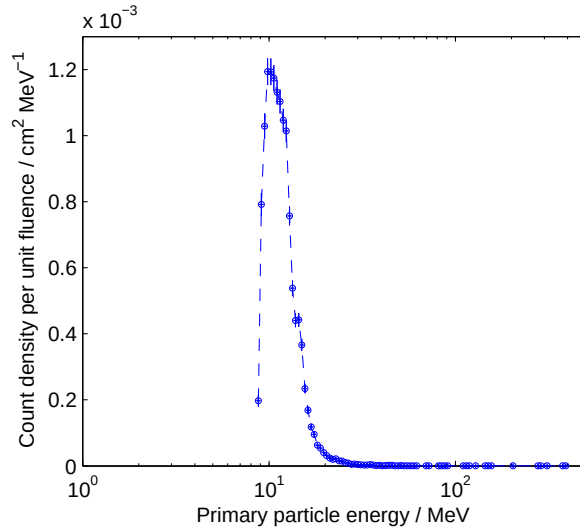


Figure 4.15: Omnidirectional response of channel 28, a pure proton channel (C3 coincidence mode). The response FWHM covers the primary proton energy range of 9.5–13.2 MeV.

Any three channels may be selected for count rate calculation; these data are included in the standard telemetry packet, while the remaining channel counts are included in the extended packet. The chosen three should be relatively pure and identify particles considered to be damaging. For example, a typical choice could be channel 19 (low energy protons for solar cell damage), channel 29 (MIP electrons and high energy protons capable of greater penetration) and channel 31 (medium energy protons for damage to imagers and lightly-shielded electronics).

Alpha particles and heavier ions

Other particle species are present in Earth orbit in addition to the electrons and protons already considered. Although the majority of HMRM simulations have studied proton and electron fluences, alpha particle (${}^4\text{He}^{2+}$ ion) identification has also been tested. As shown in Fig. 4.16, these particles are largely indistinguishable from protons — this graph was created with data from a E^{-1} power law simulation of alpha particles between energy limits of 4 MeV and 2 GeV.

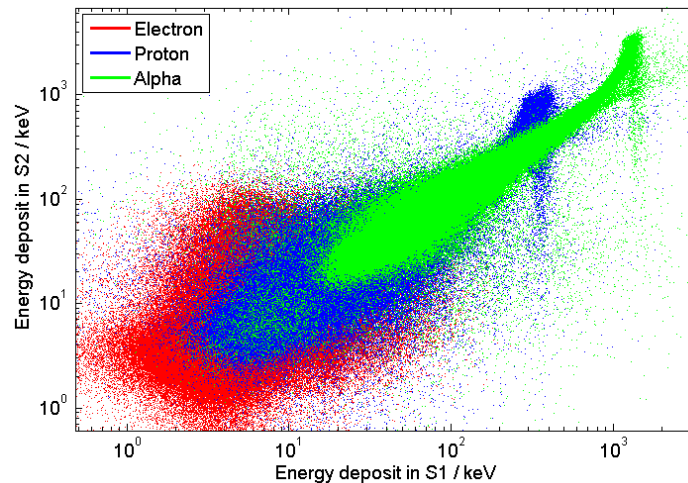


Figure 4.16: Distribution of coincident energy deposits in HMRM sensors S1 and S2 for C2 mode events due to isotropic fluences of electrons, protons and alpha particles.

This plot demonstrates that the supposed purity of the ‘proton’ channels is due to the implicit assumption that the environment consists solely of electrons and protons. Although regions of Fig. 4.16 appear to be populated exclusively by protons or alpha particles, there are again small numbers of impurity species which prevent deterministic identification based on energy deposit.

The proton/helium similarity is expected for all light hadrons with equal kinetic energy per nucleon. One possibility is to simply histogram these by energy deposit, creating a series of ‘hadron’ bins; these particles are expected to have similar damage characteristics compared to electrons, making this a useful identification. An alternative is to use prior knowledge, allowing estimation of channel purities (for example, helium ion flux is generally much smaller than proton flux).

The response to low energy heavy ions is expected to depend greatly on the saturation and charge sharing characteristics of the pixel sensors. Simulation and analysis of these ions is therefore more reliant on the device response model which will be enhanced by prototype testing.

4.4 HMRM data tables

In addition to the ID table, the two principal algorithm data tables are the pixel mask and the pixel energy weightings. These three tables are programmed via telecommand and then held in the FPGA memory for use during data analysis. Also programmed are the trim, threshold and bias settings, which are then loaded into the ASICs from the FPGA — definition of these is considered later when ASIC calibration is discussed in Chapter 7.

4.4.1 Pixel mask

The pixel mask is a series of 10200 bits (one per pixel per ASIC sensor) indicating whether to accept or reject a specific pixel’s signal. This is to be used to ignore pixels identified as anomalous or malfunctioning during the sensor calibration. A second possibility is to reduce the sensitive area of the sensor, although this has not yet been considered in detail.

When masking small numbers of pixels, the effect on the monitor’s output data may be characterised in two main ways: a reduction in effective area and an increase in misidentification probability. The former effect may be rectified, to first approximation, by a simple scaling of data products; the latter effect is addressed in this section.

The signal from single event energy deposits is in general expected to span several pixels, loss of one or more of which will result in a reconstructed energy deposit error. For a sufficiently large error, the event will be allocated to a different identification channel to that predicted by the response functions. In the case of large areas of pixel masking (e.g. to reduce effective area), the response functions should be recalculated.

To estimate the approximate misidentification probability resulting from small areas of masking, the following simplifying assumptions are made:

1. The primary particle hits only one sensor.
2. A total of n adjacent pixels are above threshold (the cluster size).
3. One pixel collects 30%–50% of the charge.
4. The remainder of the charge is shared among the remaining $(n - 1)$ pixels approximately equally.
5. The locations of masked pixels are uncorrelated.

Worst-case estimate

The worst-case situation occurs if an event is misidentified when any pixel in the event cluster is masked. In this case, with a total of m masked pixels in a sensor, the misidentification probability, P_m , is:

$$P_m = 1 - \left(1 - \frac{m}{2500}\right)^n. \quad (4.21)$$

This probability is shown as a function of m for various n in Fig. 4.17 (note that $n = 9$ is expected for most events).

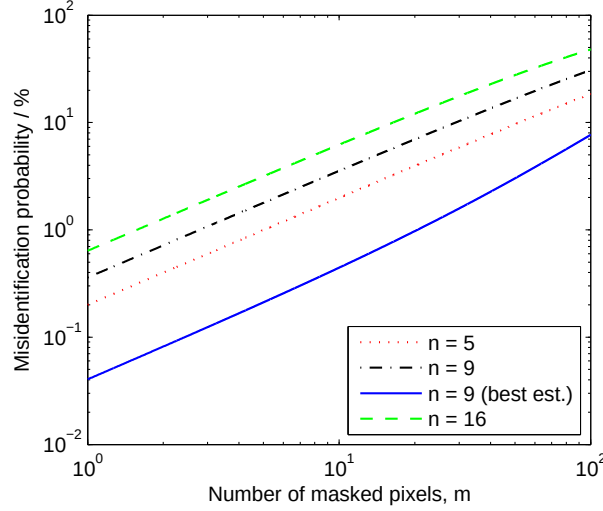


Figure 4.17: Estimated misidentification probability for events involving a single sensor, for various cluster sizes, n , as a function of the number of masked pixels. Results for $n = 5, 9, 16$ are shown for the worst-case estimate, where misidentification results from the masking of a single cluster pixel. The best estimate (misidentification caused by masking of the central cluster pixel or any other two) for $n = 9$ is also shown.

Best estimate

The worst-case estimate is, however, thought to be unnecessarily pessimistic. Instead, the best estimate is that an event is misidentified if either the central (largest charge share) cluster pixel, or at least two others, are masked. Either case ensures that more than 10% of the charge that would have been collected is lost. Assuming that the cluster size is always nine pixels, the misidentification probability becomes:

$$P_m = 1 - \frac{8m}{2500} \left(1 - \frac{m}{2500}\right)^8 - \left(1 - \frac{m}{2500}\right)^9. \quad (4.22)$$

This probability is also plotted as a function of m in Fig. 4.17.

An additional consideration is the outer pixel border of the sensor, consisting of 196 pixels. A hit to any of these is expected to result in a loss of charge (approximately equivalent to that collected by three pixels, assuming a nine pixel cluster) meaning that misidentification

of these hits is always likely. This is not accounted for in the current response model, which assumes all hits take place in a large array, far from the boundaries. Such a border hit has a probability of 7.8% — it is therefore adequate to demand a misidentification probability due to masking of smaller magnitude. This implies that fewer than ~ 90 pixels should ideally be masked (or ~ 20 under worst-case assumptions).

4.4.2 Comparator level and pixel energy weighting selection

The seven hardware comparator levels are applied to the analogue pixel signals in the ASICs, defining the 3-bit digitisation process. This is shown conceptually in Fig. 7.4, while the process is described in greater detail in Chapter 7.

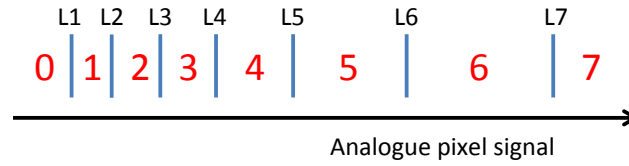


Figure 4.18: Conceptual diagram of the relation of threshold levels and digitised pixel values with respect to pixel signals of increasing magnitude. The levels (L values) are represented by vertical bars, while the pixel output values correspond to the signal ranges between the levels.

The seven levels define a series of eight charge histogram bins, the implicit minimum and maximum of which lie at zero and pixel saturation ($\sim 19,000$ e^-), respectively. The initial choice of comparator levels, prior to full ASIC calibration, is a set of geometrically spaced values between the minimum detectable signal and the saturation level. The minimum detectable signal is chosen with reference to the pixel noise (predicted to be 10 e^- ENC). For example, choosing a minimum of 50 e^- (0.18 keV) implies a false single-pixel hit probability of less than 10^{-6} , assuming Gaussian pixel noise.

The values used to weight the digitised pixel signals during the energy deposit recovery process must also be specified. These are largely determined by the choice of comparator levels. As a first approximation, the weightings are chosen to be the geometric means of the bins defined by the comparator levels. The lowest weighting, however, is zero to ensure a zero sum when all pixels are below the first level. These provisional level and weighting choices are shown in Tables 4.2 and 4.3, respectively. In use, the weightings are scaled to allow suitable resolution when using 17-bit unsigned integers for the ID table contents (this size was chosen due to FPGA memory constraints). This also determines the scaling, and hence interpretation, of the dose and dose rate data.

To test the reconstructive ability of this scheme, the ASIC response model was applied to a set of energy deposits, simulating pixellation, noise and digitisation. The resulting modelled pixel signals were summed using the weightings from Table 4.3 and each reconstructed energy deposit was plotted against the original value: this is shown in Fig. 4.19, demonstrating satisfactory predicted performance over two orders of magnitude.

Table 4.2: Provisional comparator level choices: these are chosen to form seven geometrically spaced bins between 0.18 keV (50 e^-) and the pixel saturation of 69 keV ($19,000\text{ e}^-$).

Comparator level	Threshold value / keV
L1	0.18
L2	0.43
L3	1.00
L4	2.33
L5	5.44
L6	12.70
L7	29.68

Table 4.3: Provisional pixel energy weightings: these are the geometric means of the bins defined by the adopted comparator levels. These are scaled for integer representation when implemented in the FPGA.

Pixel signal, s , relative to comparator levels	Weighting / keV
$s \leq L1$	0.00
$L1 < s \leq L2$	0.28
$L2 < s \leq L3$	0.65
$L3 < s \leq L4$	1.52
$L4 < s \leq L5$	3.56
$L5 < s \leq L6$	8.31
$L6 < s \leq L7$	19.42
$L7 < s$	45.37

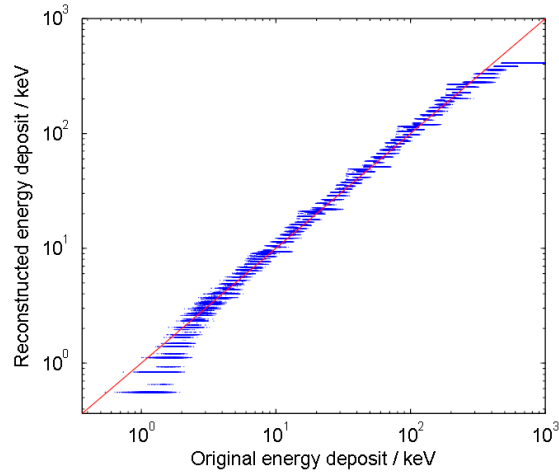


Figure 4.19: Reconstructed energy deposit as a function of the original value: this reconstruction used the comparator levels and pixel weightings provided in Tables 4.2 and 4.3. A line indicating perfect reconstruction is also shown.

4.5 Summary

A data analysis algorithm suitable for use within the HMRM has been developed. This is simple enough to be executed in real-time (event-by-event) by a small FPGA device, while delivering useful radiation data. The algorithm is also sufficiently flexible to be applied to other monitor designs, including ones with different numbers of sensors and different telemetry bandwidth requirements, with minimal changes. The algorithm requires the specification of a number of data tables, most notably the identification look-up table, which has been selected using Monte Carlo simulation data.

Particle identification with the HMRM, in common with other monitors using the dE/dx sampling method, is predicted to have mixed success. Even under ideal conditions, the identification of certain particle species/energies from energy deposit measurements is generally uncertain and must be given a probabilistic interpretation. This limitation ultimately results from the fundamental properties of the energy loss processes. The energy deposit in a thin layer, for any particle identity, is a random variable due to the stochastic nature of the interactions.

Straggling fluctuations in energy loss and range impose resolution limits on the determination of any particle property through the dE/dx method. The HMRM design has attempted to minimise these problems; for example, sensors 1 and 2 are in close proximity, minimising intervening scattering losses. Furthermore, the use of thin ASICs has allowed coincidence measurements (multiple dE/dx sampling) to occur at lower particle energies than in other monitors. However, some compromises were necessary, most notably in increased sensor separations towards the rear of the telescope.

The use of the four detection modes categorises each event in a way which is largely independent of the quality of the energy deposit measurement, providing a robust initial classification. A disadvantage of this method is that higher order modes (C3 and C4) have smaller effective areas, while the flux of higher energy particles is usually also smaller, resulting in a lower count rate for the highest energy particles. This problem, together with that posed by inter-sensor scattering losses, could be alleviated by using sensors of larger area at the rear of the telescope stack or by further reducing the inter-sensor spacing.

The probabilistic nature of the particle identification emphasises the potential importance of statistical spectrum reconstruction methods discussed in Chapter 6. Although these are expected to be executed in the ground segment, the results may be used to improve the interpretation of the monitor channel counts.

Finally, much of the monitor evaluation and testing has relied heavily on simulated results. In many instances there is no substitute for this, since knowledge of the individual incident particle characteristics is required for response function creation and algorithm testing. It is especially important, therefore, to demonstrate simulation accuracy, a topic which is addressed by the Monte Carlo validation study presented next.

Chapter 5

HMRM simulations

The design of modern radiation sensors relies on different types of device simulation which, owing to the increasing sophistication of software tools, may reduce development times significantly. This is true of the HMRM project, where simulations were used extensively in modelling various stages of the sensor response: from the incident space radiation environment, to the development of the electronic signal response and the analysis of the final digital products.

The HMRM project has made extensive use of Monte Carlo simulations with the Geant4 toolkit [84, 85], version 9.5, patch 01. Chapter 3 described how these were used to guide the design of the detector geometry, while Chapter 4 used simulated data to optimise particle identification from sensor energy deposit measurements. In this chapter, Geant4 data are used together with the HMRM physics algorithm in a simulation of the monitor response to a time-dependent flux. This allows a combined test of the monitor geometry and algorithm in an orbital environment, producing realistic output data packets.

First, however, Geant4 validation data are presented in order to demonstrate the accuracy and reliability of these simulations. This has allowed tuning of some simulation parameters and has verified the choice of physics models for the HMRM project.

Finally, an investigation of pixel charge sharing, through the use of charge diffusion simulations, is presented. A simple charge diffusion model is proposed and used to predict the characteristics of a pixel spectrum due to X-ray photon exposure.

5.1 Monte Carlo validation

Comparison of simulation results and experimental data is vital to ensure that physical processes are modelled correctly. Although the Geant4 toolkit has been extensively validated (for example, in [111], [112] and [113]), checks on individual project applications are desirable to ensure correct implementation and appropriate choice of physics models.

5.1.1 Geant4 physics

Geant4 uses several major categories of physics process, including electromagnetic, hadronic, decay and optical [114]. Multiple, often interchangeable, model implementations are available for each category and there is no single recommended or default selection. The choice of implementation may therefore be influenced by the specific particle energy ranges, computational speed criteria, or theoretical grounds relevant to the project.

The HMRM project is concerned with the measurement of charged particle ionising energy losses at energies of up to 6 MeV for electrons and 500 MeV for protons. The electromagnetic (EM) interactions are therefore expected to be especially important due to the dominance of ionisation and excitation energy loss mechanisms, with hadronics contributing significantly only for higher energy primaries.

Physics list selection

Several comprehensive packages of EM models are available as part of the Geant4 release; most commonly used is the ‘EM standard’ (G4EmStandardPhysics), a baseline collection of wide applicability which is used in the Geant4 reference physics lists. Also tested here are three packages with nominally higher accuracy: ‘EM Standard option 3’, ‘Penelope’ and ‘Livermore’, all of which are optimised for low energy electromagnetic interactions and optionally include atomic de-excitation. These investigations used the EM low energy library version 6.23.

EM treatment at low energies (below ~ 1 MeV) relies on phenomenological approaches rather than the theoretical models used at higher energies. The database libraries employed for this contain cross sections, stopping powers and other physical data.

A suitable collection of hadronic interaction models was obtained from a Geant4 study of cosmic ray spacecraft charging [115, 71]. This had been undertaken for LISA and LISA Pathfinder by the Imperial Group and is similar to that now packaged under the label ‘QGSP_BIC_HP’. This, combined with transportation and decay processes plus any one of the EM packages, forms the complete HMRM physics list.

The hadronic list features the Quark-Gluon String (QGS) theoretical model for high energy nucleons, pions and kaons (>20 GeV) and an intra-nuclear Binary Cascade (BIC) for lower energies (<10 GeV), which is proposed as most appropriate for the low energy primaries (less than 1 GeV) of greatest importance in this project [116]. An older parameterised model (LEP) is used to cover the intermediate range. For primary energies below ~ 100 MeV — or following hadron inelastic scattering — the highly excited residual nucleus is handled by a Pre-compound model and then by equilibrium models which deal with fragment or photon evaporation and fission, for example, until nuclear equilibrium is established. Neutron transport below 20 MeV is treated with the data-driven ‘NeutronHP’ models [2]. The Binary Ion Cascade is used for alpha, deuteron, triton, ^3He and generic ion (heavier ion) projectiles up to ~ 1 GeV/n.

Photonuclear and electronuclear processes are applied to gamma-rays, electrons and positrons.

The energy spectrum of muons produced in spacecraft structures is too soft to justify the inclusion of the muon-nucleus interaction, which becomes important only above a few GeV. The QGS model is implemented for high-energy gammas (>3 GeV), operating together with the CHIPS (Chiral Invariant Phase Space) nuclear fragmentation model [2].

Production cut parameter

Particles produced in Geant4 interactions are tracked to the end of their range in a medium. However, secondary particle production is controlled by the specification of a ‘production cut’ parameter. If the mean range of the secondary is less than this value, it is not produced and its energy is deposited locally instead [84]. The value assigned to this cut parameter is therefore expected to influence simulated energy deposit measurements, especially for thin layers.

Range parameter

Initial simulation step sizes of a particle entering a new volume are limited to a fraction of the particle range. This fraction, denoted F_R , presents another adjustable simulation parameter. Without this limit, a large first step will cause an excessive energy loss, leading to subsequent smaller steps. This affects the angular scattering distribution and backscattering yields because the particle may not be able to return to the surface before being absorbed [117]. A default value of $F_R = 0.02$ is provided, which may be changed by the user.

5.1.2 Experimental data

A literature search for experimental particle data published over the past 40 years has been conducted. The selected experiments mostly provide measurements of beam-target interactions at mildly relativistic and non-relativistic energies. These include electron, proton and ion primaries incident on solid state layers, with measurements of the energy deposits, particle energy loss and angular scattering. Special emphasis is placed on experiments dealing with thin layers of material (<100 μm). The HMRM includes several such layers, for example, the first ASIC die (50 μm), the sensor epitaxial layer (12 μm) and the aperture window (8 μm). The accurate simulation of these is important in assessing the monitor’s capabilities, such as its low energy sensitivity limits.

Where possible, multiple measurements of the same phenomena have been sought in order to provide independent verification and allow comparison between different experimental datasets. However, due to the large parameter space, this is not always possible. Furthermore, detailed reporting of the experimental setup and uncertainties (especially systematics) is rare. This meant that some datasets were rejected because insufficient details prevented their accurate reproduction in simulation.

5.1.3 Simulation method

All of the selected experiments used a beam-target arrangement where a parallel, monoenergetic beam of single particle species is incident on a planar, homogeneous material layer. This allowed a highly automated system of simulation with each experimental run being described by just five parameters: (1) beam species, (2) beam energy, (3) incidence angle of the beam on the target layer, (4) layer material, (5) layer thickness.

The data saved from each event included the energy deposit in the layer, the primary particle energy on exit, and the species, energy and direction of all secondaries leaving the layer. The simulated data were then analysed to produce quantities for direct comparison with the experimental values. Sufficient data were usually obtained to give statistical uncertainties which were small compared to those of the experimental data (and generally of order 1%).

This simulation, analysis and comparison process was repeated using various combinations of EM package, F_R and production cut values. A selection of the results are presented below.

5.1.4 Electron and positron backscattering

Simulations of electron backscattering reproduced the large-scale features of the experimental distributions, with peak locations within 10° of the measured values. An example is shown in Fig. 5.1 using the Penelope package, although the four baseline EM packages all gave approximately identical results. However, the absolute agreement with experiment is relatively poor, with most simulated values approximately 10% too small.

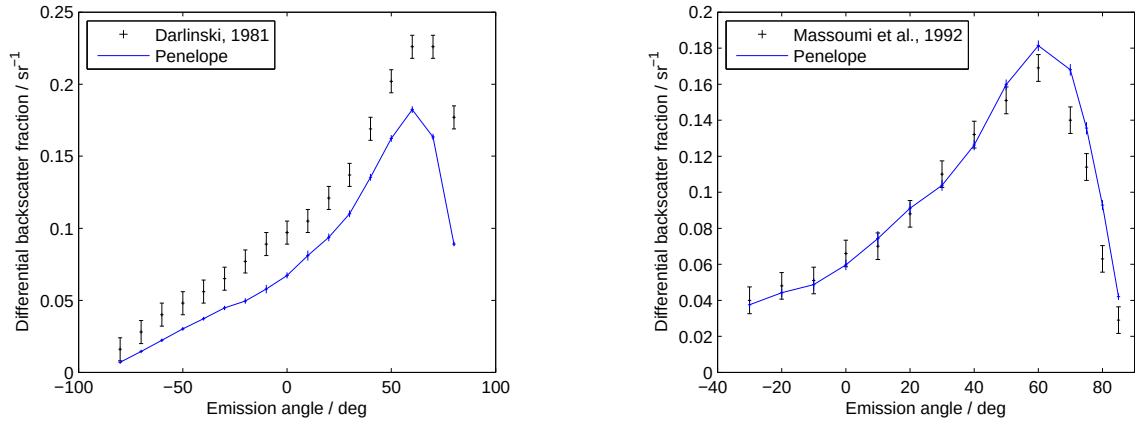


Figure 5.1: Directional backscatter from aluminium. Left: 30 keV electrons [118]. Right: 35 keV positrons [119]. The incidence angle was -60° in both cases (with 0° defined as the surface normal). These simulations used Penelope EM physics, with $F_R = 0.01$ and 10 μm cut.

The simulated results for backscattering of positrons are a better match than for electrons with similar energy and target material. Furthermore, the simulated results for positrons and electrons are essentially indistinguishable, whereas the experimental results differ in

peak location. This suggests a possible oversimplification in the simulation models, rather than a systematic experimental error; both electron and positron multiple scattering use the ‘G4UrbanMscModel95’ model [117].

Figure 5.2 shows the effect of varying F_R , using the Penelope package: larger F_R decreases the peak amplitude (a similar result was obtained for positrons). This reduction is expected because a larger F_R allows the particle to enter further into a volume on its first step and is thus less likely to re-emerge. This trend of decreasing backscattering with increasing F_R is also found to increase with the atomic number of the target, presumably due to the greater stopping power. Figure 5.2 indicates that F_R values in the approximate range 0.01–0.02 are suitable and is consistent with the default value of 0.02. Variation of production cut over the range 1 μm to 1000 μm was found to have no significant effect on any simulated backscattering distribution.

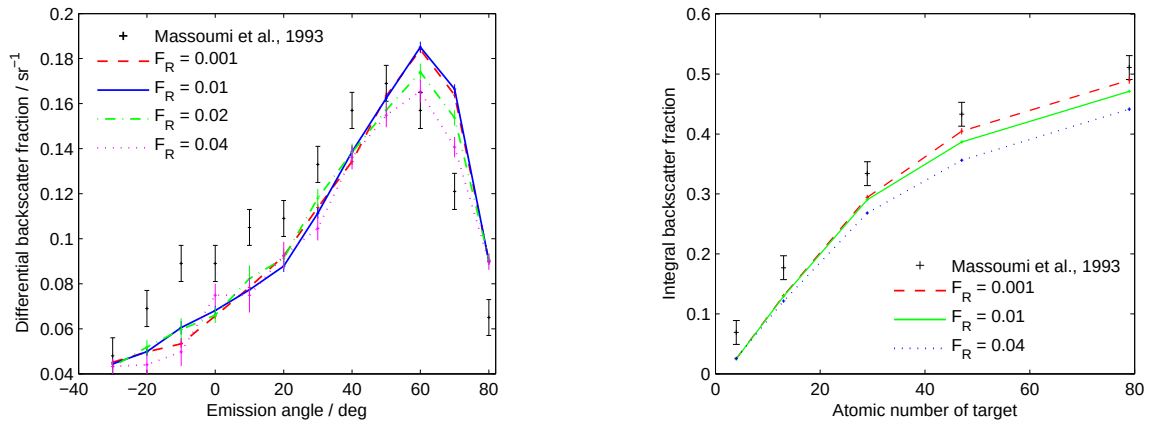


Figure 5.2: Backscatter data for 35 keV electrons [120] compared with simulation results using different F_R values (Penelope EM physics, 10 μm cut). Left: directional result from aluminium. Right: Integral result from various elements (beryllium, aluminium, copper, silver and gold).

5.1.5 Energy losses in thin layers

Generally good agreement with experiment is evident in Fig. 5.3, which considers energy losses in extremely thin layers (of order 0.1 μm). Small differences between the four EM packages were found, with Penelope performing best. Production cut dependence is insignificant until reduced to 1 μm , where there is a reduction in mean energy loss for particles emerging close to the surface normal. Additionally, there is a significant increase in mean energy losses for lower energy particles; overall, reduction of the cut length appears to improve the match with experimental results.

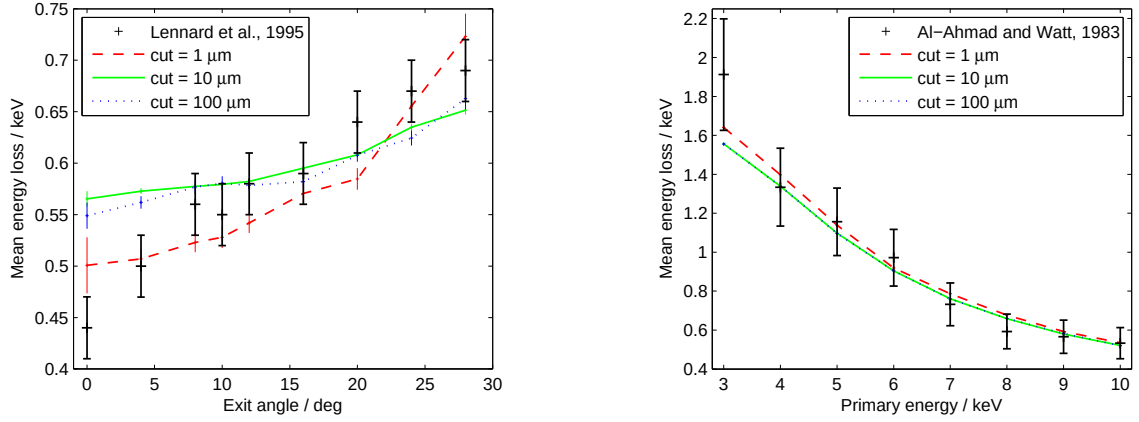


Figure 5.3: Mean energy losses of electrons passing through thin aluminium layers, compared with simulations using different production cuts (Penelope EM physics, $F_R = 0.01$). Left: directional result for 20.4 keV primaries incident on a 0.226 μm layer [121]. Right: as a function of primary particle energy for a 0.1035 μm layer [122]. Both experiments used normal incidence beams.

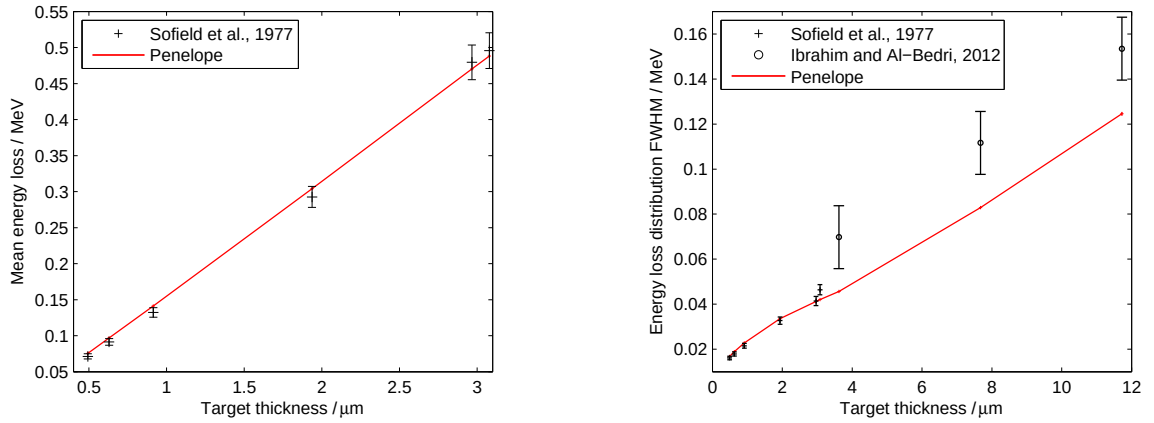


Figure 5.4: Energy losses of 5.49 MeV alpha particles passing through thin aluminium layers at normal incidence. Left: mean energy loss [123]. Right: energy loss straggling [123] [124]. The simulations used Penelope EM physics, with $F_R = 0.01$ and cut = 10 μm .

The mean energy loss results for alpha particles in thin foils (Fig. 5.4) show excellent agreement with experiment. The energy loss straggling data (distribution FWHM) shown here are a rare example of directly comparable experimental measurements. The simulation results give an excellent match with the data reported by Sofield *et al.* [123] but are significantly lower than those of Ibrahim and Al-Bedri [124]. This is indicative of a systematic error in the latter experimental investigation, possibly a failure to fully remove the broadening effects of the measurement apparatus from the energy loss distribution. Due to the brevity of the article, however, this suggestion cannot be verified.

5.1.6 Multiple scattering

Figure 5.5 shows the results of scattering measurements for proton beams passing through beryllium and aluminium. In these experiments a narrow, normal incidence, monoenergetic beam of 158.6 MeV protons was allowed to pass through material layers of different thicknesses. The angular distribution of the emerging particles was measured and a Gaussian distribution fitted to these data, from which the width was obtained. The simulation results accurately match the experimental data, both in shape and in absolute magnitude, with typical differences of just 5%.

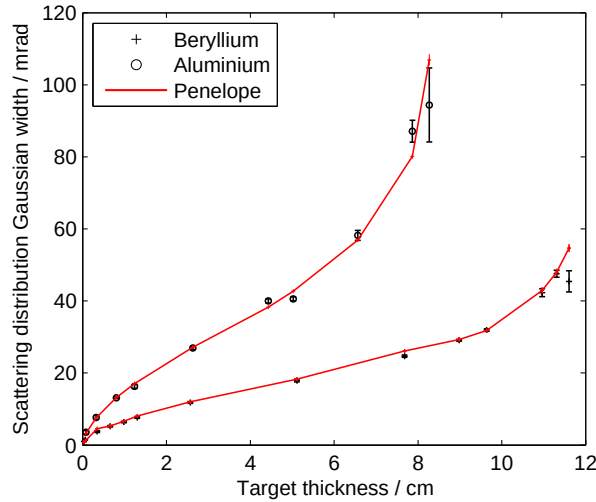


Figure 5.5: Angular scattering distribution widths of a 158.6 MeV proton beam after passing through various thicknesses of beryllium and aluminium [125]. The width measurement was obtained via a Gaussian distribution fit. The simulations used Penelope EM physics, with $F_R = 0.01$ and cut = 10 μm .

5.1.7 Depth-dose curves

Simulated comparisons have been made with experimental data taken from an extensive investigation of electron energy deposits at various depths in extended media [126]. These measurements were obtained using a calorimetric method, for particle energies up to 1.0 MeV, with an electron beam incident on pure element layers.

The simulations showed generally excellent agreement with the experimental data. However, as shown in Fig. 5.6, the EM Standard results are consistently less accurate by a small but significant amount, whereas the other three physics packages gave similar values.

Cut and F_R dependence are shown in Fig. 5.7: the former is relatively minor for these datasets, although a slight improvement is obtained with a larger value (1000 μm). Large values of the F_R parameter, however, cause a clear oscillation of the simulated energy deposition at depths up to about 400 μm . The oscillation amplitude decreases as F_R is decreased, indicating suitable F_R values of 0.01 or less: a significant result given that 0.02 is the default parameter value. Similar effects were found in other simulations using $F_R = 0.04$, at depths up to 200 μm . These oscillations may be due to a feedback effect between the particle energy loss and the resulting step size.

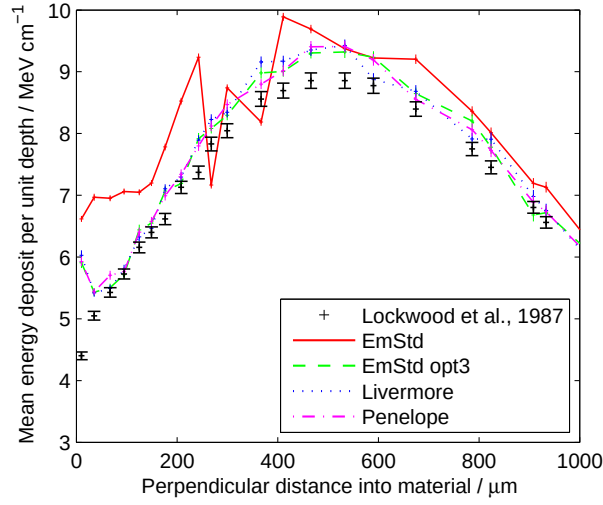


Figure 5.6: Depth-dose curve in aluminium for a 1.03 MeV electron beam at normal incidence [126]. The data are compared with simulation results using various EM physics packages ($F_R = 0.01$ and cut = 10 μm throughout).

Higher energy proton and ion modelling was tested by simulating parallel, monoenergetic beams incident on high density polyethylene. Measurement of the dose delivered to a series of sensitive volumes at different depths in the material allowed the construction of depth-dose curves. These were compared with experimental data gathered using an ion beam at the NASA Space Radiation Laboratory, Brookhaven [127]. These data are available in relative units only, requiring the use of a best-fit scaling to match the simulated data to the experimental measurements; note, however, that the depth values were not scaled. The Bragg peak location and relative curve shape are therefore relevant for these comparisons. As illustrated in Fig. 5.8, the agreement is generally excellent (the experimental uncertainties are unknown).

Figure 5.9 shows comparison of the experimental data with simulations with and without the hadronic list enabled (EM package always enabled). This confirms that hadronic interactions become important at primary energies of order 100 MeV and higher.

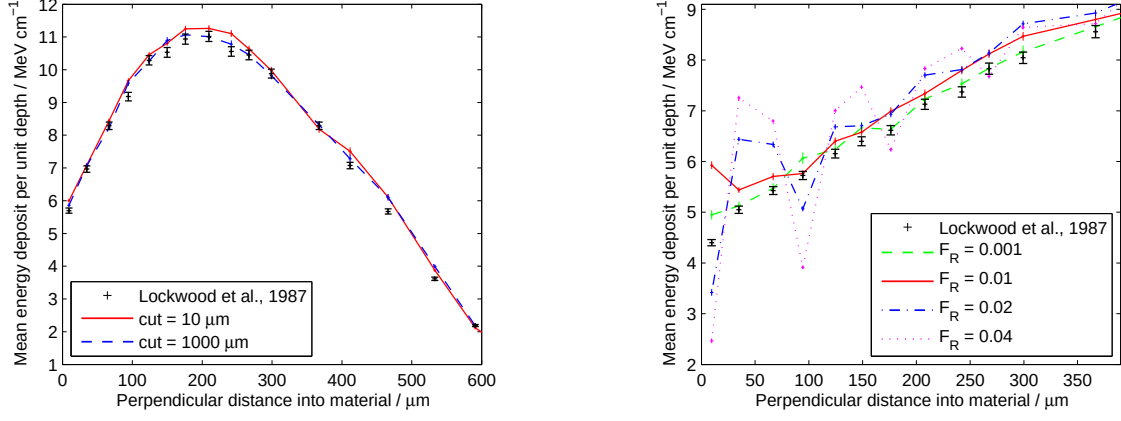


Figure 5.7: Depth-dose curves in aluminium for electron beams at normal incidence [126]. Left: 0.52 MeV, compared with simulations using two production cut values ($F_R = 0.01$). Right: 1.03 MeV, compared with simulations using different F_R values (cut = 10 μm). All simulations used Penelope EM physics.

The result for the ^{56}Fe ions with hadronics enabled shows good agreement, especially in the region beyond the Bragg peak, demonstrating accurate modelling of knock-on and secondary particles. The region at smaller depths (up to ~ 10 cm) shows a maximum discrepancy of 20%; this may be a result of the simplistic simulation of the experimental arrangement, including the simplified composition of the beam itself.

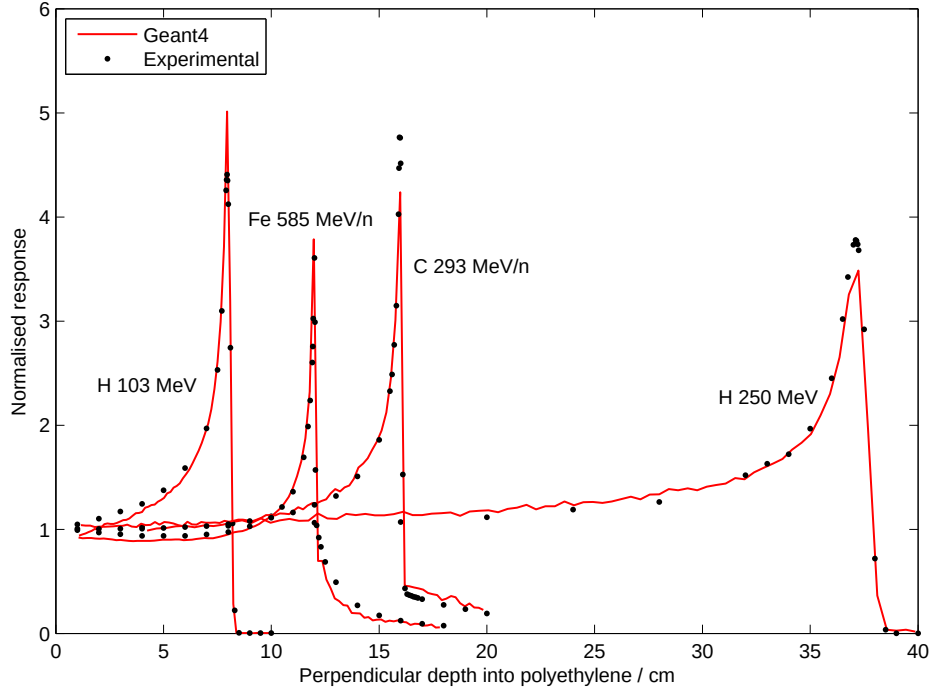


Figure 5.8: Depth-dose curves in high density polyethylene due to ion beams at normal incidence: data for protons at 103 MeV and 250 MeV, ^{56}Fe at 585 MeV/n, and ^{12}C at 293 MeV/n are shown [127]. The simulations used the Penelope EM package ($F_R = 0.01$, cut = 10 μm) and the HMRM hadronic list; a best fit constant response scaling was applied to the simulated data.

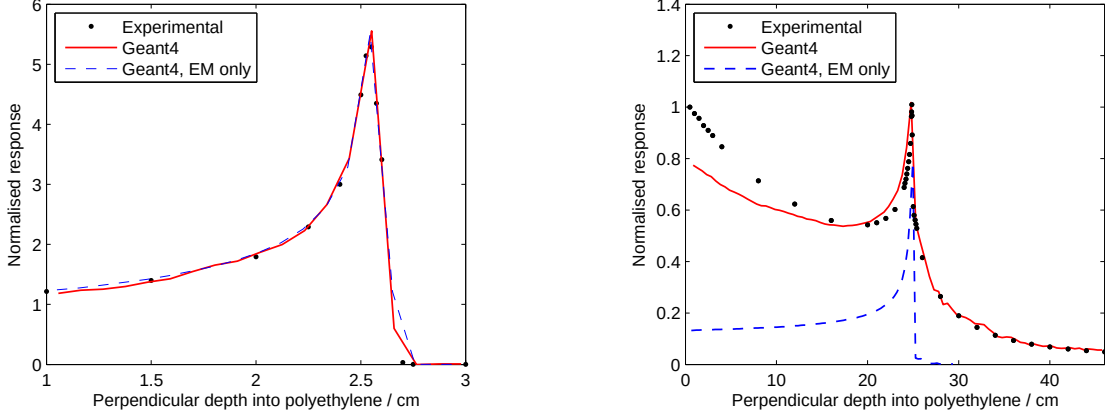


Figure 5.9: Depth-dose curves in high density polyethylene due to particle beams at normal incidence [127]. Left: 55 MeV protons. Right: 963 MeV/n ⁵⁶Fe ions. The experimental data are compared with simulation results using the Penelope EM package ($F_R = 0.01$, cut = 10 μm), both with and without the HMRM hadronic list enabled. A best fit constant response scaling was applied to the simulated data.

5.1.8 Energy deposits in silicon sensors

The modelling of energy deposits in the thin silicon HMRM sensor layers is especially important. Experimental data giving the most probable energy deposit and energy deposit fluctuation (distribution FWHM) due to energetic electron and proton primaries has been collected. These datasets cover sensor thicknesses between $\sim 1 \mu\text{m}$ and $\sim 1000 \mu\text{m}$ at particle energies in the MeV to GeV range. The simulations again used the Penelope package (with $F_R = 0.01$ and 10 μm cut), although this defaults to the EM standard at energies above 1 GeV.

Electron data from several experiments are shown in Fig. 5.10, while similar proton data are shown in Figs. 5.11 and 5.12. In addition, Figs. 5.13 and 5.14 show several proton measurements from single experimental datasets. The data are generally reproduced accurately by the simulations although small, consistent discrepancies from a single dataset may suggest the possible presence of systematic experimental errors. For example, both the most probable and the FWHM data shown in Fig. 5.13 are systematically higher than the simulated values, although the trends with primary energy show a good match. This discrepancy may be caused by additional variance in the experimental data (e.g. equipment related) which widens the energy deposit distribution. Since the distribution is asymmetric, this also has the effect of displacing the peak value towards higher energy.

The proton most probable deposits in Fig. 5.11 show a better match, although in this case the simulated values for the thinnest sensors appear too large. Since these data are from a single experiment [128], systematics cannot be excluded and additional data for very thin sensors ($\sim 10 \mu\text{m}$) should be sought to investigate this further. Lower proton energies ($\sim 10 \text{ MeV}$) are considered in Fig. 5.14, once again showing excellent agreement (196 μm silicon sensor). The

results for proton FWHM values (Fig. 5.12) also show small but consistent differences. Here, simulated values for sensors thicker than 100 μm are consistently low, whereas the values for sensors below 100 μm are consistently high. Further investigations of these effects are planned, with emphasis on energies of less than 1 GeV in thin sensors.

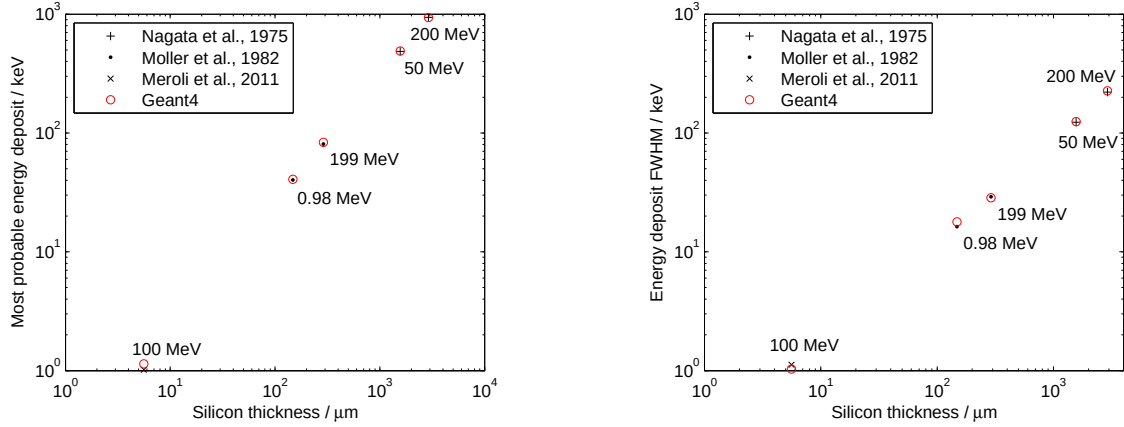


Figure 5.10: Energy deposit data for electrons incident on silicon sensors, showing the most probable value (left) and the distribution FWHM (right). The data are organised by sensor thickness, with the electron energy indicated for each datum [128, 129, 130].

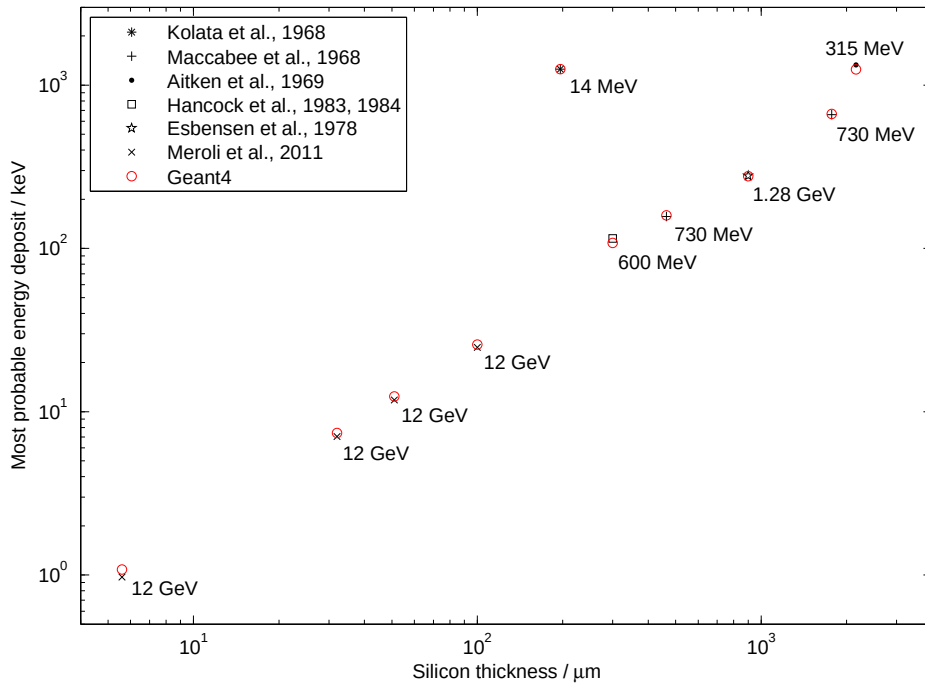


Figure 5.11: Most probable energy deposit data for protons incident on silicon sensors. The data are organised by sensor thickness, with the proton energy indicated for each datum [130, 128, 131, 132, 133, 134, 135].

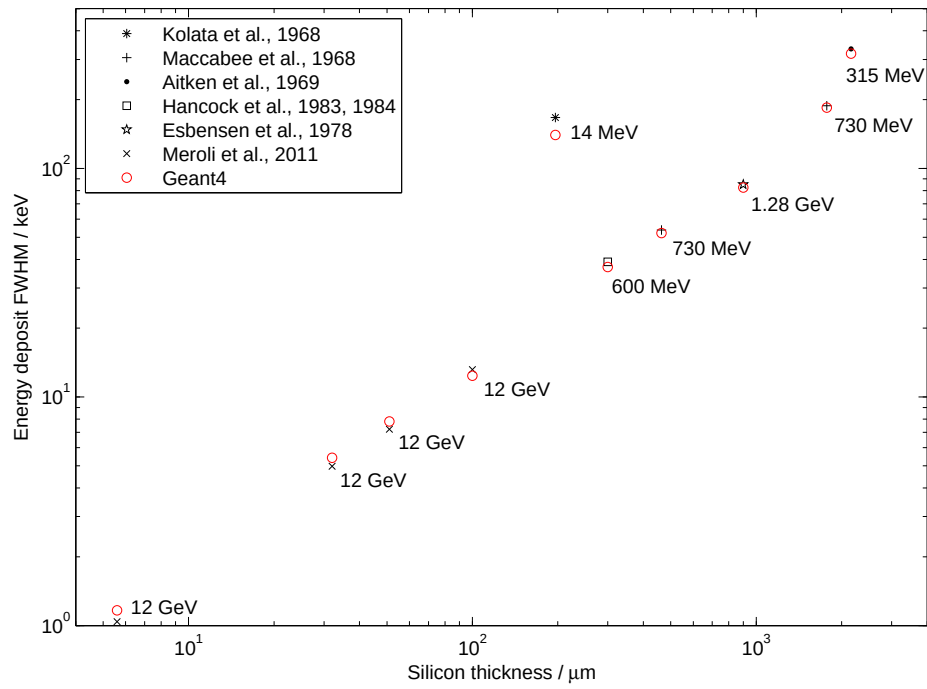


Figure 5.12: The FWHM of the energy deposit distributions for protons incident on silicon sensors. The data are organised by sensor thickness, with the proton energy indicated for each datum [130, 128, 131, 132, 133, 134, 135].

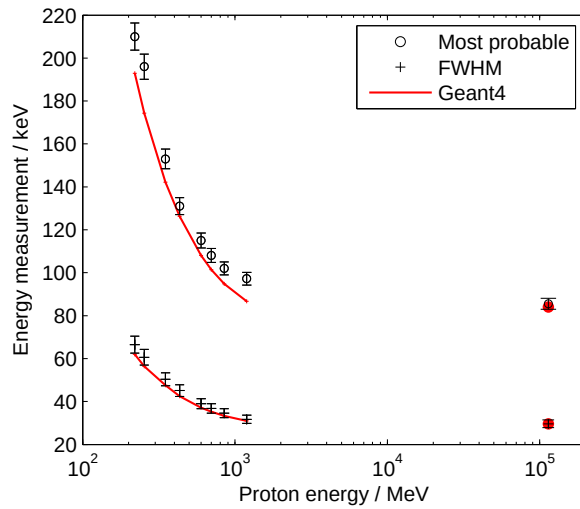


Figure 5.13: Energy deposit data for protons of various energies incident on a 300 μm thick silicon sensor, showing both the most probable values and the distribution FWHMs [131].

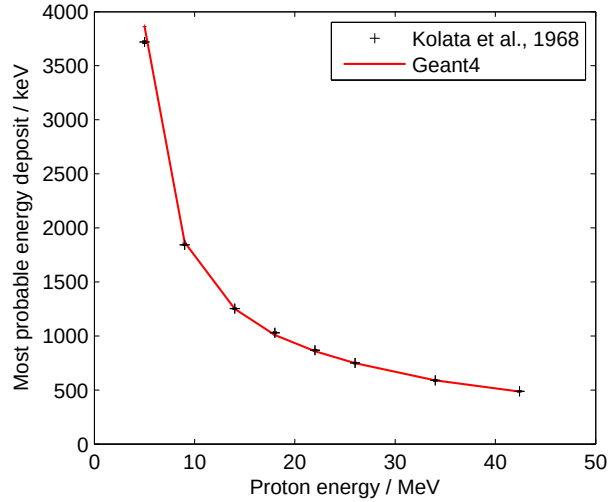


Figure 5.14: Most probable energy deposit values for protons of various energies incident on a 196 μm thick silicon sensor [132].

5.1.9 Validation summary

A simple computational speed test was performed by running the same simulation (300 keV electrons onto an aluminium target) with each of the four EM packages enabled. The EM Standard package was found to be nearly 25% faster than Penelope and EM Standard option 3, as expected, while Livermore was $\sim 5\%$ slower than these two. For the HMRM project, with its small geometry and relatively low energy particles, these differences in speed are not considered significant.

The Penelope low energy EM package has been identified as more accurate than the EM Standard package in several comparisons with experimental data. Furthermore, a production cut of 10 μm and a range factor of $F_R = 0.01$ have been selected as most appropriate for the simulation of energy deposits in thin layers.

5.2 HMRM orbit simulations

Although Geant4 simulations produce accurate predictions of energy deposits in the monitor sensors, these are treated as independent events; typically, Monte Carlo simulations do not consider the distribution in time. This means that the basic modelled response is independent of the flux magnitude, whereas effects such as pile-up are expected to affect the real monitor.

Operation in orbit is characterised by transient changes in flux and energy spectra as the spacecraft moves through the Earth's magnetic field — more detailed simulations are therefore required which involve time-dependent effects. These will help determine whether the planned exposure and integration times are suitable, given that a modelled flux may be assumed for

a particular orbit. This also provides a simulated test of the electronic shuttering system algorithm and generates realistic output data to be used for analysis testing.

Electron and proton time-dependent energy spectra in the reference orbits have been obtained from the AE-8 and AP-8 models [7, 6]. These provide estimates of the omnidirectional flux due to magnetospherically trapped particles; other species and flux components (such as SEP events) will be added in future work. The simulation procedure for each HMRM readout frame is:

1. Fold the monitor response functions with the instantaneous incident spectra.
2. Monte Carlo sample the resulting distributions to obtain event energy deposits (combining if necessary to model a pile-up event).
3. Apply the monitor response model to obtain the digital pixel values (including random noise).
4. Execute the HMRM physics algorithm.

The particle spectra are currently applied as isotropic fluxes, although the monitor response functions are unidirectional and therefore allow future extension to anisotropic flux models.

Example results from the simulation of 100 minutes in reference orbit A (a 700 km altitude SSO at 98° inclination) are presented below. The incident integral fluxes for electrons and protons in this orbit are shown in Fig. 5.15.

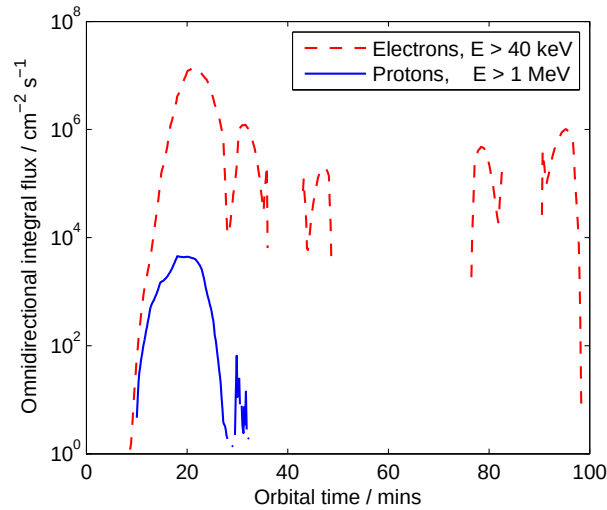


Figure 5.15: The omnidirectional integral fluxes for electron energies above 40 keV and proton energies above 1 MeV in reference orbit A, according to the AP-8/AE-8 models.

Reference orbit A: count rate and shuttering

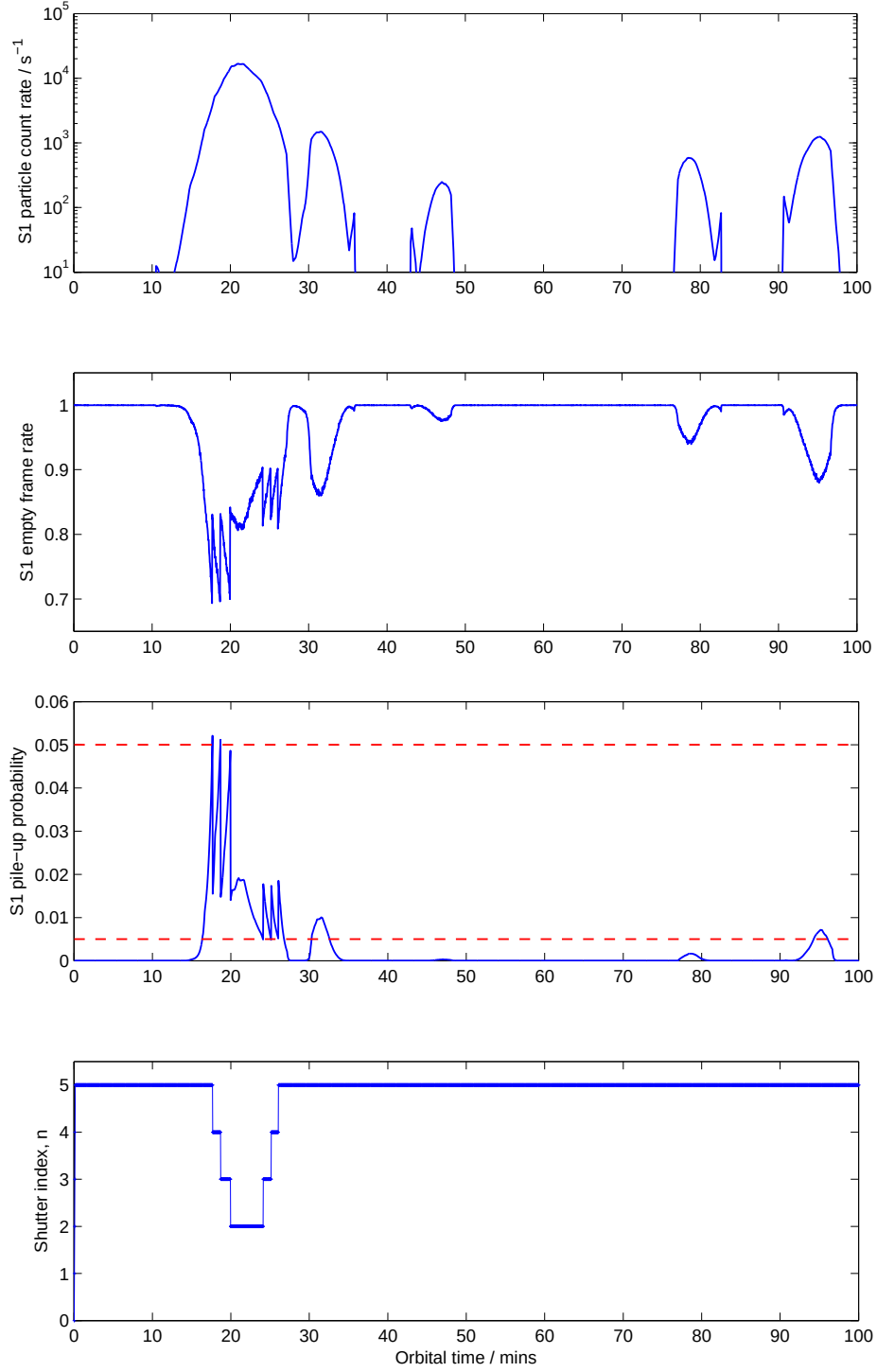


Figure 5.16: Simulation results for reference orbit A, a 700 km SSO. The graphs show measured count rate, empty frame rate (readout frames where no particle hit is detected), pile-up probability and the electronic shutter index, which controls the charge integration time per exposure frame. This is shown for approximately one orbit.

Reference orbit A: dose rates

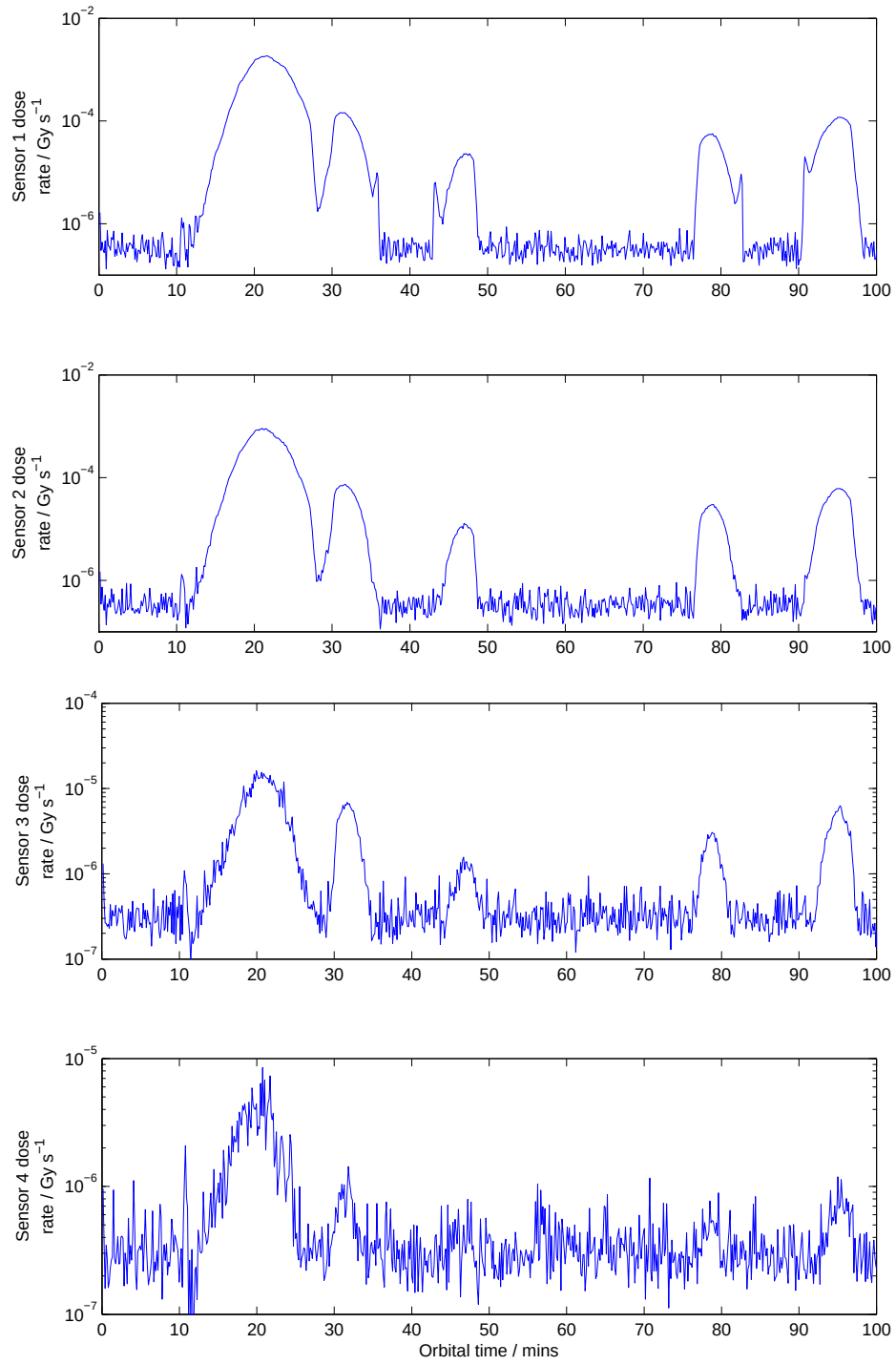


Figure 5.17: Simulation results for reference orbit A, a 700 km SSO. The graphs show the measured ionising dose rate in each of the four HMRM sensors as a function of time: approximately one orbit is shown.

Figure 5.16 shows five large peaks in the measured count rate: these correspond to passages through the SAA (after ~ 20 minutes), the southern polar horn (at 32 mins and 47 mins) and the northern polar horn (78 mins, 95 mins). This figure also shows that the shuttering algorithm is effective in maintaining pile-up between the chosen limits of 0.5% and 5%. This is important because particle misidentification (defined as allocations not quantified by the response functions) will be at least as large as the pile-up rate.

Shuttering (reduction of sensor charge integration period) is also required to ensure that the simple total particle count rate measurement remains accurate in high flux environments. This is illustrated in Fig. 5.18, which compares the true simulated rate with that measured by the monitor algorithm, showing how the error is reduced when the shutter changes to a shorter exposure time in an increasing flux.

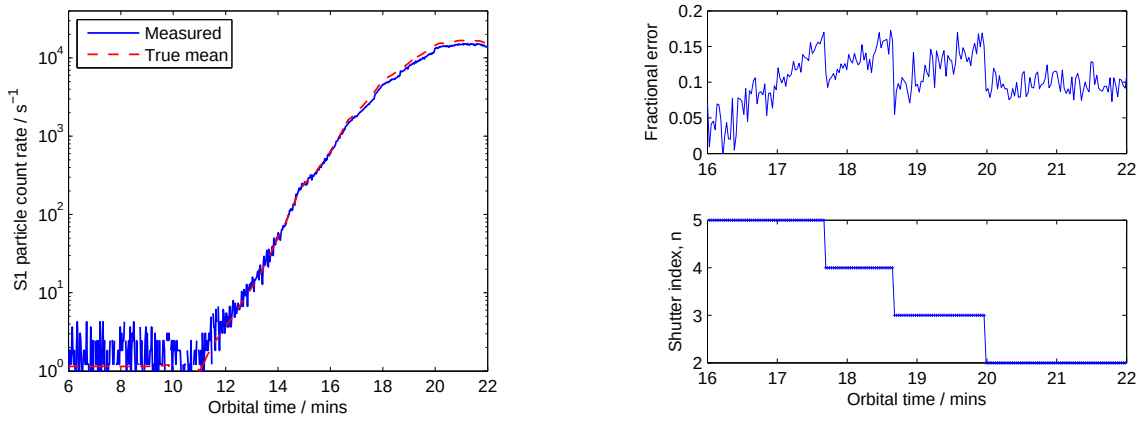


Figure 5.18: Left: comparison of the count rate measured by the HMRM algorithm with the ‘true’ simulated rate. Right: fractional difference between these rate values in an increasing flux (up to fluxes in excess of $10^7 \text{ cm}^{-2} \text{ s}^{-1}$), showing the effect of shuttering.

Integration of the four sensor dose rates in Fig. 5.17 predicts total ionising doses (TIDs) for this 100 minute period of 658, 313, 8.6 and 3.6 mGy for S1–4, respectively. These doses are absorbed in the surface pixel sensor regions of the ASICs (28 μg of silicon); however, other regions of the ASICs will also experience a similar exposure. Note that this is a worst case orbit, having been selected to pass through the most intense portion of the SAA (whereas the SAA is only encountered in $\sim 25\%$ of these orbits).

Reference orbit A: event identification count rates

Figure 5.19 shows the count rate measured for four identification channels. The ID table used for this simulation is provided in Appendix E and differs from the optimised table presented in Chapter 4. Channels 9 and 22 detect MIPs (energy deposits of $\sim 1.5\text{--}10 \text{ keV}$), with little species discrimination. Channel 18 shows high detection efficiency for electrons and also peaks in the polar horns and SAA.

Channel 15, however, is associated with low energy hadrons (in this case, only protons are

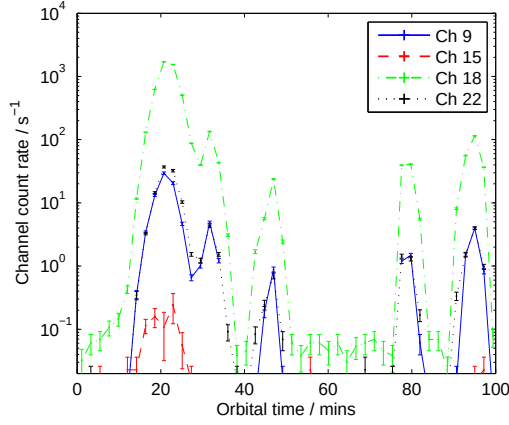


Figure 5.19: Simulated event identification count rates for four ID table channels in reference orbit A.

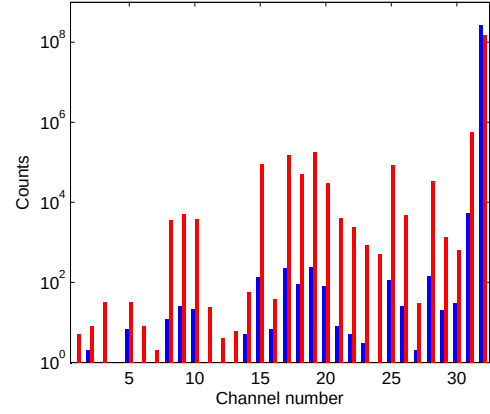


Figure 5.20: Simulated event identification counts in all channels for two consecutive 14 minute periods in orbit A.

simulated) and shows significant counts only in the SAA — proton flux is negligible in the polar horns. This channel selects events with large deposits (~ 100 keV) on sensor 1 only. Protons of low energies are especially damaging to solar cells, making this a useful channel despite its low count rate (high species purity has been obtained at the expense of detection efficiency).

Figure 5.20 provides the total channel counts for two consecutive 14 minute acquisition periods. The changing particle energy spectra lead to changes in channel counts which may be used to trigger alerts to the host spacecraft. These data are also used as the input to the offline (executed in ground segment) spectrum reconstruction methods which are discussed in Chapter 6.

5.3 Simulated charge sharing investigation

A conceptual model of an active pixel sensor is shown in cross-section in Fig. 5.21. Energy deposition results in the creation of electron-hole pairs which move through the epitaxial layer both in response to electric field and due to thermal diffusion. Charge collection, however, takes place only in the depletion region under the n-well collection diodes. Radiation interactions in the region between diodes will therefore generally result in partial collection by more than one diode and possible loss of charge. Due to this charge sharing between pixels, the single-pixel spectrum will not match the radiation energy deposit spectrum.

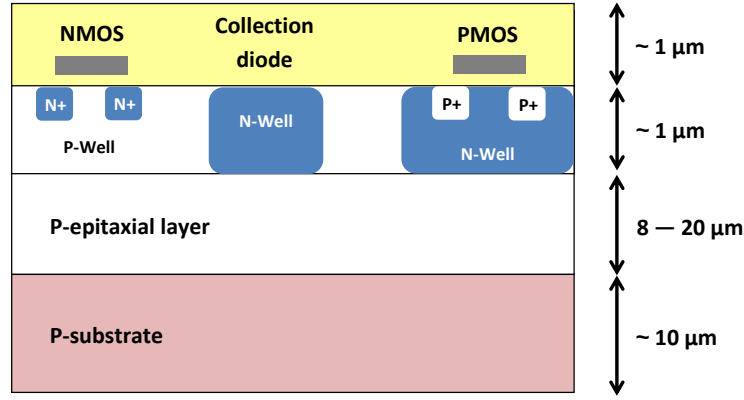


Figure 5.21: Simplified conceptual diagram of a cross-section through a CMOS active pixel sensor (not to scale), with typical layer thicknesses indicated. The epitaxial layer and substrate are shown, together with a charge collection diode and NMOS and PMOS transistors [92].

Assuming negligible charge loss and a large measurement resolution per pixel, the total event deposit may be recovered by summation of pixel ‘clusters’ (groups of adjacent pixels with non-zero signal). The cluster spectrum should thus reproduce the energy spectrum of the absorbed radiation. However, this cluster summing may be imperfect in monitors with limited resolution per pixel, such as the HMRM. A prediction of the effect of the charge sharing between pixels is therefore a desirable part of the device response model. Simulations presented in this section are preliminary only and are not yet intended to provide a quantitative description of the HMRM sensors.

5.3.1 Sensor layout

As illustrated in Fig. 5.21, a lightly p-doped epitaxial layer, which may be up to 20 μm in thickness, is situated between the substrate and surface process layers. Particle ionising energy losses in the epitaxial layer generate electron-hole pairs, constituting the detected signal, although interactions in the upper region of the substrate may also contribute. The surface layers consist of metal circuit tracks, silicon dioxide insulation and passivation layers.

The substrate is more heavily doped than the epitaxial layer, forming a potential barrier that confines signal electrons to the latter. NMOS transistors are located in p^+ wells formed in the first few micrometres of the epitaxial layer, also creating a confinement potential. Most of the epitaxial layer is not depleted and is free of electric field [136], thus the charge carriers move predominantly by thermal diffusion. Collection occurs in the small depletion region immediately surrounding the diode n-wells.

PMOS transistors cannot be implemented in the normal way (shown in Fig. 5.21) since the necessary n-well will also collect charge, preventing it from appearing as signal. One possible remedy for this is to create a deep p-well implant below the PMOS transistor. This shields the n-well from the epitaxial layer, reinstating the potential barrier. This results in efficient charge collection at the diode while allowing true CMOS electronics within the pixel. One such example of this strategy is the INMAPS process [137].

The HMRM APS uses a 12 μm thick epitaxial layer, in a 0.18 μm process with up to six metal circuit layers. Versions have been created with low and high ($>1 \text{ k}\Omega \text{ cm}$) resistivity substrates for further exploration of performance.

5.3.2 Diffusion model

It is proposed that the charge spreading process be modelled as the diffusion of a single fluid through the epitaxial layer. Pixel diodes should act as a sink for this fluid, where the output signal is assumed to be proportional to the total charge sunk during the integration time.

By simulating the diffusion and collection for different radiation interaction locations, total event energy deposits (which may be obtained through Monte Carlo simulation) can be ‘pixellated’ into a set of pixel charge signals. For added realism, each pixel signal should have a randomly sampled noise charge added — this also includes pixels that do not collect the particle-induced charge (noise-only pixels). This process allows the conversion of an energy deposit spectrum into a predicted pixel spectrum.

A simple diffusion equation [138] may be written as:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \vec{u} = \beta\rho + \gamma, \quad (5.1)$$

where ρ is the charge density, t is time and $\vec{u}$ is the current density. The source/sink term $\beta\rho + \gamma$ contains two constants, either or both of which may be zero. Assuming that the medium is isotropic, homogeneous and linear, the charge and current densities may be related by a constant diffusion coefficient, α , according to:

$$\vec{u} = -\alpha \nabla \rho. \quad (5.2)$$

The simplest model for the sensor is to assume a single medium for the epitaxial layer, where the source/sink term is zero, through which charge may diffuse. Diodes are represented by a periodic array of subregions where there is a non-zero sink and charge is removed from the geometry. Initially, a cloud of charge (the particle interaction point) exists at a chosen location, with the charge density made zero everywhere else. Spatial boundary conditions define the charge density to be zero at large distances from the interaction point at all times. Furthermore, charge is assumed to be conserved and $\rho \geq 0$ always.

5.3.3 Computational solution

Equation 5.1 may be integrated to obtain the charge density throughout the array as a function of time and hence deduce the fraction collected in each diode. However, analytical solution of this equation is a challenging task even for such a simple geometry. Instead, a computational approach was followed and the equation was solved using the Partial Differential Equation Toolbox of the MATLAB numerical computing environment [139]. This software allows the solution of time-dependent partial differential equations in two spatial dimensions using the finite element method. The user must describe the geometry of the problem using simple

shapes and assign numerical coefficients (diffusion, sink, etc.) to these different regions. The software then assigns a triangular mesh over the geometry, which may be later optimised if necessary. The solution, ρ , is approximated as a piecewise linear function over the nodes of this mesh.

Solution with this software requires that the problem be represented in two dimensions (2-D). This was judged to be an acceptable simplification because the pixel features, such as the diode array, are largely two dimensional as they are situated on the thin process layers near the sensor surface. As described in Section 5.3.1, charge is confined by a potential well to the continuous epitaxial layer which extends across the whole sensor. Lateral diffusion, over anticipated distances of up to 30 μm , is thus assumed more important than vertical diffusion (over the 12 μm depth of the epitaxial layer). This expected diffusion distance, obtained from studies of similar devices [88], indicates that a simulation of nine pixels, in a 3×3 grid, may be sufficient.

Within the solution toolbox, boundary conditions may be specified in two ways: Dirichlet and Neumann, which are shown in Equations 5.3 and 5.4 respectively. In these equations, a , b and c are constants to be specified at the boundary.

$$\rho = a, \quad (5.3)$$

$$\hat{n} \cdot (a\nabla\rho) + b\rho = c. \quad (5.4)$$

Since only a small portion of the pixel array is simulated, the solution is required to behave at the boundary as if the geometry continued outwards. This cannot be achieved with the restrictive boundary conditions provided by Equations 5.3 and 5.4.

Neumann boundary conditions on a small grid are often used in charge sharing investigations, inadvertently causing reflection of charge back into the geometry and an overestimation of the collection efficiency. One proposed remedy [140] is to surround the geometry with a region of strong sink coefficient which consumes any charge reaching the boundary, preventing reflection. However, this is also imperfect because the modification of the charge density will propagate back to the central region and lead to underestimation of the charge collected. Instead, a large grid of 9×9 pixels was simulated, with the charge density constrained to be zero at the outer boundary. This array was found to be large enough to prevent significant charge density reaching the outer boundary during the simulation time, thus making the boundary conditions less important.

Initial charge deposits were located within the central pixel and only the central set of nine pixels were used for data collection. The arrangement of these nine pixels is shown in Fig. 5.22, together with the initial locations of 36 simulated charge deposits. The diodes are modelled as a central square area in each pixel, with the surrounding region representing the epitaxial layer. The whole geometry is defined by three free parameters:

1. **The diffusion coefficient, α .** This applies to both the epitaxial layer and the diodes.
2. **The sink term, $\beta\rho + \gamma$.** This applies to the diode areas only. Initially, β was chosen to be zero, thus giving a constant charge sink rate per unit area. However, this allowed the charge density to become negative, which cannot be prevented within the software

environment. As an *ad hoc* solution, $\gamma = 0$ was instead chosen, resulting in a single sink parameter, β . Thus the sink rate becomes negligible as ρ approaches zero, preventing negative charge density.

3. **The diode width.** This is an effective width and so cannot be assumed equal to the actual diode size.

Note also that no boundary conditions or constraints were imposed at the diode and pixel borders (since no physical boundary exists).

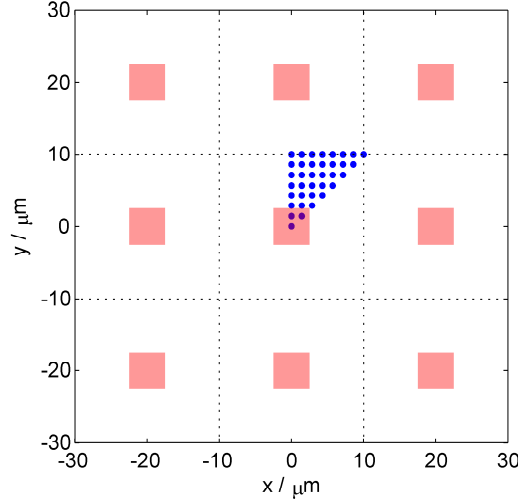


Figure 5.22: Central portion of the charge diffusion simulation geometry, showing nine pixels (the full grid consists of 81 pixels). Pixel diodes are indicated by shaded regions. Thirty-six separate simulations were run, each with initial charge located at one of the points indicated in the central pixel. Borders between pixels are indicated for clarity, although no boundary is simulated at these positions.

Initially, the charge sharing model is to be used to predict the qualitative features of the pixel spectrum resulting from an exposure to a soft X-ray source, with a photon energy of 5.9 keV. At this energy, the interaction and ionisation process for each event is expected to occur in a small volume of order $1 \mu\text{m}^3$ (based on the mean range in silicon of the photoelectron produced). The charge cloud from the particle interaction is therefore modelled as a two-dimensional Gaussian distribution with standard deviation width of $1 \mu\text{m}$, centred at the hit point. This spot is scaled to give a total integral of 1616 electrons.

The MATLAB software provides the charge density solution at the mesh grid points throughout the geometry as a function of time. This must be numerically integrated over the diode boundaries to find the net current into each diode at each point in time. Using the integral of this function, together with the charge present within the diode area at the start and end times, the total collected charge was calculated. A total simulated time of $5 \mu\text{s}$ was allowed, although all nine central diodes reached a steady state after approximately $2 \mu\text{s}$ in all cases. One such simulation requires a computation time of ~ 10 minutes on a personal computer.

5.3.4 Choice of simulation parameters

The diffusion coefficient may be related to the charge carrier mobility, μ , by the following equation [141]:

$$\alpha = \frac{\mu k T}{q}, \quad (5.5)$$

where q is the carrier charge, k the Boltzmann constant and T the absolute temperature. Using a value of $\mu = 0.14 \text{ m}^2 \text{ V}^{-1} \text{ s}^{-1}$ for the electron mobility in intrinsic silicon [142] at room temperature ($T = 298 \text{ K}$), this predicts that $\alpha = 36 \text{ cm}^2 \text{ s}^{-1}$.

Using the results of preliminary simulations, β and the diode size were adjusted in order to reproduce approximate characteristics which have been reported for similar sensor designs. These include a charge collection time of order 100 ns [88], and a central pixel collection fraction of 25% to 50% [143]. This was rapidly achieved with a choice of $\beta = 10^2 \text{ } \mu\text{s}^{-1}$ and a diode effective width of 5 μm (slightly larger than the actual physical dimension of 3 μm); further tuning has not yet been attempted.

5.3.5 Charge sharing response function

Examples of the simulation results for a single hit point at coordinates of (4.3, 5.7) μm are shown in Figs. 5.23 and 5.24. Figure 5.23 shows the calculated net current flowing into the diode located at (−20, 20) μm as a function of time, together with the cumulative integral of this function. This shows that, with this choice of parameters and deposit location, most of the signal is collected in less than 1 μs .

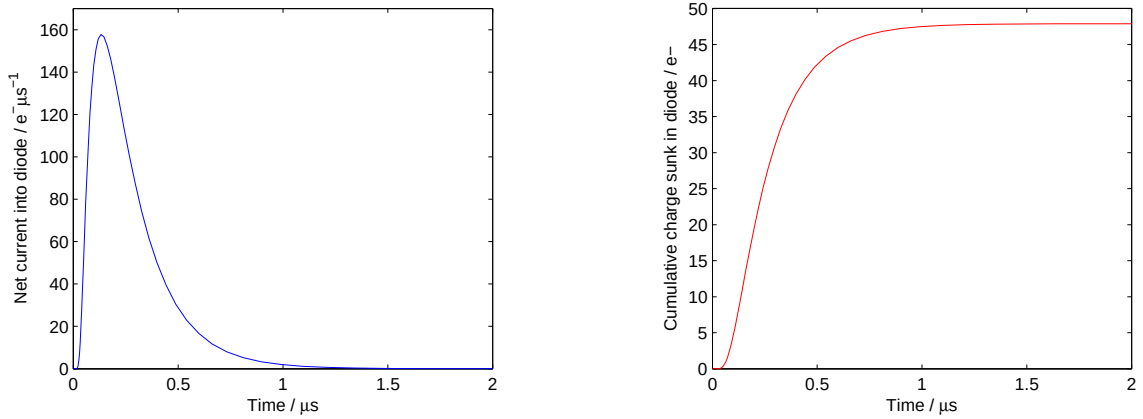


Figure 5.23: Simulated inward net current (left) and cumulative charge collected (right) for the diode at (−20, 20) μm as a function of time. A charge of 1616 e^- was released at (4.3, 5.7) μm at zero time.

Fig. 5.24 shows in bar chart form the charge collected in each of the central nine diodes for this simulation, expressed as a fraction of the initial charge present (1616 electrons). These values were taken after a simulated period of 5 μs .

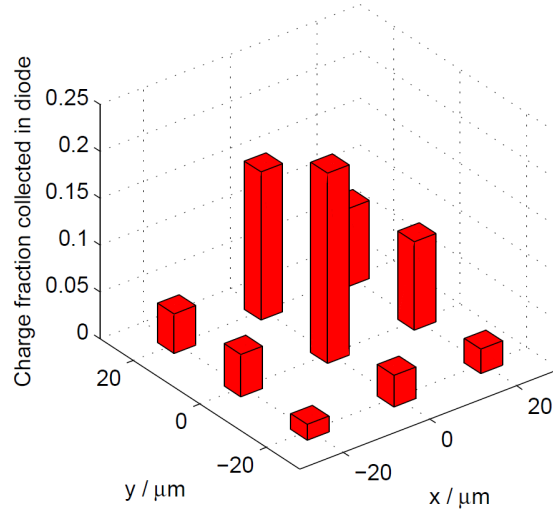


Figure 5.24: The charge fraction collected by the nine test diodes after a period of 5 μs . A charge of 1616 electrons was initially located at (4.3, 5.7) μm .

The simulation and calculation process was carried out a total of thirty-six times, using the initial deposit locations shown in Fig. 5.22. For each location, the charge share obtained by each of the nine diodes was determined. Due to the symmetry of the array, the results for this one-eighth portion of the pixel may be reflected and duplicated to cover the entire central pixel. Thus the charge share collected by all nine surrounding pixels is known for any hit location (after linearly interpolating between the locations tested). These data have been used to construct the basic response function shown in Fig. 5.25. This gives the charge fraction collected by a diode as a function of the hit location relative to that diode and is shown over a $60 \times 60 \mu\text{m}^2$ area, corresponding to a 3×3 pixel group.

Figure 5.26 shows the total charge collected by the nine diodes closest to the initial deposit point, as a function of hit position inside the central pixel. Note that little attempt has been made to tune the parameters at this stage, and the true peak value of this function for the HMRM sensor may be greater than the $\sim 80\%$ shown. This function shows a central peak surrounded by a broad plateau at values of approximately 0.72, with minima at each pixel corner. These features are due to the square pixel grid layout and the inter-diode spacings that result from this pattern.

The charge collection time has been assumed small compared to the total exposure time of one readout frame (100 μs). This means that charge collection is likely to be complete when the pixels are read, with only a small probability of the collection being interrupted by the readout. Figure 5.27 confirms this assumption, showing that the majority of the pixel's charge is collected in under 1 μs at all hit points.

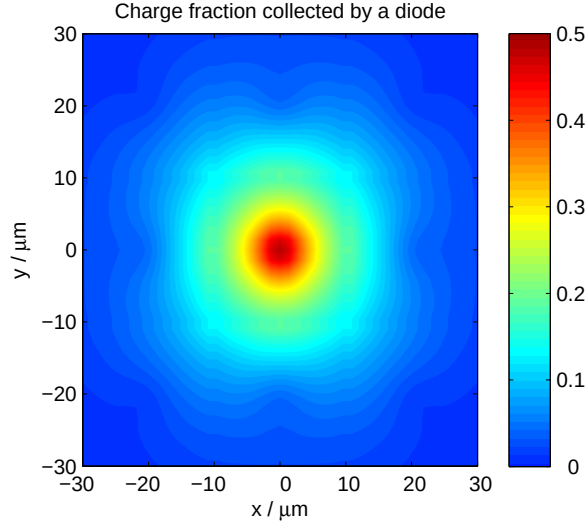


Figure 5.25: Charge sharing response function, showing the fractional charge collected by a single pixel diode. This is shown for hits occurring anywhere within a 3×3 pixel grid centred on the test diode.

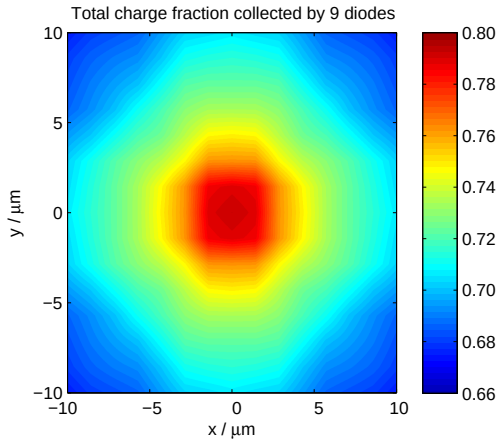


Figure 5.26: The total fractional charge collected by the nine diode grid (as a function of hit location within a single pixel).

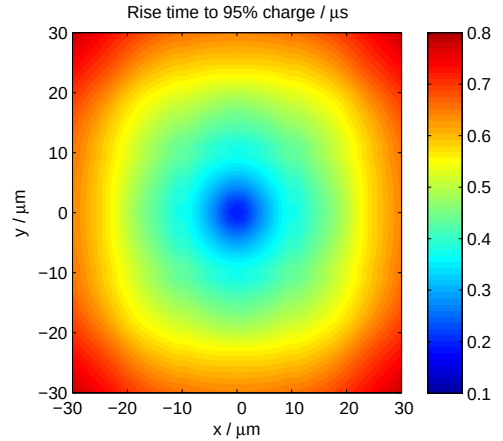


Figure 5.27: The time taken for a diode to collect 95% of its long-term charge (as a function of hit location within a 3×3 pixel grid).

5.3.6 Implementation: 5.9 keV X-ray pixel spectrum

The results obtained in the previous section have been applied to the detection of 5.9 keV photons; this is one of the X-ray energies emitted by the decay of the ^{55}Fe isotope and is frequently used for sensor calibration. Two results are sought: the single pixel spectrum and the ‘charge sharing noise’. The former provides a qualitative prediction for the appearance of a ^{55}Fe calibration spectrum for a pixel sensor, including any spectrum peaks or features.

The latter quantity is defined as the standard deviation of the cluster sum distribution for a constant input signal (photon energy), when charge diffusion is the only source of signal variance. This may be compared with other noise sources (e.g. thermal, Fano, shot), obtained from separate considerations, to assess the importance of charge sharing in determining the overall sensor performance.

As previously discussed, the whole photon energy is expected to be deposited in a small volume at each interaction point, which occur randomly over the sensor array and at all depths in the epitaxial layer. A significant disadvantage of 2-D simulation is the inability to account for absorption at these different depths. Hits in the third (depth) dimension can occur at increasing distances from a given diode without being nearer to other diodes, unlike the situation in a 2-D geometry. This fundamental difference will affect the pixel spectrum, although the size of this discrepancy is currently unclear.

A further consequence of photon exposure is the occasional complete collection of an event by a single diode, assumed to occur when the photon interacts inside the N-well diode depletion region. Since the charge in this case is not collected by diffusion, these events are not described by this model. Instead, these may be included by assuming a simple cross-section for this interaction and adding the corresponding number of full-energy pixel signals (plus noise) to the final spectrum. This has been estimated to affect $\sim 0.1\%$ of the interacting photons [88]. If included, this direct absorption probability constitutes a fourth parameter of the model.

The charge sharing response model consists of the following steps for each primary energy deposit value:

1. Convert the energy deposit, E_{dep} , into ionisation charge: $Q = E_{dep}/W$, with $W = 3.65$ eV.
2. Randomly sample a hit location within the arbitrary ‘central’ pixel.
3. Using the hit location, allocate a charge share to each of the nine nearest pixels according to the function shown in Fig. 5.25.
4. Add a randomly sampled, uncorrelated, Gaussian-distributed noise value to each pixel charge share.
5. Generate an appropriate proportion of noise-only pixels (according to the expected photon flux and sensor area).
6. If required, generate a number of full-energy pixels according to the direct-absorption probability (and add noise).
7. Combine all of the resulting pixel values into a single probability distribution.

This process makes several simplifications, including the neglect of cluster pileup (two or more 3×3 regions overlapping), hits near the array border or near defective pixels, and hits occurring within $1 \mu\text{s}$ of the end of each exposure time. Note that although the diffusion simulation itself is deterministic, the hit location and added noise are Monte Carlo sampled.

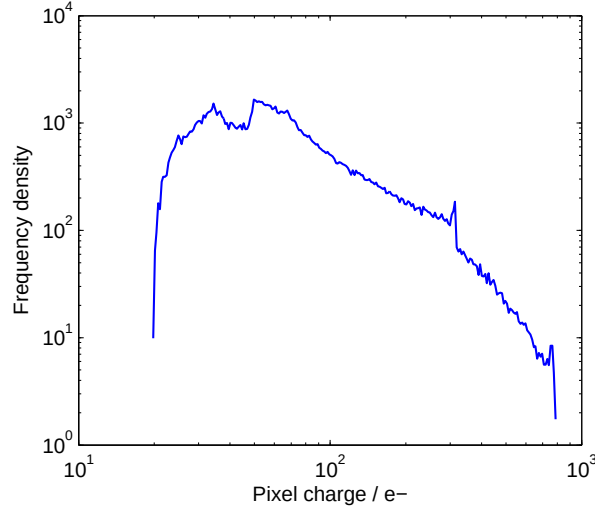


Figure 5.28: Distribution of pixel charge shares simulated with incident 5.9 keV photons (modelled as a constant 1616 e^- initial charge). Pixel noise has not been added.

The result of step 3, the ‘pixellation’ stage, is shown in Fig. 5.28; four peaks are observed in this spectrum. The two most probable peaks, at approximately 33 e^- and 50 e^- , correspond to charge shares of 2–3%. These occur when a photon is absorbed close to a pixel diode, meaning that the surrounding eight of the cluster all collect a similar charge share, seen in Fig. 5.25 to be 0–5%. The double peak arises from the slight difference between the four corner diodes (collecting less) and the four side diodes (on the x and y axes, collecting more). Meanwhile, the central pixel in these cases collects up to 50% of the charge, resulting in an endpoint peak at $\sim 800 e^-$.

The peak at $\sim 3000 e^-$ is a result of a further symmetry of the array. This corresponds to the case where a photon is absorbed near to a pixel corner (equidistant from four diodes) or near the centre of a pixel edge (equidistant from two diodes). In both cases, the diode group (two or four) receive similar charge shares of 15–20%, giving a small probability increase at these values.

Figure 5.29 shows the final result, having simulated 15,000 photon absorptions and 170,000 noise-only pixels. The noise convolution has caused the two low-charge peaks to merge; however, the other features are still visible. Note that no direct absorption events (diode hits) were added and so this distribution does not contain a full-energy photopeak.

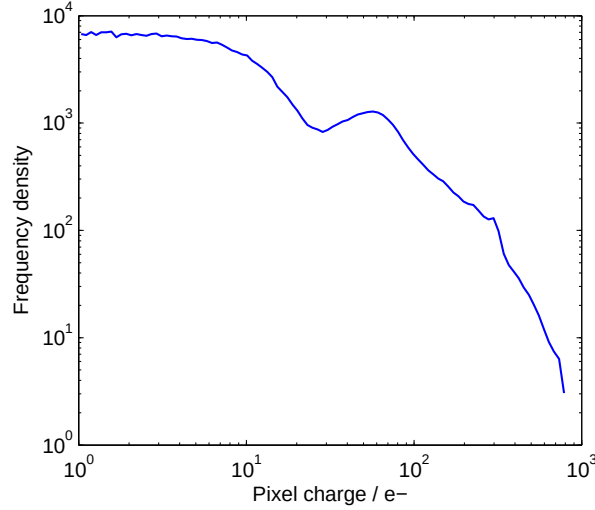


Figure 5.29: Distribution of pixel charge shares simulated with incident 5.9 keV photons after adding a pixel noise of 10 e^- RMS. Noise-only pixels have also been included.

A sensor exposed to ^{55}Fe X-ray emission is predicted to produce a single pixel distribution similar to Fig. 5.29. In addition, the measured distribution is expected to contain a photopeak and increased variance due to the other noise sources. These will depend on the individual sensor and have not been included in order to reveal the contribution due only to the charge sharing. The simulation parameters may be adjusted to obtain a close fit to experimental measurements. The model may then be used to determine quantities, such as pixellation resolution limits or cluster sum ‘noise’, that cannot be measured directly.

Cluster summation

In HMRM radiation monitoring mode, the whole array of pixel values is summed to provide an estimate of the total energy deposited by a particle. The variance of this reconstructed sum, for a constant input signal, determines the sensor resolution. This is influenced by many factors including the pixel noise, charge collection and digitisation system, as well as aspects of the interaction itself, such as the hit location and particle path. However, this analysis seeks only the contribution due to charge sharing. Non-cluster pixels are ignored so that only the nine pixels closest to the interaction point are summed. This leaves incomplete charge collection and the artificially added pixel noise as the only sources of variance in the simulated data.

Figure 5.30 shows the distribution of cluster sums from the X-ray photon simulation already considered. A function fit shows that this distribution has an approximately Gaussian shape, with a mean collected charge of 1191 e^- (74% of the total). The function standard deviation gives the effective noise inherent in the cluster sum process due to both the individual pixel noise and the charge sharing.

The added pixel noise, σ_p , is expected to contribute $9\sigma_p^2$ to the variance of a cluster sum of

nine pixels. This leaves an effective ‘charge sharing noise’, σ_c , of:

$$\sigma_c = \sqrt{\sigma_s^2 - 9\sigma_p^2}, \quad (5.6)$$

where σ_s is the total cluster sum noise. Using the fitted value for σ_s , the value of $\sigma_c = 43.8 \text{ e}^-$ is obtained for the charge sharing noise. This is greater than the Fano noise for photons of this energy ($\sim 13 \text{ e}^-$) and is expected to be similar in magnitude to the typical thermal and fixed pattern noise combined.

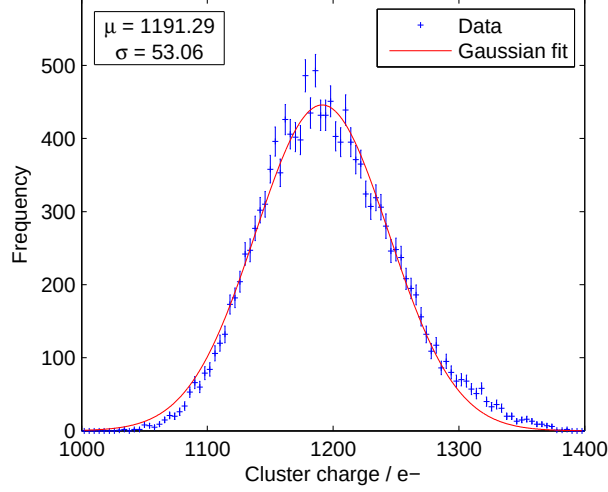


Figure 5.30: Distribution of the pixel cluster sums after simulated charge sharing of 5.9 keV photon absorption events (assuming a constant 1616 e^- initial ionisation charge per event). Each event was shared among nine pixels and a randomly sampled 10 e^- RMS equivalent noise charge added to each pixel value.

5.3.7 Charge sharing conclusion

The charge sharing model has predicted an individual pixel spectrum peaking at an intermediate value ($\sim 60 \text{ e}^-$), in addition to the previously expected noise and full-energy peaks. This third, ‘charge sharing’ peak has been observed experimentally in other investigations (see, for example, [144] and [145]), demonstrating that the model has at least qualitative utility.

It should be noted that the numeric results obtained in this section are not expected to apply exactly to the HMRM. However, an attempt will be made to tune the model parameters to match measurements taken with the HMRM sensors, although quantitative use may be limited by the reduction to two dimensions. Future work could investigate a 3-D approach or use a commercial simulator such as Synopsys’ *Sentaurus* [146] or Silvaco’s *Atlas* [147]. These achieve greater accuracy through simulation of the semiconductor doping and electric field.

Model calibration may also be facilitated using a scanning laser [92] with a spot size of order $1 \text{ }\mu\text{m}$: this can be used to map the charge sharing response as a function of spot position. Finally, an extension to the description of charged particle interactions will require

a different initial charge distribution. High energy particles will cause ionisation along their path through the epitaxial layer and substrate, with possible secondaries also contributing. Initial spot and line charges, however, may be sufficient as a first approximation for MIPs near normal incidence.

5.4 Summary

Comparisons of experimental particle data with results generated by the Geant4 toolkit have demonstrated that these simulations are generally very accurate. Energy deposits and particle energy losses in thin and thick material layers are well modelled, with typical discrepancies which are consistent with experimental uncertainty. Multiple scattering of proton beams and depth-dose curves of high energy ion beams have also shown excellent agreement. Electron backscattering simulations show the correct distribution features but have larger errors of typically $\sim 10\%$.

These results confirm that Geant4 Monte Carlo simulations may be relied upon to evaluate the monitor design and to provide accurate response functions. Validation of the monitor geometry implementation must still be undertaken by comparing experimental HMRM data (e.g. from particle beam exposure) with simulation results.

Time-dependent simulations of the monitor in a modelled Earth orbit environment are useful for the assessment of dynamic effects such as the electronic shuttering algorithm. The suitability of integration period choices and the success of the whole physics algorithm are also tested in this way. The generated data packets may be used as inputs to test post-processing analysis methods, such as energy spectrum reconstruction.

Finally, charge diffusion simulations have provided insight into the response characteristics of pixel sensors; further calibration work may allow this model to be applied to the HMRM.

Chapter 6

Reconstruction of particle energy spectra

A complete characterisation of a radiation environment is provided by the energy spectra of all particle species present. In general, the unidirectional differential spectrum of each species shows the notional instantaneous mean flux, as a function of energy, direction and time. In consequence, the spectra may be used to derive information about radiation effects on any object in that environment.

This chapter explores methods which may be used to convert monitor channel counts into reconstructed spectra. A popular ‘matrix method’ is first described and tested in various situations. Several deficiencies in this method, however, lead to a new approach, proposed in this chapter. Although still in development, preliminary tests of this new algorithm are presented using modelled HMRM output data.

6.1 The inverse problem

As introduced in Chapter 3, the monitor response functions may be used to predict the particle counts in each of the 32 monitor channels in a given radiation environment. This is a simple evaluation of the formula:

$$\mu_k = \sum_s \int_0^T \int_{4\pi} \int_0^\infty \epsilon_k^s \alpha^s f_u^s dE d\vec{\Omega} dt, \quad (6.1)$$

where μ_k is the mean count expected in channel k in an exposure time T . Here, the response functions ϵ_k^s (identification efficiency) and α^s (effective area) are unidirectional and require the unidirectional flux, f_u^s , to be specified. The sum is over all particle species, where superscript s indicates species dependence.

Although this forward evaluation is simple, the inverse problem, where the measured counts (N_k) are provided and the f_u^s must be found, is more challenging. An additional problem lies in the fact that the N_k are integer random variables, while Equation 6.1 is stated in terms of the corresponding underlying mean values, μ_k . The solution method must therefore contend with statistical fluctuation in any finite measurement sample.

6.1.1 Assumptions

A number of simplifying assumptions are introduced in order to make the solution of Equation 6.1 more tractable — these are used for all analyses presented in this chapter.

1. **Flux isotropy.** Incident fluxes are assumed isotropic as observed by the monitor, allowing fluences and response functions to be represented by their omnidirectional integral values.
2. **Particle species.** Analysis is limited to electrons and protons only, as the two most generally abundant species.
3. **Response linearity.** Monitor channel counts are assumed to be proportional to the incident fluence. This means that the response functions are independent of the flux magnitude and that counts are, on average, proportional to exposure time in any constant flux. This is equivalent to assuming that pile-up effects are negligible.

These three assumptions allow Equation 6.1 to be reduced to simple terms for electrons and protons:

$$\mu_k = T \int_0^\infty R_k^e f_o^e dE + T \int_0^\infty R_k^p f_o^p dE, \quad (6.2)$$

where, for example, R_k^e is the electron omnidirectional response function for channel k and f_o^p is the omnidirectional proton flux; both of these quantities are functions of particle energy.

6.2 Matrix method

The simplest solution of Equation 6.2 assumes that the incident fluxes, f_o^s , are piecewise constant over a series of primary particle energy bins. This allows the integrals to be replaced by sums:

$$\mu_k = T \sum_{i=1}^{m_e} C_{ik}^e F_i^e + T \sum_{j=1}^{m_p} C_{jk}^p F_j^p, \quad (6.3)$$

where:

$$C_{ik}^s = \int_{E_{i,\min}^s}^{E_{i,\max}^s} R_k^s dE, \quad (6.4)$$

and:

$$F_i^s = \int_{E_{i,\min}^s}^{E_{i,\max}^s} f_o^s dE. \quad (6.5)$$

The flux and response functions are represented by their integrals over discrete energy bins for each species, of which there are m_e and m_p for electrons and protons, respectively. The resulting set of linear equations is usually expressed as a matrix equation giving a vector of mean count rates, $\vec{n}$, having normalised for the exposure time:

$$\vec{n} = C \vec{F}, \quad (6.6)$$

where $\vec{F}$ is a vector of integrated fluxes, each given by Equation 6.5. The response matrix C , of size P by $(m_e + m_p)$, where P is the number of monitor channels, is independent of flux

and is characteristic of the monitor itself. Note that the two species have been combined into a single set of simultaneous equations.

This ‘matrix method’ is extremely common in radiation monitor analyses (see, for example, [148, 149, 150, 151]). However, this matrix notation is not especially useful since the inverse matrix solution cannot, in general, be used. This is partly because an inverse of C does not usually exist but is also because measured count rates are used as inputs, rather than the true population means. Therefore, each linear equation contains a random error, necessitating a fitting procedure to find the $\vec{F}$ that best agrees with the data.

Fitting is usually achieved via the method of least squares, with weightings indicating the variance of the measured channel counts. In this case, the weighted least squares parameter, χ^2 , for which a minimum is sought, is given by:

$$\chi^2 = \sum_{k=1}^P \frac{(N_k - \hat{\mu}_k)^2}{\hat{\mu}_k}, \quad (6.7)$$

where it is assumed that the channel counts are Poisson distributed and that the count predictions, $\hat{\mu}_k$, are large (note that χ^2 is undefined if any channel is predicted to contain zero count). There are a total of P monitor channels: for the HMRM, $P = 32$. Extensions to prevent negative fluxes are also possible [150]; note, however, that these alterations to the basic least squares procedure mean that the problem is no longer linear. Even this simple spectrum model requires the use of more complex non-linear optimisation routines to search for the function minimum.

A possible alternative to least squares is to use a maximum likelihood estimate as the fitting objective function [152]. This may be especially useful as it allows explicitly for the statistics of small count samples, including channels with zero counts. For example, assuming a simple Poisson probability model for each channel count, the negative log-likelihood parameter is given by:

$$-\ln L = \sum_{k=1}^P [\hat{\mu}_k - N_k \ln \hat{\mu}_k + \ln N_k!]. \quad (6.8)$$

6.2.1 Test implementation

Performance of the matrix method is tested in this section using a simplified HMRM modelling procedure, where Geant4-derived energy deposit data are used directly in the ID table look-up for allocation to a monitor channel. This omits the HMRM response model, where the effects of sensor noise and pixellation are added. The results are therefore idealised and correspond to the situation where the sensors are perfect detecting devices; this is sufficient to investigate the most salient aspects of the spectrum reconstruction process. The modelled channel counts are used to reconstruct the proton and electron spectra in a weighted least squares fit.

An example of a set of channel counts is shown in Fig. 6.1, obtained using a simulated E^{-1} differential energy spectrum for each species and the ID look-up table presented in Chapter 3. The total generated particle fluences were $3.5 \times 10^6 \text{ cm}^{-2}$ over the range 0.04–6.0 MeV for electrons and $8.7 \times 10^4 \text{ cm}^{-2}$ over the range 1.0–500 MeV for protons, resulting in a total of 8400 electron counts and 600 proton counts.

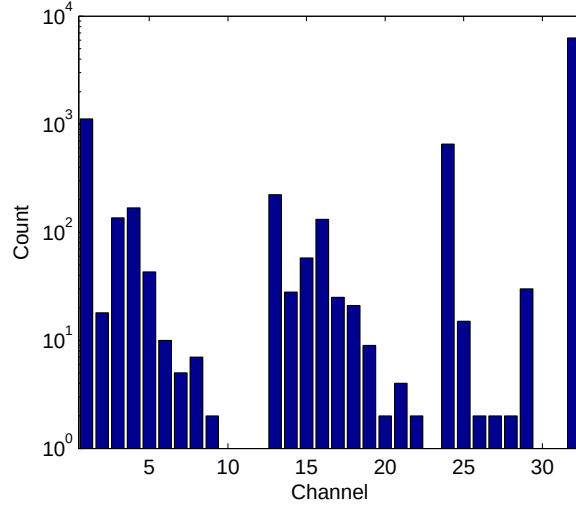


Figure 6.1: HMRM channel counts allocated in response to a simulated flux of protons and electrons with E^{-1} differential energy spectra (the total fluences generated were $8.7 \times 10^4 \text{ cm}^{-2}$ for protons and $3.5 \times 10^6 \text{ cm}^{-2}$ for electrons).

Further sets of counts were randomly sampled, using the same equivalent exposure, and each set was used in a separate reconstruction of the particle energy spectra. For each reconstruction, the energy bin edges were chosen to be geometrically spaced with five bins for each species. The resulting spread of reconstructions is shown in Figs. 6.2 and 6.3 together with the simulated spectrum functions.

Also shown in these figures is the mean number of counts of each species as a function of primary energy. The reconstructed spectra are deconvolutions of this count spectrum so that particular features do not correspond directly between the two functions. However, small numbers of input counts across an extended part of a species energy range generally translate into greater uncertainty in the reconstructed spectrum over the same range, indicated by the greater spread in the reconstructions. While this uncertainty is mostly due to the greater fractional variance of small counts, an additional problem is the reduced particle discrimination over certain energy ranges, or between species, as discussed in Chapter 3.

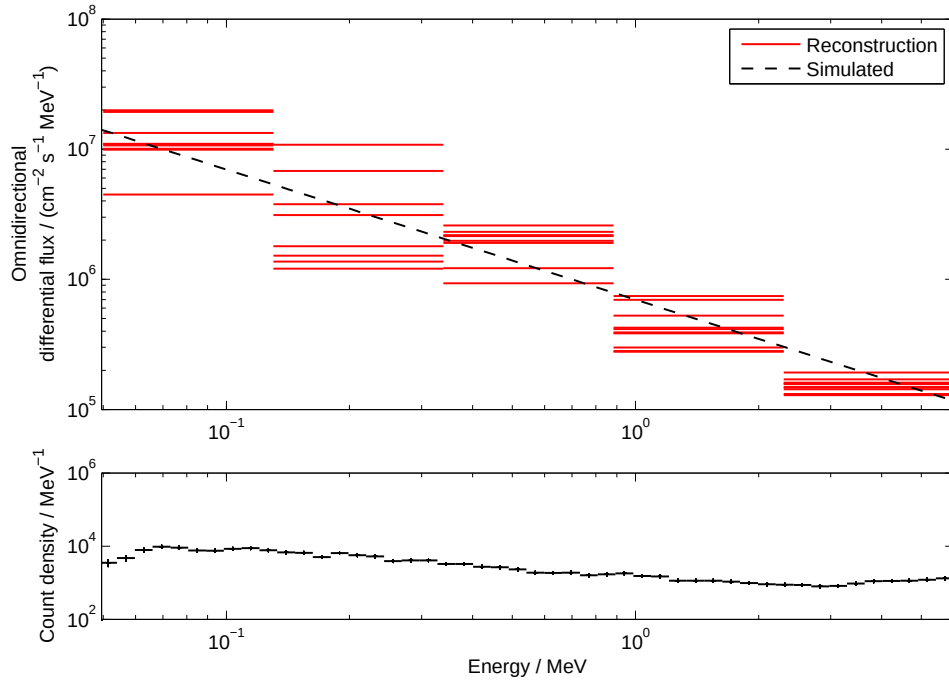


Figure 6.2: Matrix method spectrum reconstruction results for electrons, using an incident mixed fluence of $8.7 \times 10^4 \text{ cm}^{-2}$ protons and $3.5 \times 10^6 \text{ cm}^{-2}$ electrons. The results from ten separate data samples are shown, each using five primary energy bins. The lower panel shows the true number of counts as a function of energy.

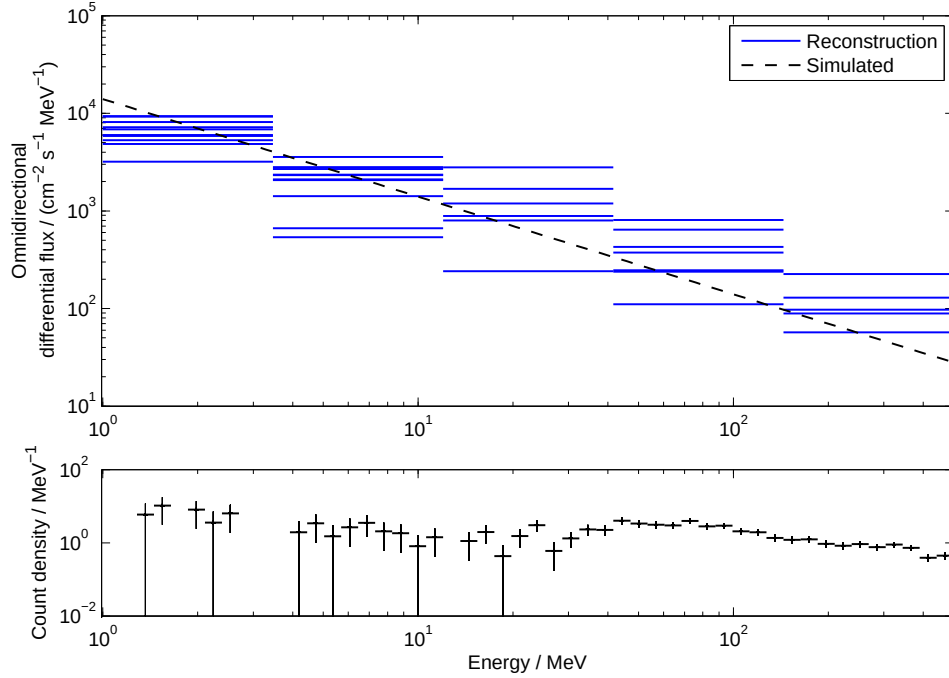


Figure 6.3: Matrix method spectrum reconstruction results for protons, using an incident mixed fluence of $8.7 \times 10^4 \text{ cm}^{-2}$ protons and $3.5 \times 10^6 \text{ cm}^{-2}$ electrons. The results from ten separate data samples are shown, each using five primary energy bins. The lower panel shows the true number of counts as a function of energy.

6.2.2 Size of the data sample

The amount of data required for successful spectrum reconstruction is an important consideration due to its influence on the channel count variance. This requirement will depend partly on the spectrum shape, although the general dependence is investigated here using different integrated fluences of E^{-1} spectra. This is approximately equivalent to varying the monitor exposure time before reconstruction is attempted. This cannot be extended arbitrarily due to time variation of the flux and movement of the spacecraft in orbit.

Figure 6.4 shows reconstruction results using fluences that are 30 times greater than those used for Figs. 6.2 and 6.3, while Fig. 6.5 shows the results from fluences 30 times smaller. There is a clear change in the reconstruction ability, with increased uncertainty spread for the lower fluences.

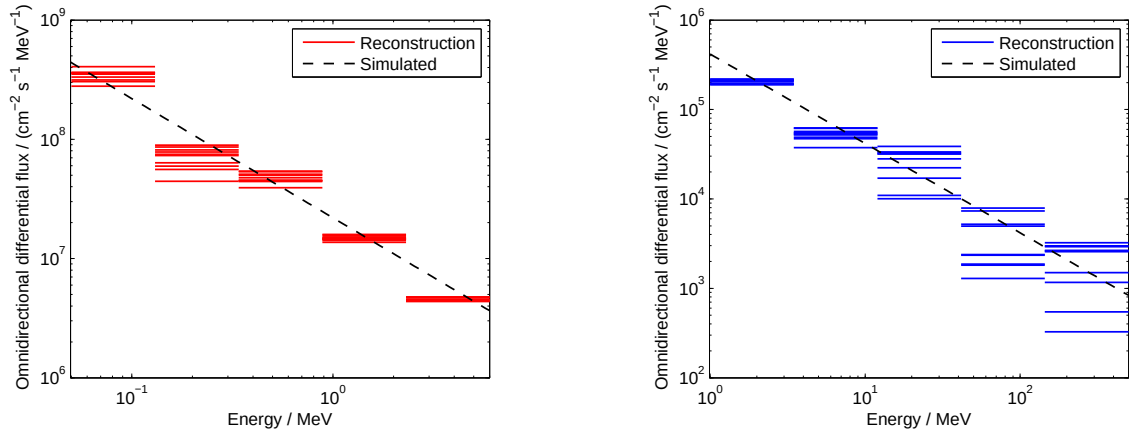


Figure 6.4: Matrix method spectrum reconstruction results for electrons (left) and protons (right), using an incident mixed fluence of $2.6 \times 10^6 \text{ cm}^{-2}$ protons and $1.1 \times 10^8 \text{ cm}^{-2}$ electrons. Ten sample reconstructions are shown, each using five primary energy bins per species.

The RMS difference between the true spectra and the fit result provides a figure of merit for the reconstruction (this RMS error should not be confused with the least squares fitting parameter, defined in Equation 6.7). This quantity is shown as a function of the total input channel counts in Fig. 6.6, having maintained a constant fluence ratio between the two species and a fixed number of energy bins (twelve per species) throughout. This graph shows that the reconstruction quality follows the usual $N^{-\frac{1}{2}}$ dependence, and is thus expected to display a $T^{-\frac{1}{2}}$ dependence on exposure time, assuming a constant incident flux.

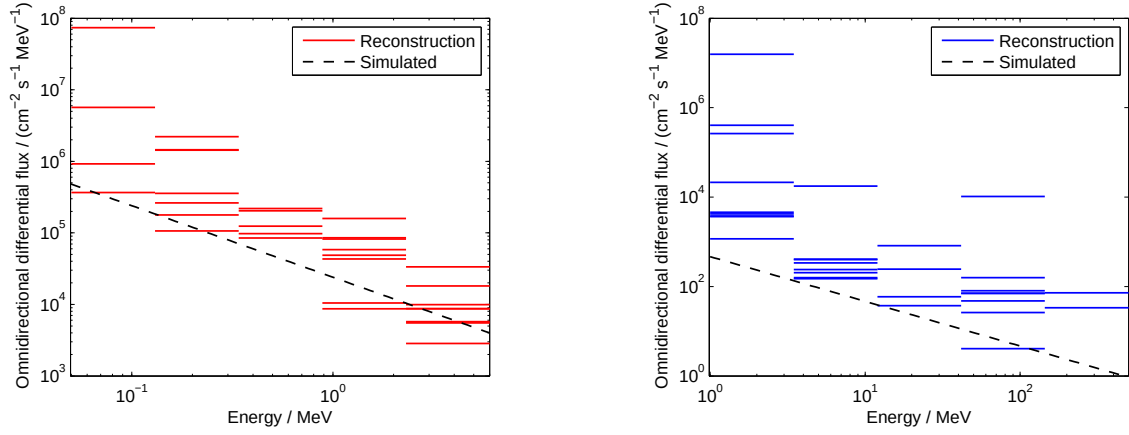


Figure 6.5: Matrix method spectrum reconstruction results for electrons (left) and protons (right), using an incident mixed fluence of $2.9 \times 10^3 \text{ cm}^{-2}$ protons and $1.2 \times 10^5 \text{ cm}^{-2}$ electrons. Ten sample reconstructions are shown, each using five primary energy bins per species.

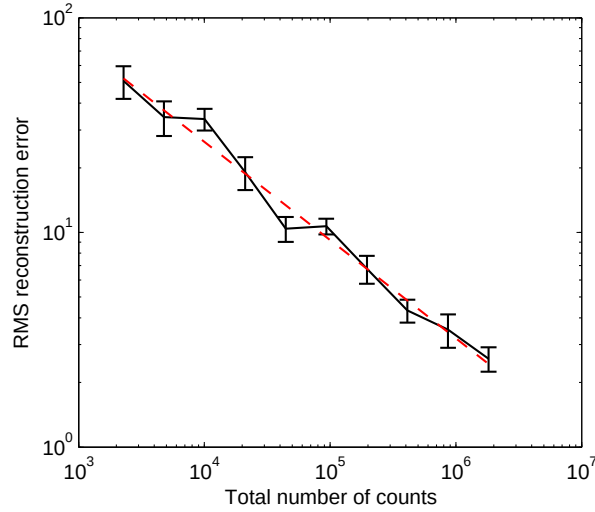


Figure 6.6: Combined fractional RMS errors for proton and electron E^{-1} spectra reconstructed using varying quantities of input count data. The dashed line indicates a power law function with index equal to $-\frac{1}{2}$.

6.2.3 Species flux ratio

A further concern is the extent to which each species' energy spectrum can be reconstructed in a mixed species environment. This is especially important in Earth orbit, where electron fluxes may outnumber protons by many orders of magnitude. Figure 6.7 shows the combined and separate RMS errors of the electron and proton spectra in E^{-1} particle fluences of varying relative amplitude. This is shown as a function of the electron/proton count ratio while maintaining a constant overall total number of counts. It is also shown as a function of total

proton count in the presence of a constant electron count (in which case the total number of counts changes). These graphs confirm that, for these spectra and choice of ID table, the best results are achieved with approximately equal numbers of counts. Meanwhile, the reconstruction of one species can be improved by increasing the number of its counts, although this generally comes at the expense of the other species' reconstruction.

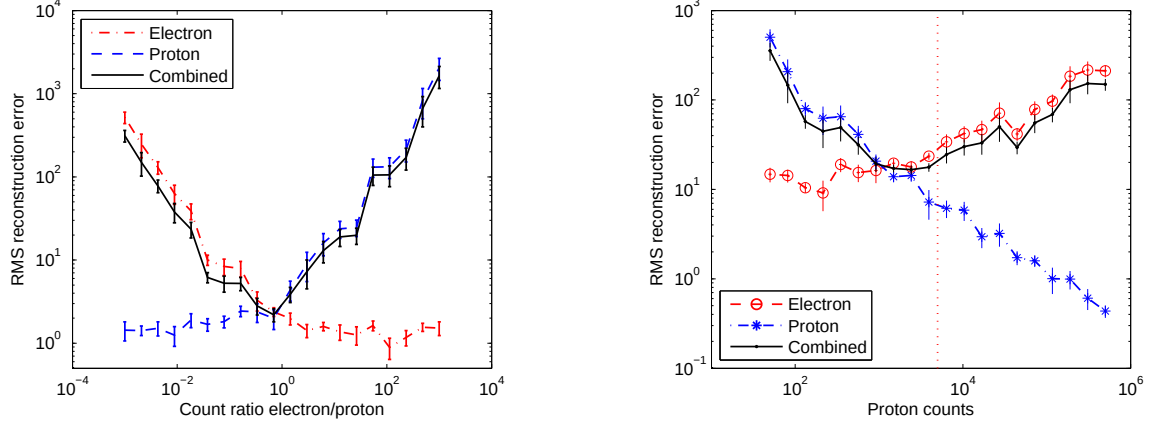


Figure 6.7: Fractional RMS errors for E^{-1} energy spectra reconstructed using varying quantities of input count data. Left: error as a function of electron/proton count ratio, while maintaining a total of 10^5 counts. Right: error as a function of the total proton count with a constant 5×10^3 electron count (vertical dotted line).

6.2.4 Number of spectrum energy bins

The number and size of the spectrum primary energy bins are important choices within the matrix spectrum reconstruction method. A large number of bins promises a better agreement with the true spectrum and a higher-resolution response matrix. However, the fixed number of input data (32 channel counts) implies a maximum number of allowed free parameters before the problem becomes under-determined.

The best fit flux for some spectrum energy bins, in low fluences, may be a negative value due to statistical fluctuations. As well as being unphysical, this has the general consequence of causing a compensatory increase in the amplitude of the adjacent bins, appearing as an oscillatory pattern. Non-negative least squares fitting algorithms, although preventing a negative flux, cannot fully solve this problem as the fitted fluxes may still vary by many orders of magnitude.

Oscillation is interpreted as indicating that the problem is under-determined and that the number of spectrum energy bins is too large for the available input data. This may either be due to small particle counts or insufficient differences in monitor response between these energy bins. In these situations, the fit may be attempted with fewer bins, maximising both the counts per bin and the differences between the monitor response matrix elements.

However, in reducing the number of bins the piecewise constant function model becomes

an increasingly poor fit to real spectra. This is demonstrated by a simulated exposure in the modelled mean radiation environment of reference orbit E, a 35,786 km geostationary orbit. Figures 6.9 and 6.10 show ten repeated reconstructions using six and three bins per species, respectively. The input data sample corresponds to a short exposure of 110 s in orbit (assuming the monitor electronic shutter limits the frame exposure to the 3.125 μ s minimum); the channel counts are shown in Fig. 6.8.

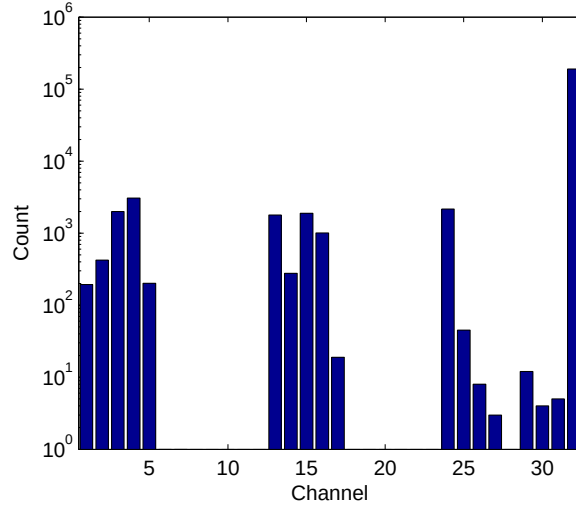


Figure 6.8: HMRM channel counts allocated in response to a 110 s simulated exposure to the mean radiation environment of reference orbit E.

Using six energy bins, an excellent result is obtained for electrons at energies up to 0.5 MeV. However, the fourth and sixth bins have negative or zero values, while the fifth bin has a greater amplitude to compensate. The proton result is poor, with a large spread of results at low energy (1–3 MeV) and extreme oscillation at higher energies.

In contrast, the reconstructions with three bins (Fig. 6.10) show less variation but remain generally inaccurate. The proton and high energy electron spectra show rapid changes with energy that cannot be represented accurately by a piecewise constant model.

6.2.5 Limitations of the matrix method

Despite some success, the matrix method has a number of disadvantages. The limitation of the number of spectrum energy bins leads unavoidably to limited resolution. However, the piecewise constant model also restricts the range of functions that can be fit effectively. This also introduces discontinuities (undershoot/overshoot) at the bin edges, giving a poor fit to real spectra which are assumed to be generally continuous functions of energy. These problems are compounded by the loss of resolution in the response functions due to integration (Equation 6.4) when formulating the response matrix. In this way, the spectrum model forces the assumption of a constant monitor response over each energy bin whereas, as discussed in Chapter 3, the response functions actually vary rapidly with particle energy.

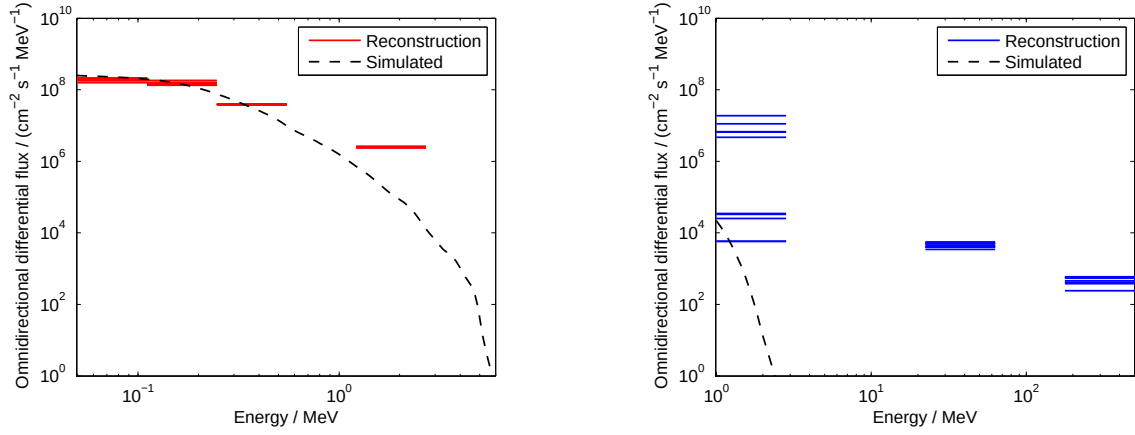


Figure 6.9: Matrix method spectrum reconstruction result for electrons (left) and protons (right) for a 110 s simulated exposure in reference orbit E using six energy bins per species.

6.3 New method

A new method of spectrum reconstruction is proposed here to address some of the disadvantages of the matrix method. The overall procedure consists of many elements, some of which may allow several possible implementation methods. A full exploration of all of these possibilities has not yet been undertaken.

In common with the non-linear versions of the matrix method, this new method is solved through iteration. Electron and proton trial spectra are folded with the monitor response functions to produce a set of predicted channel counts, which are compared with the actual HMRM data. The spectra are then modified and the process repeated until an optimal reproduction of the measured data is obtained.

6.3.1 High resolution response functions

Central to the new method is the preservation of the full monitor response function resolution, typically with the order of 100 energy bins per species. This is achieved by assuming that each spectrum is a continuous function of particle energy and may be modelled as a piecewise continuous power law.

N free parameters are interpreted as a set of flux ‘nodes’ at predetermined energy values. These are interpolated by $(N - 1)$ fully-determined power law function segments to produce a trial spectrum. This spectrum is then used to modulate the response function, which may be of arbitrarily high resolution, to produce a set of channel count rate estimates (the ‘folding’ process). This is illustrated in Fig. 6.11 compared to the analogous procedure for the matrix method; in both cases the spectra are modelled using the same number of free parameters (four matrix bins or four flux nodes). In addition to allowing a greater response function resolution, it is proposed that this new model provides a better fit to real spectra.

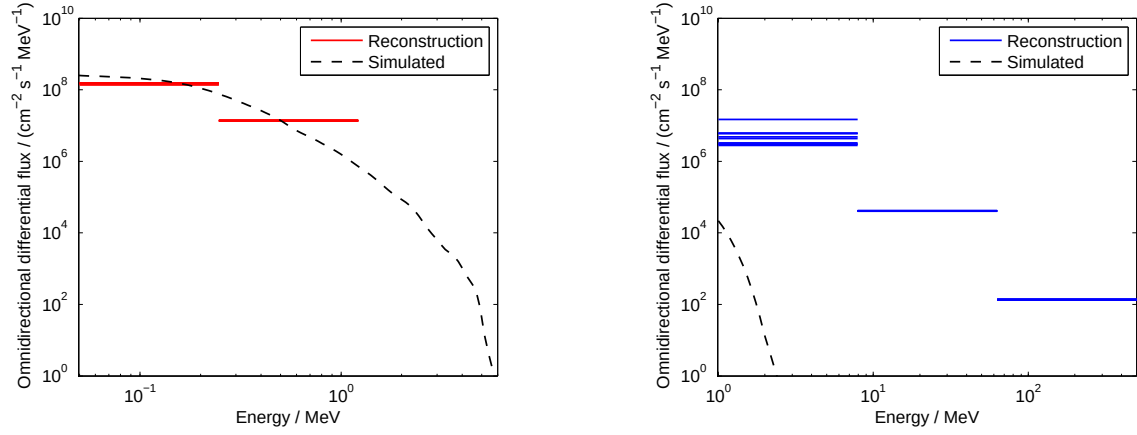


Figure 6.10: Matrix method spectrum reconstruction result for electrons (left) and protons (right) for a 110 s simulated exposure in reference orbit E using three energy bins per species.

Spline interpolation

The change from a constant flux model to a linear dependence per energy bin (on a logarithmic scale) corresponds to an increase in polynomial order. It is appealing to increase the order further still and use continuous spline interpolation. However, two additional free parameters would then be required to constrain the endpoint gradients. In addition, splines often exhibit oscillation and overshoot, creating spurious spectral features which are a function of the model rather than the input data.

6.3.2 Function optimisation

As for the matrix method, the count rate estimates produced by the folding process are compared with monitor measurements in a least squares or maximum likelihood iterative fit. Both species must be reconstructed simultaneously since, in general, each monitor channel may count both species (mixed efficiency) and thus the two trial spectra constitute a combined hypothesis. In each iteration, evaluation of the objective function is a simple extension of the forward method. However, the system of optimisation (variation of the trial spectra to approach the objective, usually a function minimum) is a major determinant of the efficiency and success of the reconstruction.

The new spectrum model means that the problem is non-linear, requiring more sophisticated function minimisation routines, a number of which are available for general use. The MATLAB ‘fminsearch’ function, which uses a Nelder-Mead simplex direct search method [153], has been used for the reconstructions presented in this chapter.

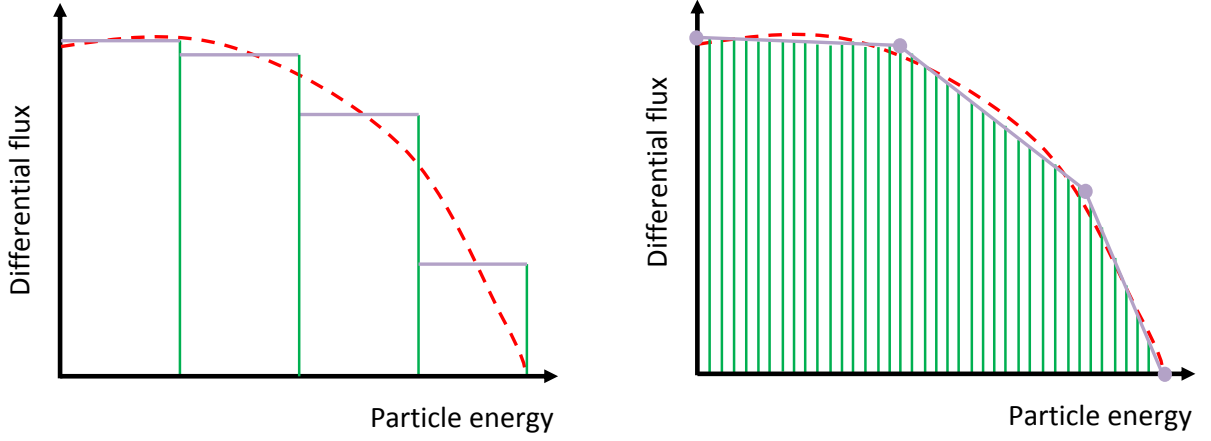


Figure 6.11: Illustration of the spectrum modelling and response function modulation in the matrix method of reconstruction (left) and the new method (right). The number of response function bins greatly exceeds the number of spectrum nodes (free parameters) in the proposed new method, whereas the response is constrained to be piecewise constant in the matrix method. The true spectrum is shown as a dashed line (graph axes are shown with logarithmic scales).

6.3.3 Flux limits and parameter scaling

The function minimisation routine cannot be used to control the trial spectrum nodes directly since any value (including negative values) may be returned. Instead, the following formula is used to convert each optimiser variable, v , to a spectrum differential flux value, f , within a limited range:

$$f = \exp \left[\ln \left(\frac{f_2}{f_1} \right) \frac{\text{atan}(v)}{\pi} + \frac{\ln(f_1 f_2)}{2} \right], \quad (6.9)$$

where f_1 and f_2 are the chosen minimum and maximum differential flux limits. This also has the effect of producing a geometric change in flux value from a linear change in optimiser variable. This is particularly effective in this application as spectrum fluxes often vary by many orders of magnitude, whereas the optimiser function proceeds generally through linear parameter variations. In the following reconstructions, limits of $f_1 = 10 \text{ cm}^{-2} \text{ s}^{-1} \text{ MeV}^{-1}$ and $f_2 = 10^{10} \text{ cm}^{-2} \text{ s}^{-1} \text{ MeV}^{-1}$ have been used.

6.3.4 Node energy sampling

Even if the measured counts equal the mean values expected from the underlying spectra (random fluctuations are negligible), a perfect reconstruction is not possible, in general, due to the limited number of nodes. As a result, the objective function minimum does not correspond to the true spectrum. Instead, the reconstruction obtained is dependent on the choice of number of nodes and node energy locations. The location of the nodes is therefore a potential source of bias and may directly influence whether a spectrum feature can be reconstructed. Multiple reconstructions (using the same input data) with randomly sampled node energies are used to avoid this bias, typically providing a spread of reconstruction estimates. Note that this is in contrast to the spread of results presented for the matrix method in Section 6.2,

where different estimates were obtained using different input data samples.

However, an entirely random selection of node locations across the particle energy ranges may lead to reconstruction attempts where all the nodes are clustered, giving poor results. Node sampling has thus been restricted to smaller ‘windows’ of the full energy range. In addition, the end nodes are fixed at the maximum and minimum energies chosen for reconstruction (1–500 MeV for protons and 0.05–6 MeV for electrons, where the minima are the approximate 50% particle detection thresholds).

6.3.5 Spectrum averaging and uncertainty

Multiple reconstructions of a single set of input data, using different node locations, produces a spread of individual spectrum estimates; an average of these constitutes a smoothed function which may be a better fit to the underlying spectrum. Spectrum averaging may be achieved through Robust Locally Weighted Regression, a method sometimes used for smoothing scatterplots [154]. In the present application, this procedure is used to create a single best estimate pair of spectra. These are also piecewise linear power law functions with the same number of nodes as the constituent spectrum estimates. However, the nodes are placed at the centres of the energy windows (not random) and each flux value (the ‘smoothed point’) is assigned through an iterative smoothing of the spectrum estimate coordinates (the ‘input coordinates’). This procedure is as follows:

1. Each input coordinate is assigned a local weight depending on its proximity in energy to each smoothed point.
2. The total weighting of each input coordinate is equal to the product of its local weight and its robust weight (the latter initially equal to the inverse of the objective function value obtained for that spectrum estimate).
3. Each smoothed point is estimated as the weighted mean of the input coordinates.
4. The robust weights are updated according to the proximity in flux of each input coordinate from the smoothed points.
5. Steps 3 and 4 are repeated, usually converging in ~ 3 –10 iterations.

The width of the local weighting window is chosen to equal the node spacing. The local and robust weighting functions are the ‘tricube’ and ‘bisquare’ respectively, given in [154].

Calculation and communication of the fitted spectrum uncertainties are also important considerations. The uncertainty in the fit may be estimated by constructing a confidence interval for the parameters: for example, this may be given by the RCF inequality [155]. Otherwise, uncertainties may be estimated (less rapidly) by a Monte Carlo study. An alternative (and currently unquantified) indication of uncertainty is the spread in the individual reconstruction estimates.

6.3.6 Example implementation

Figure 6.12 shows this new method of reconstruction applied to the spectra for reference orbit E. This example uses the same input data and number of free parameters (six nodes per species) as the reconstruction presented in Fig. 6.9, allowing a direct comparison with the matrix method.

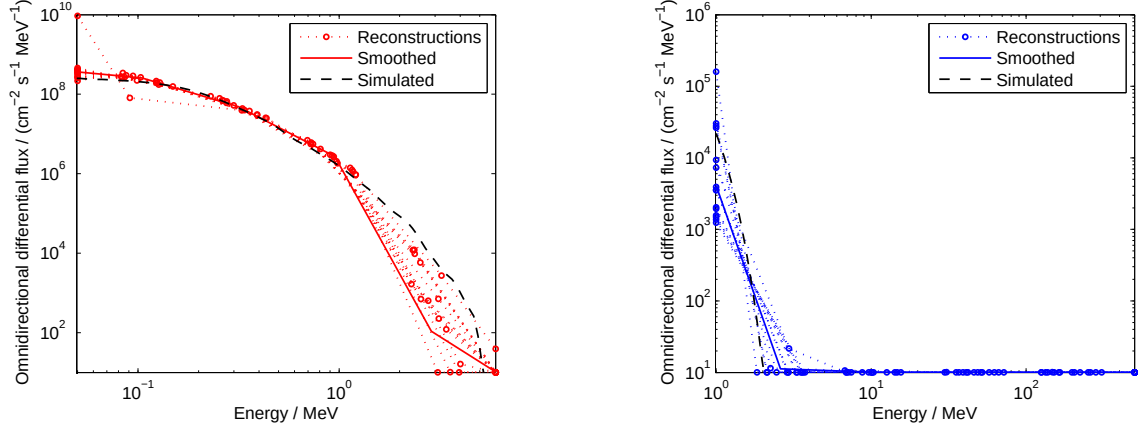


Figure 6.12: New method spectrum reconstruction result for electrons (left) and protons (right) for a 110 s simulated exposure in reference orbit E using six spectrum nodes per species.

The result of each individual spectrum estimate is shown together with the overall smoothed best estimate. An improved reconstruction is obtained, especially for protons, although it should be noted that this is a relatively short exposure (resulting in greater uncertainty at high energies where fluxes are lower). Further reference orbit results are presented in Section 6.5.

6.4 Reconstruction in the presence of sensor noise

The reconstructions presented above have used randomly sampled event data (sensor energy deposits) from Geant4 simulations of the full HMRM geometry. However, these deposit values have then been used directly in the identification algorithm, where each event is counted in one of the 32 monitor channels. The sensor response model, which includes noise and pixellation behaviour, has not been included. Both of these effects will generally increase the variance of the monitor response to any particle type, meaning that spectrum reconstruction fits will be less well constrained. To investigate this effect, in the absence of a fully calibrated sensor response model, reconstructions have been undertaken using a simple Gaussian noise model. Each Geant4 sensor energy deposit has a randomly sampled noise energy added prior to the ID table look-up and channel allocation.

Figure 6.13 shows how an increasing sensor noise affects the channel counts resulting from the same input energy deposit data. The change in allocation when a small amount of noise is first introduced is considerable. In contrast, a further doubling or tripling of the noise

level leads to smaller fractional changes. Prior to experimental measurements with the sensor prototypes, the total noise is assumed to be approximately 30 e^- per readout (clusters of nine pixels with 10 e^- noise per pixel).

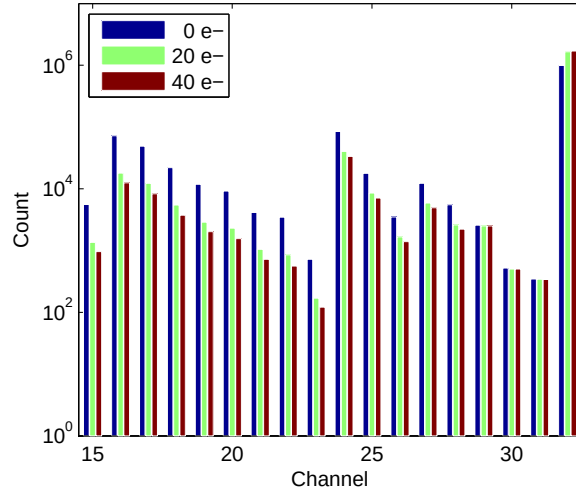


Figure 6.13: Counts allocated to channels 15–32 in response to a simulated flux of protons and electrons with E^{-1} differential energy spectra, after the addition of different levels of random Gaussian noise.

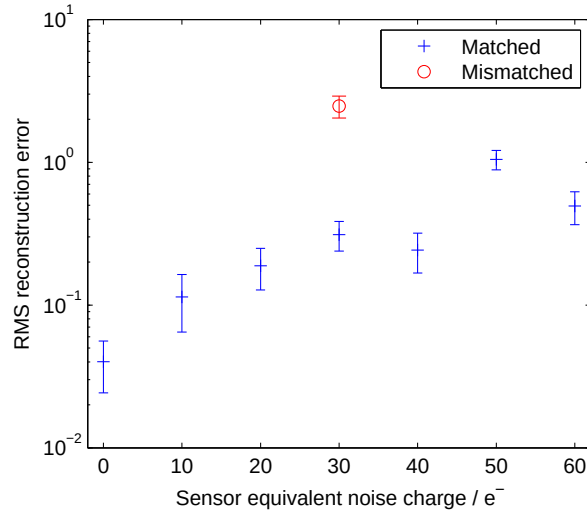


Figure 6.14: Fractional spectrum reconstruction error for E^{-1} spectra after the addition of different levels of random Gaussian noise to the event data. In the ‘matched’ case, the response functions contained the same noise as added to the event data; also shown is an example of the ‘mismatched’ case, where the response functions contained 10 e^- noise compared to 30 e^- in the event data.

A second feature of Fig. 6.13 is the pattern of count redistribution in the presence of noise. Independent of the spectrum, the channel 32 count is always significantly increased by an

increase in sensor noise. This is because the other 31 channels cover only a small portion of the possible sensor energy deposit combinations, while channel 32 counts all remaining ‘unidentified’ events. Fluctuations are therefore far more likely to result in transfer of counts from channels 1–31 to channel 32 than the reverse.

Noise must also be added when formulating the response functions. It is initially assumed that the noise level included in the response functions is exactly equal to that included in the event data sample. Figure 6.14 shows the RMS reconstruction error for E^{-1} spectra when assuming different levels of sensor noise, demonstrating that increased noise produces a worse reconstruction. Also shown on this plot is the case where the response functions assumed a noise of 10 e⁻ while a noise of 30 e⁻ was applied to the data. This results in a worse fit than if the sensor noise were twice as large but correctly described in the response functions. This illustrates the importance of accurate device modelling for successful spectrum reconstruction.

6.5 Preliminary reference orbit results

Example reconstructions are provided in this section and in Appendix F for the reference orbits identified in Chapter 1 as being of particular interest for HMRM operation. In each case, the reconstruction input data were obtained from a Geant4 simulation of the monitor in proton and electron fluxes corresponding to the AP-8 and AE-8 model mean omnidirectional energy spectra for the orbit. Each simulated sensor hit event was identified via the monitor ID table look-up procedure and allocated to one of the 32 channels. This single sample of 32 counts was used as the input to ten reconstructions (each with different random node positions), producing multiple spectrum estimates. These were averaged using the procedure described in Section 6.3.5 to produce best estimate spectra for each orbit.

Although the full sensor response model has not been applied to the data at this preliminary stage, a random noise of 110 eV (30 e⁻, as introduced earlier) was added to each sensor energy deposit prior to the table look-up; this noise is also included in the monitor response function.

Note that the simulated exposure times used for these examples are short (of order 100 s) and are approximately equivalent to the integration period of one HMRM extended data telemetry packet. In normal operation, reconstruction may be performed using accumulated data from several consecutive packets (or even multiple orbital passes), reducing the statistical uncertainty. Furthermore, the assumption of *mean* orbital fluxes provides an especially harsh test of the reconstruction ability since peak fluxes are at least an order of magnitude greater than these mean values. This is illustrated by Fig. 6.15 which compares the mean spectra for orbit A with those occurring at the point of peak integral proton flux in the SAA: both spectra are significantly increased in amplitude. Figure 6.16 shows that these high fluxes are expected to persist for several minutes as the spacecraft passes through this region of the orbit, providing sufficient time for data acquisition.

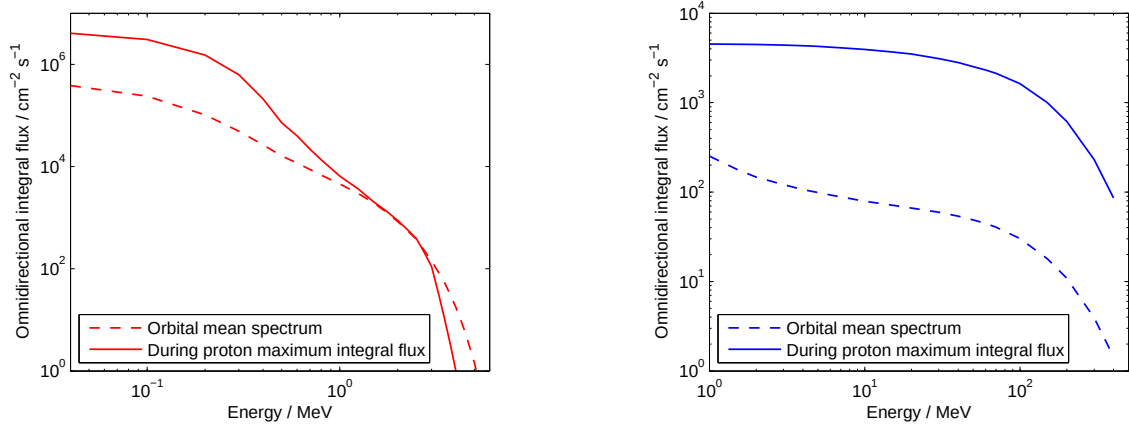


Figure 6.15: Comparison of the orbit A mean integral energy spectra for electrons (left) and protons (right) with the spectra obtained at the point of maximum proton integral flux (occurring at the SAA).

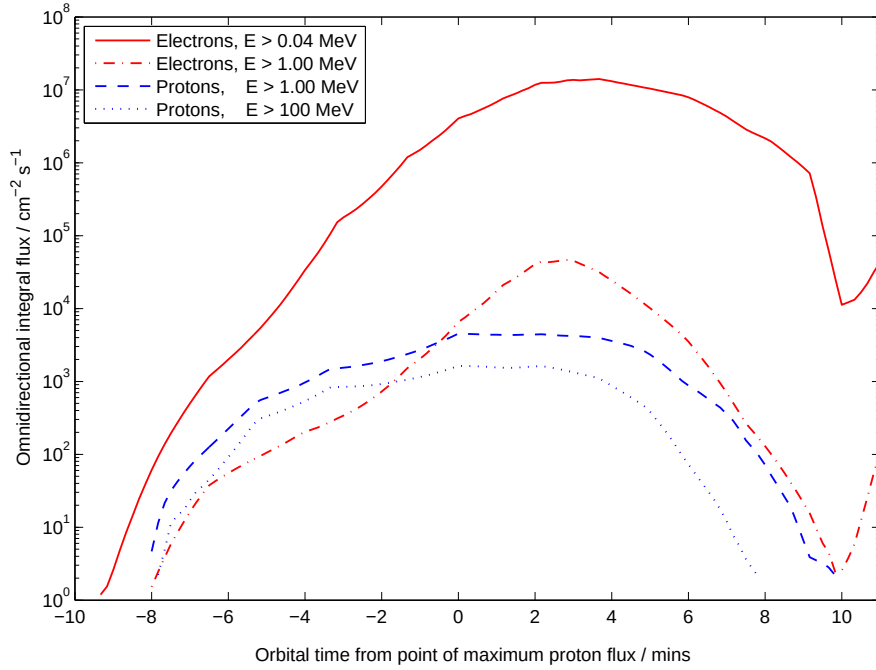


Figure 6.16: Total integral fluxes of electrons and protons above selected energies as a function of time in reference orbit A.

Figures 6.17 and 6.18 show the reconstructions for orbit A: the electron result is good at low energies, while protons and high energy electrons suffer from low count rates. Peak fluxes (greater by a factor of ten) would perhaps be a fairer test for this orbit, as proton fluxes are negligible outside of the SAA. In contrast, Figs. 6.19 and 6.20, the reconstructions for orbit D, show an excellent result despite an exposure time one quarter of that used for orbit A. This improvement is due to the increased fluence (by a factor of ~ 100) provided by the harsher environment of this orbit. The reconstructions presented in Appendix F for the remaining reference orbits show results between these two extremes, giving generally very good performance.

6.5.1 Orbit A: 700 km, 98° inclination, sun synchronous orbit

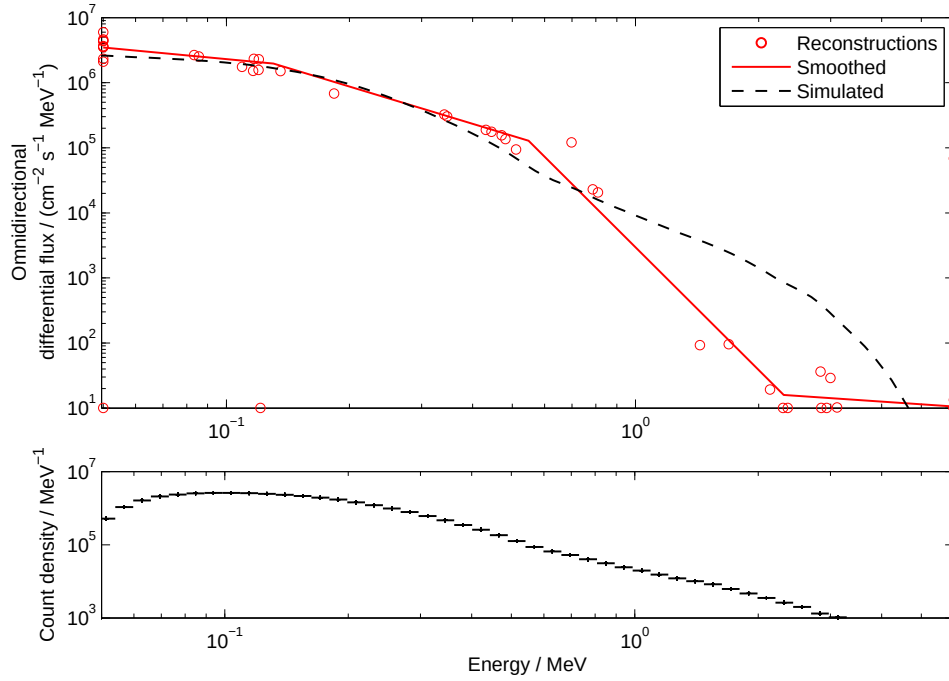


Figure 6.17: Electron spectrum reconstruction for a 950s simulated exposure in reference orbit A with five spectrum nodes per species.

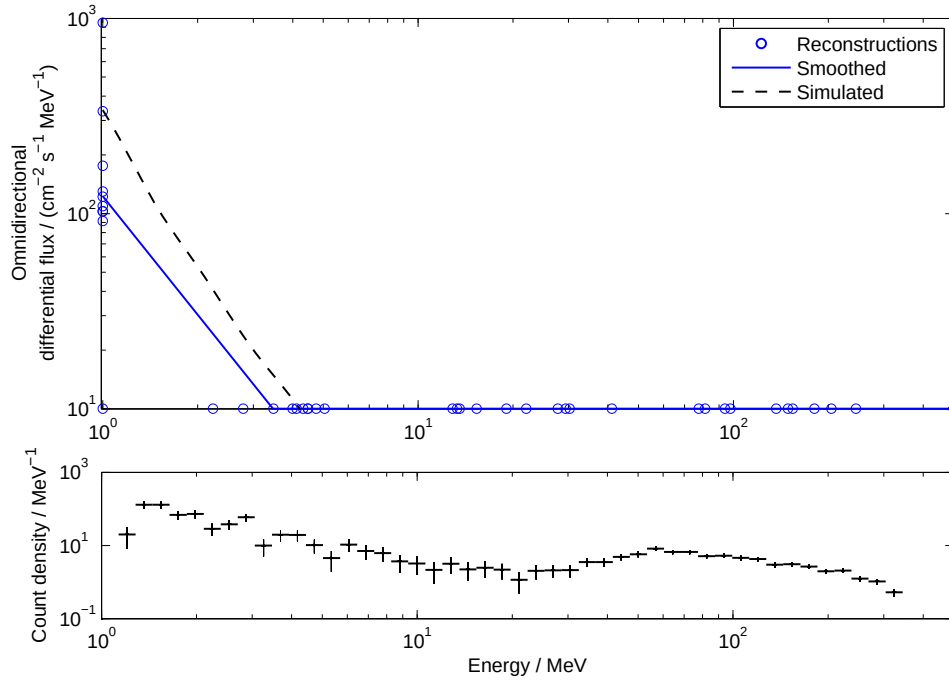


Figure 6.18: Proton spectrum reconstruction for a 950 s simulated exposure in reference orbit A with five spectrum nodes per species.

6.5.2 Orbit D: 23,222 km, 56° inclination, medium Earth orbit

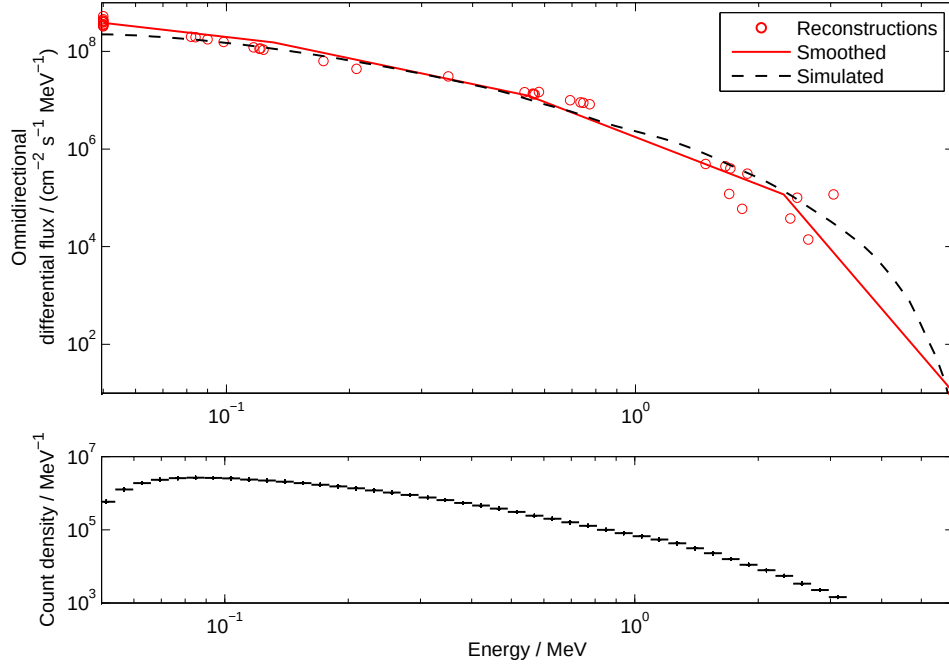


Figure 6.19: Electron spectrum reconstruction for a 225 s simulated exposure in reference orbit D with five spectrum nodes per species.

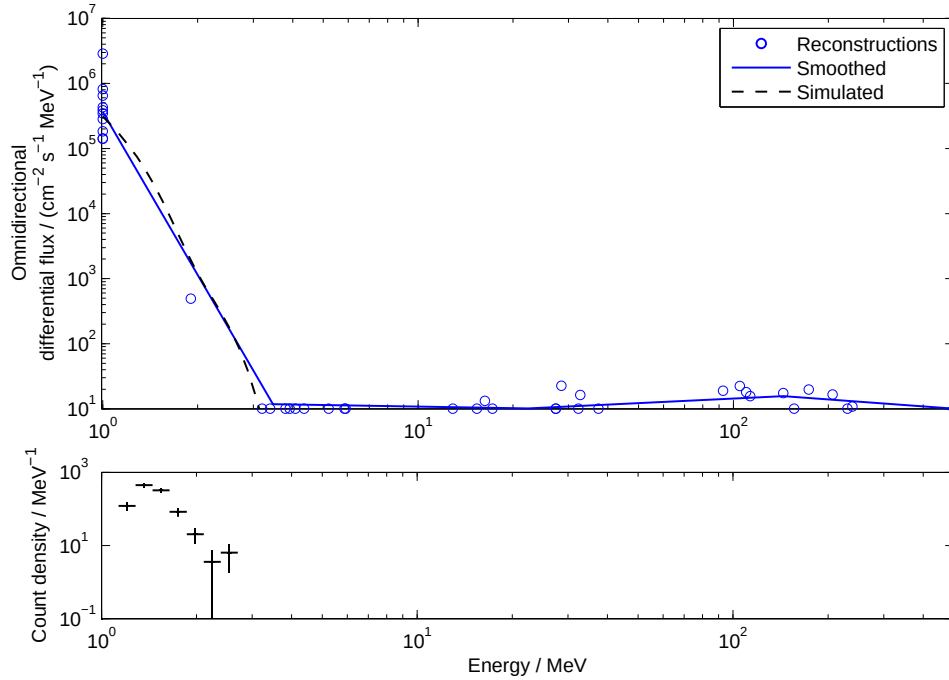


Figure 6.20: Proton spectrum reconstruction for a 225 s simulated exposure in reference orbit D with five spectrum nodes per species.

6.6 Future development

Several aspects of the proposed new method require further development. These include the quantification of spectrum uncertainty estimates and testing of alternative optimisation routines. The appropriateness of different numbers of spectrum nodes, especially in relation to the quantity of data available, should also be addressed.

6.6.1 Method extensions

In addition to development of the reconstruction method (including increasing the efficiency and effectiveness of the optimisation routine), aspects of the response and simulation model may be improved to enhance realism. The full sensor response model (describing the pixel charge sharing and limited digitisation resolution), which was neglected in this chapter, should be applied to the input data. Furthermore, output data from the orbit simulator (described in Chapter 4) may be used to assess spectrum reconstruction in a time-dependent radiation environment. Finally, reassessment of the assumptions stated in Section 6.1.1 should be considered:

1. **Response linearity.** A simplified model of pile-up may be introduced by assuming that these events are always assigned to channel 32 (non-specific event counting), resulting in the loss of two counts from the other channels. The channel count rate estimates produced from the trial spectra may be used to estimate the average pile-up rate and thus move a mean fraction of counts accordingly.
2. **Particle species.** A third particle (e.g. helium ions) may be introduced, usually in smaller fluences. This would allow an assessment of the effect on the electron and proton reconstructions, as well as setting of a sensitivity limit on detection of these small fluences. This may be a smaller consideration than pile-up, which is expected to reach $\sim 10\%$ in some orbits.
3. **Flux isotropy.** The presence of the Earth's magnetic field will generally create anisotropy in orbit. This may be addressed by using a parameterised anisotropy model, such as the Badhwar-Konradi [156, 11]. Extra information (the spacecraft coordinates and attitude, or magnetic field measurements) may be used to reduce the number of additional free parameters required in the fit.

6.6.2 Response function validation

Monitor response functions should be validated experimentally, where possible, as they are relied upon by all spectrum reconstruction methods. This requires the monitor to be exposed to a particle source of accurately known energy spectrum, ideally approximately monoenergetic and of single species, such as a particle beam. Predicted channel counts may then be compared directly with those measured.

The response function accuracy is dependent on the Geant4 geometry model and the sensor response model (the Geant4 physics list having been validated satisfactorily in Chapter 4).

Improvements in these two models may be required to ensure optimal description of the constructed device, allowing a realistic end-to-end test of the spectrum reconstruction process.

6.7 Summary

A reconstruction of the incident particle energy spectra is an attractive goal for radiation monitoring. This information allows any radiation effect to be deduced through later simulations.

Given accurate monitor response functions, the forward problem of predicting channel counts in an environment with known spectra is a simple calculation. This can be used to solve the more difficult inverse problem by iteratively varying a set of trial spectra and comparing the predicted counts with monitor data. The simple ‘matrix method’ provides a well constrained model of a piecewise constant spectrum which allows fast solution. This model does not, however, allow a good fit to real spectra or monitor response functions, which are generally detailed, continuous functions of energy.

A new spectrum model of a piecewise continuous power law has been proposed to provide a better fit for the same number of free parameters. Random energy node selection reduces the implicit bias created when setting up the optimisation problem.

Spectrum reconstruction accuracy is found to increase with larger input data samples, and reduce with greater sensor noise. Performance is greatly worsened, however, if the response functions do not match the true sensor behaviour.

Preliminary results show that satisfactory reconstructions can be obtained with exposures of order 100 s in a varied selection of important Earth orbits. This is despite a large imbalance in the total flux of electrons and protons.

Chapter 7

Preliminary testing of the HMRM active pixel sensors

This chapter presents the results of initial testing of the sensor ASICs following the assembly of the first HMRM prototype in December 2012. The limited time available for this testing was due to the targeting of a flight opportunity in 2013 and this work was carried out in parallel with ASIC probe-testing undertaken by a commercial electronics company.

A pixel signal model is proposed and used to guide the calibration process, allowing the isolation and measurement of each model parameter. While this work does not constitute a full calibration of the ASIC, comparator offsets, pixel pedestals and pixel noise have been measured in relative units with preliminary conversion to equivalent noise charge. These values have also been used to select comparator trim settings and pixel masks for use in monitor operation.

7.1 The HMRM ASIC

Each HMRM ASIC incorporates a 50×50 APS array with a pixel pitch of $20 \mu\text{m}$ (sensitive area of 1 mm^2). The pixel circuitry follows a four transistor design, allowing reset noise reduction via the correlated double sampling method [157]. The pixels are expected to show approximately 10 electrons RMS equivalent noise charge and a full-well saturation of 19,000 electrons. A photograph of an HMRM ASIC, wirebonded to the monitor PCB, is shown in Fig. 7.1. Further details were provided in Chapter 3, while the high-level ASIC design diagram is presented in Appendix B.

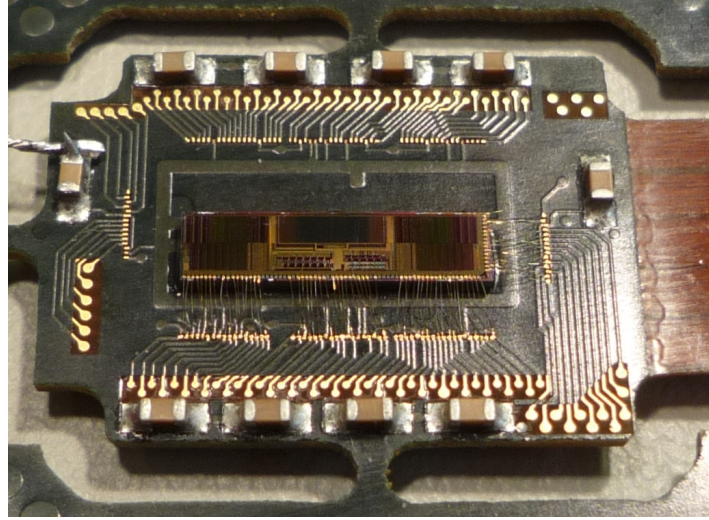


Figure 7.1: Photograph of an HMRM ASIC mounted on the monitor front PCB section. The 250 μm thick ASIC die measures 10.3 mm by 2.4 mm.

7.1.1 Pixel design

Active pixel sensors are distinguished from the earlier passive pixel sensors (PPS) in their inclusion of a buffer or amplifier within each pixel, resulting in improved noise performance [143]. The simplest APS design uses three transistors per pixel ('3T'), following the general architecture shown in Fig. 7.2. Here, the sense node D1 is a photodiode or photogate. A reset pulse applied to the gate of T1 reverse-biases the diode; after removal of the reset signal, a charge remains due to the diode capacitance. After an integration period, when ionisation charge is collected by the diode, switch T3 is closed by application of the row select signal. This allows the diode voltage to be read through the source follower T2, providing a low output impedance. The source follower load (current sink I) and column select transistor Tc are external to the pixel (and may be shared by all pixels in a column). Thus, in the 3T design, the diode is both the charge collecting element and the readout node.

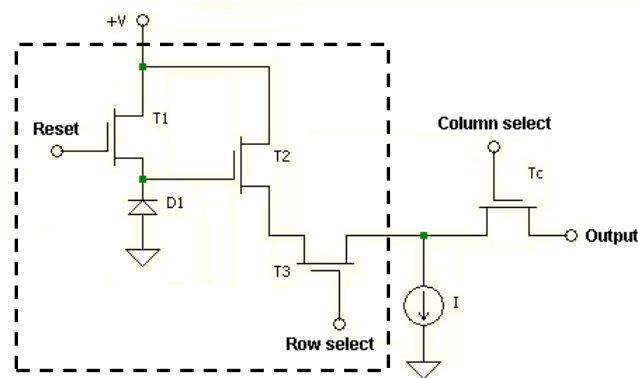


Figure 7.2: Simplified circuit diagram of a three transistor ('3T') pixel [143]. The dashed box indicates the components that exist within each pixel, including the three transistors (T1, T2 and T3) and the photodiode (D1). The current sink (I) and column select transistor (Tc) may be shared by each column of pixels.

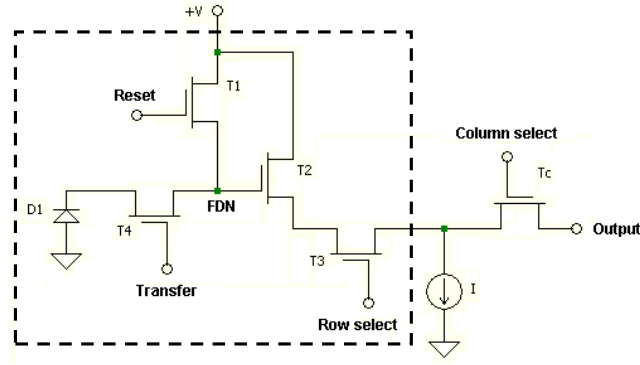


Figure 7.3: Simplified circuit diagram of a four transistor ('4T') pixel [82]. The transfer transistor (T4) isolates the floating diffusion node (FDN) from the pinned photodiode (D1). The dashed box indicates the components that exist within each pixel.

A disadvantage of the 3T design is that, after reset, an unknown, thermally generated noise voltage is left on the diode. To reduce this noise, many modern CMOS APS designs use a pixel containing four transistors ('4T'), as shown in Fig. 7.3. This design typically uses a pinned photodiode as the charge collector which, when held at the pinning voltage, becomes fully depleted [82]. During integration, the additional transistor, T4, isolates the diode. Meanwhile, the region between T4 and T2, known as the floating diffusion node (FDN), is also isolated. This has capacitance to ground and is charged, via T1, to a voltage higher than the pinning voltage during the reset process. To read the pixel, the FDN voltage is first sampled through T2 and T3 as before. Switch T4 is then closed, allowing charge transfer between diode and FDN. A second reading is then made of the FDN voltage. These two measurements represent the reset voltage and the exposed voltage, both of which contain the same reset noise. These may thus be subtracted (either in further pixel circuitry, or externally) to remove this noise, a technique known as correlated double sampling. The lower capacitance, C , of the floating diffusion node also results in lower thermal noise (proportional to $\sqrt{TC}$, with T the absolute temperature) and higher conversion gain.

7.1.2 Pixel readout and digitisation

The HMRM pixels are read out simultaneously ('snapshot mode') and digitised in a 3-bit analogue-to-digital conversion process involving seven comparator levels. Each comparator level is programmed as a 7-bit threshold setting, allowing customisable, non-linear pixel digitisation schemes. These levels are converted from digital to an analogue voltage on chip.

The comparator level voltage step size increases with level number, L1 to L7, giving progressively larger ranges with lower resolution. Each level voltage is applied to the comparator in sequence, giving a series of seven output bits. These are converted to the final 3-bit pixel value by a digital encoder, according to the truth table shown in Table 7.1. The level voltages must be programmed to be monotonically increasing with level number to ensure that one of the standard sequences shown in Table 7.1 results. Sequences not listed in this table generate an undefined 3-bit output value.

Table 7.1: Truth table for creation of the pixel values by digital encoder. The result of comparison of the analogue pixel value and each threshold level L1–7 is shown as a 1 (pixel signal exceeds threshold) or 0 (signal below threshold).

Signal above threshold?							Output code	Value
L1	L2	L3	L4	L5	L6	L7		
0	0	0	0	0	0	0	000	0
1	0	0	0	0	0	0	001	1
1	1	0	0	0	0	0	010	2
1	1	1	0	0	0	0	011	3
1	1	1	1	0	0	0	100	4
1	1	1	1	1	0	0	101	5
1	1	1	1	1	1	0	110	6
1	1	1	1	1	1	1	111	7

One comparator is provided for every half-column of 25 pixels; each comparator has an unknown input offset error, due to unavoidable manufacturing variations, which is assumed to be constant for any given device. This may be compensated by the application of a trim voltage programmed as an 8-bit trim value. An extra (51st) column of null pixels was included to allow convenient readout over the 9-bit bus. This means that there are a total of 102 comparators and 2550 pixels per chip.

In addition to the threshold and trim values, four programmable bias currents are generated on chip, each of which requires a 6-bit value (default values were provided by the probe-testing). These bias currents have the following functions:

1. **biasCOL** loads the output source follower of the pixels.
2. **biasSF** loads an additional source follower in the pixel analogue readout chain.
3. **biasTH** biases the digital to analogue conversion for the comparator threshold generation.
4. **biasCOMP** biases the digital to analogue conversion for the comparator trim generation.

The trim, threshold and bias values constitute a full set of programmable parameters for the ASIC. These values are loaded serially into two shift registers within the ASIC (one register accepts the trims, the other accepts both the threshold and bias settings) by the HMRM FPGA.

7.1.3 HMRM ASIC version 1

The first ASIC production run for the HMRM was completed in 2011. However, initial tests indicated that the bias/threshold shift register was not fully functioning, preventing values from being loaded reliably. This meant that this initial device version could not be used; after modification of the design, a second version was produced in 2012. This second ASIC version is the subject of the characterisation work presented here.

7.2 Sensor noise and signal model

The calibration of any sensor involves the determination of the relationship between the output signal and the physical quantity being measured. In this instance, from an electronics perspective, the desired quantity is the charge collected by each pixel diode, whereas the output signal is a voltage (which is further abstracted by digitisation, as previously discussed). Without calibration, this output voltage or digital value is physically meaningless as a measure of radiation interactions.

As well as finding the transfer function relating charge input and signal output, the calibration should quantify the variability of the output signal experienced with a constant input. This variation is generally referred to as *noise* and, for the HMRM, it is modelled as having three main components: fixed pattern noise, dark noise, and interaction noise. Note that the word ‘signal’ is used here to refer to any analogue value from the pixel and is not used to exclude the unwanted ‘noise’ component. Thus, in the absence of radiation interactions, the pixel signal may consist solely of noise.

7.2.1 Fixed pattern noise

Fixed pattern noise (FPN) results from the non-uniformity of a pixel array. It is only apparent if the pixels are treated as a set of identical sensors, rather than a collection of separate sensors which must be calibrated individually. This is usually a practical necessity, since a separate conversion of each pixel signal into a calibrated measurement is not generally feasible for arrays of thousands of pixels.

Due to manufacturing variability, location within the pixel array, and other effects, each pixel will have a different gain and mean dark signal (the latter here referred to as *pedestal*). Rather than measure these quantities for each pixel and use them in normal operation, the array may be characterised by the overall mean (best estimate) and standard deviation (the noise). Thus the distributions of the gain and pedestal across all pixels are interpreted as probability distributions when applied to a single pixel selected at random. However, this is merely a simplification and the gain and pedestal non-uniformity can in principle be corrected via flatfield and offset corrections, respectively.

Pedestal correction is partially achieved through the correlated double sampling (CDS) method, whereby the voltage immediately after reset is subtracted from the final integrated pixel signal. This process also allows reduction of the thermal (random) noise added in the reset process as well as the constant offset [158]. Pixel offset may also be introduced after the CDS application and thus a unique pedestal is still expected for each pixel. Since the HMRM calibration will treat each pixel independently, these pedestals may be measured independently of the random noise. However, during radiation monitoring, any residual pedestal remaining after trimming will appear as a noise due to the necessary assumption of pixel indistinguishability.

The mean measured signal, m_i for each individual pixel may therefore be expressed as:

$$m_i = g_i s_i + d_i, \quad (7.1)$$

where s_i is the mean charge collected by the pixel from the radiation interactions. The gain and pedestal are represented by g_i and d_i , respectively.

In addition to the constant offset (pedestal) associated with each individual pixel, an offset is also expected to arise at each half-column comparator. This results randomly from process variations (e.g. doping concentration and layer thickness) and also systematically due to location of the half-column in the array. Readout lines from some columns are longer, due to the circuit layout, causing differences in capacitance and voltage drop.

The comparator offset will combine with the pixel pedestal to affect the measured pixel signal. In absolute terms, these two offsets are inseparable when measurements are made from the comparator output. However, the comparator offset will be the same for each pixel in a half-column and so d_i may be reinterpreted as the *effective* pixel pedestal, defined such that:

$$\frac{1}{25} \sum_{i=1}^{25} d_i = 0, \quad (7.2)$$

where the mean is taken over the 25 pixels of a comparator group (a half-column). Any common component of the true pedestal is combined into the comparator offset, c_j , giving:

$$m_i = g_i s_i + c_j + d_i, \quad (7.3)$$

where the j^{th} comparator digitises pixel i and 24 others.

7.2.2 Comparator trim

The offset of each comparator may be compensated by the application of an 8-bit trim setting, p_j . This integer is assumed to be linearly converted to a voltage with a trim gain a_j of unknown value. Although the value of a_j may differ between comparators, this does not appear as a noise since it can be taken into account when choosing the operational trim settings. The full model for the mean pixel signal thus becomes:

$$m_i = g_i s_i + c_j + a_j p_j + d_i. \quad (7.4)$$

Trim selection corresponds to choosing p_j such that:

$$c_j + a_j p_j = 0. \quad (7.5)$$

However, this cannot be achieved exactly because p_j has limited resolution, being an 8-bit integer. This means that there will in general be a small digitisation error. In addition, there exist the untrimmable pixel pedestals, d_i (with a mean of zero). These two quantities form a residual pedestal for each pixel which will remain after trimming — in this case:

$$m_i = g_i s_i + \delta_i, \quad (7.6)$$

where δ_i is the residual pixel pedestal. Under the assumption of indistinguishable pixels in radiation monitoring mode, the spread of these values will appear as a noise.

7.2.3 Interaction noise

While FPN presents the same spatial pattern across the array on each readout, many noise sources are random so that a different value is obtained for a given pixel on each readout. Random noise components may be divided into those that depend on particle interactions (*interaction noise*) and those that are independent of these interactions (collectively referred to as *dark noise*).

An example of interaction noise, signal shot noise is the fluctuation in particle arrivals which may be described by the standard deviation of the Poisson distribution. Additionally, variation in the energy deposited by identical particles will be observed (see the discussion of energy loss fluctuations in Chapter 2). Sharing of a radiation event energy deposit between neighbouring pixels occurs due to charge diffusion. The resulting spread of pixel values from a single event constitutes a noise source; this charge sharing behaviour was discussed in detail in Chapter 5. Finally, the conversion of deposited energy to charge carriers within the semiconductor implies a Fano noise: this was also introduced in Chapter 2.

7.2.4 Dark noise

Dark noise is defined as all random noise components that do not depend on the magnitude of the charge collected from a radiation interaction. This noise is therefore present in every readout pixel signal. Important components of the dark noise originate inside the pixel; amongst others, these include:

1. **Reset noise.** Excluding FPN, this is the dominant noise source in CMOS monolithic active pixel sensors [143]. It may, however, be greatly reduced through CDS. The equivalent noise charge in electrons is related to the sense node capacitance, C_N , and temperature, T , by:

$$\sigma_R = \left(\frac{kTC_N}{e} \right)^{\frac{1}{2}}, \quad (7.7)$$

with e the electron charge [159]. This noise can therefore be reduced by decreasing the diode capacitance. Note, however, that this may make the pixel more susceptible to other effects, such as random telegraph signal noise [160].

2. **Thermal white noise in the channel of the source-follower MOSFET.** This is also dependent on the diode capacitance [42]. The exact functional form depends on the readout circuit implementation.
3. **Leakage current shot noise.** As well as producing a fixed pattern noise, the leakage current I_l causes a shot noise, σ_s , which depends on t , the integration time:

$$\sigma_s = \sqrt{I_l t}. \quad (7.8)$$

4. **Random telegraph signal noise.** This is caused by trapping and release of charge carriers at a defect site and is observed as discrete transitions in the output signal [158]; it is commonly associated with the in-pixel source follower transistor.

HMRM dark calibration will be carried out using a fixed integration time and at approximately constant room temperature. The overall dark noise is thus assumed to be constant for each pixel and is modelled as having a Gaussian distribution.

Following from Equation 7.4 (the model for the mean pixel signal), the overall pixel noise is modelled as:

$$\sigma_{m_i}^2 = g_i^2 \sigma_{s_i}^2 + \sigma_i^2, \quad (7.9)$$

where σ_i is the dark noise and σ_{s_i} is the interaction noise. When in the dark, $s_i = \sigma_{s_i} = 0$ is assumed and hence $\sigma_{m_i} = \sigma_i$. Note that measured noise resulting from threshold variation is often also included in noise models. This is not added here since the same levels are shared by all comparators. Any threshold variation must therefore occur at the comparator and so the mean component is assumed to be included in the comparator offset, while fluctuations are included in the pixel dark noise. Note that Equation 7.9 considers each pixel individually and thus excludes fixed pattern noise, as explained in Section 7.2.1.

7.2.5 Calibration model

The noise and mean signal models (Equations 7.4 and 7.9) have been used to guide the initial testing of the HMRM sensors. This involves measuring d_i and σ_i for every pixel and c_j and a_j for every comparator. The data obtained are used to select suitable trims, p_j , decide whether to mask pixels (prevent their use in radiation measurement) and characterise the sensor by quantifying the overall effective noise. Since this calibration is done using dark readouts, all interaction noise sources are zero and the pixel signal gains, g_i , are not observed. Note that cosmic rays and local radioactivity background constitute a very minor source of dark counts in such a small sensor.

A relative unit of measurement, the ‘digital number’ (DN) is used for all quantities in the model equations, with the exception of p_j (an integer in the range 0–255) and g_i (a dimensionless number). The DN unit is defined to be equal to the step size of the lowest threshold level, L1. This is a convenient choice because a direct measurement of a pixel signal may only be made relative to the thresholds.

The label ‘Lx’ is used to denote the x^{th} lowest comparator level: the value of this is set by specifying a threshold integer value, n_x , in the range 0–127. The magnitude of the level, in DN, is given by:

$$Lx = n_x h_x, \quad (7.10)$$

where h_x is the *inter-level gain*, the ratio of the steps of the comparator levels, defined such that $h_1 = h_2 = 1$. Nominal values for the inter-level gains are available from the ASIC design and are shown in Table 7.2. Note that levels L1 and L2 have equal step size and are thus both variable in steps of 1 DN.

The DN scale does not relate directly to the eight possible pixel values. Instead, these values are a consequence of a particular choice of the seven threshold settings. In general, an operational choice of thresholds may be represented as shown in Fig. 7.4.

Table 7.2: Nominal inter-level gains for the HMRM ASIC comparator levels.

Level	Nominal inter-level gain, h_x
L1	1.0
L2	1.0
L3	2.0
L4	4.0
L5	7.1
L6	14.0
L7	27.9

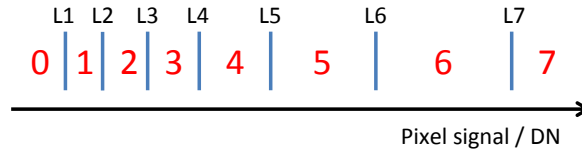


Figure 7.4: Conceptual diagram of the relation of threshold levels and digitised pixel values with respect to pixel signals of increasing magnitude. The levels are represented by vertical bars, while the pixel output values correspond to the signal ranges between the levels.

In normal operation, the voltages of each level are arranged to increase monotonically with level number. The resulting pixel values are indicated on the diagram; this shows how the levels act like a set of histogram bins. The levels (bin edges) may be approximately linearly or geometrically spaced, or have an entirely customised arrangement. For any particular threshold set, the same pixel value will result from a range of analogue signal voltages.

In contrast, the DN scale is an analogue measurement scale, assumed proportional to voltage, with which continuous quantities may be measured. However, these quantities must be measured using the threshold levels, which can only occupy discrete values. Thus a quantity which is assumed to be continuous, such as a pixel pedestal, must be estimated from a distribution of discrete measurements. This is achieved using the threshold scan method.

7.2.6 Threshold scans

Each pixel signal distribution may be deduced over many readouts by scanning a comparator level across all threshold values, yielding the integral of the probability distribution. In general, this will have the form of an ‘S-curve’ (sigmoid) as the probability of the pixel signal being below threshold changes from zero for very low thresholds to one at very high thresholds. Without radiation, if the assumption of Gaussian distributed dark signals is correct, this curve will be well approximated by the Error function, defined as:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt, \quad (7.11)$$

which is equal to twice the integral of the Gaussian distribution with zero mean and variance of 0.5. Continuous quantities, such as the mean and standard deviation (pedestal and noise), may be extracted by fitting an Error function to the threshold scan data.

7.3 Calibration overview

The calibration procedure must be completed for each individual ASIC due to unavoidable variation in the manufacturing process. The full calibration [161] includes several parts:

1. **Inter-level gain measurement** — These should be measured directly and the nominal values in Table 7.2 verified or amended as necessary.
2. **Offset measurement** — This is done for each individual pixel and is defined as the mean pixel signal in the absence of radiation. The comparator offset is obtained by averaging the pixel offset over the 25 pixels of each comparator group. The deviation of each individual pixel offset from this comparator mean is an untrimmable pedestal and will introduce dispersion into any measured signal.
3. **Trim gain measurement** — This gives the size of the trim step relative to the comparator threshold step.
4. **Trim selection** — Once the offsets and trim gain are known, it is a trivial matter to choose the optimal trim setting for each comparator such that: $c_j + ap_j = 0$. In general this cannot be achieved exactly because the trims, p_j , are integers and there will therefore be a small digitisation error.
5. **Pixel noise measurement** — In the absence of radiation this is the previously defined dark noise. The variance of a detected signal is assumed to contain this noise, plus additional components due to fluctuations in the charge collection process and in the signal itself. If the detected signal is derived from multiple pixel signals, the overall variance will increase with the number of pixels involved and also depend on their non-uniformity in pedestal and gain.
6. **Pixel mask selection** — Pixels that behave anomalously (showing a significantly large noise or offset) may be selected for masking. Their signal will then be ignored by the FPGA when in radiation monitoring mode. However, this should be done sparingly since it may result in incomplete charge collection from a radiation event. Masking of pixels at the boundary of the array is less problematic since these already face this problem.
7. **Trim and mask verification** — After selection and testing in software, the trim and mask may be applied in hardware (the trims loaded into the ASIC itself, and the mask in the FPGA) and their effect confirmed directly.
8. **System gain measurement** — The above measurements (noise, offset, trim gain) are made directly in DN units. At this stage, signals measured with the device are also in these relative units. Absolute calibration of the device, where the value of 1 DN is determined in a physical quantity such as eV, may be achieved by measuring a signal of

known energy per event. This will be attempted using ^{55}Fe which emits X-rays of known energy: when an X-ray photon is absorbed close to a pixel diode, the full photon energy may be collected by a single pixel. This produces a recognisable full-energy photopeak in the pixel spectrum (measured in DN) allowing determination of the overall system gain.

9. **Comparator level threshold selection** — Once the steps listed above are complete, the settings to be used during radiation monitoring may be chosen. These include the comparator level threshold settings (which are loaded into the ASIC) as well as software quantities which are used in the FPGA physics algorithm (lookup tables and alert thresholds).

Note that this procedure does not include the calibration of the HMRM as a radiation monitoring instrument and instead only considers the individual ASIC component. Furthermore, the data presented in this chapter do not cover all of the points listed above, as the sensor calibration is ongoing. Full monitor calibration, after component-level testing of the ASICs, will involve additional particle beam and radioactive source tests.

7.4 Data acquisition system

Once integrated into the HMRM, the detector ASICs must be read out using the monitor FPGA and the data acquired through the CAN telemetry interface. Rather than the usual radiation monitoring mode (in which pixel data are summed), characterisation of the ASICs must be done in transparent mode, where all 2550 pixel values from each installed sensor are transmitted by the FPGA on the CAN bus. The bandwidth of the CAN bus limits the data rate to approximately 4 frames per second (compared to a readout rate of 7.8 kHz). The transparent frame data have been acquired and analysed using a preliminary data acquisition system developed for this initial testing phase of the project. This uses real-time processing with a C++ program running on a personal computer.

The bias, threshold and trim settings may be edited individually and uploaded by the user. Single readout frames may be viewed from a selected sensor; for maximum speed, the pixel values are displayed as a grid of numbers on screen (an example is shown in Fig. 7.5). Also displayed are the chosen trim, comparator level and bias settings. Multiple frames may also be readout and the average displayed (each pixel mean is rounded to the nearest integer).

7.5 Initial tests

Before attempting the calibration, preliminary tests were carried out to check that the basic ASIC functions were operating as expected and to ensure that communication between the ASIC and the FPGA was correct. These included testing the two shift registers used to input the trim, bias and threshold settings into the ASICs, and also obtaining pixel values from the device.

7.5.1 Shift register tests

Test patterns of known bit sequences were clocked into the two ASIC shift registers using the FPGA. By programming the registers twice, the ASIC is made to shift the initially programmed data back out on a test pin. This allows the ingoing and outgoing data to be observed with an oscilloscope and compared, demonstrating that the shift registers are functioning on a simple level. Note, however, that this method does not determine whether the ASIC interprets the data correctly.

The comparator trim register is 102 bytes long; a series of test values were loaded into the FPGA via CAN telecommand and from there shifted into the ASIC twice. The input and output data from the shift register were captured on a digital storage oscilloscope, a small portion of which is shown in Fig. 7.6. The two bit sequences were extracted using a MATLAB script and were found to match with no bit errors.

This process was repeated for the 73-bit bias/threshold register, again demonstrating successful operation. These tests confirm that the shift register malfunction discovered in the first ASIC version has been rectified.

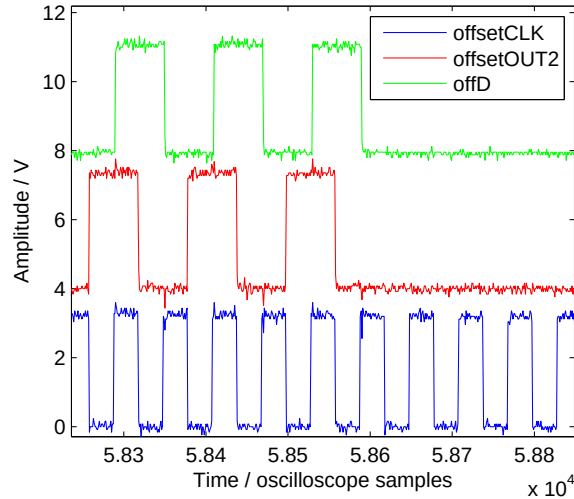


Figure 7.6: Oscilloscope data capture showing the FPGA-generated clock signal and input data (offsetCLK and offD), compared with the data shifted out of the ASIC (offsetOUT2) for the trim register. The oscilloscope sample period was 0.1 μ s.

7.5.2 Monitor power consumption

The power supply current (at 5 V) for the HMRM with two installed ASICs was measured in the three different operational modes and used to calculate the total power consumption. This information is shown in Table 7.3, compared to the consumption with no ASICs installed (thus comprising just the FPGA and support electronics). An internal linear voltage regulator produces a 3.3 V supply for the ASICs and other electronics (no 3.3 V supply is available directly on the spacecraft), constituting $\sim 34\%$ of the power dissipation. This information has been used to calculate the approximate power consumption per ASIC. Note that standby mode is not intended as a low power mode and is instead used to halt monitor operation while telecommands are issued and settings programmed.

The large current drawn with two ASICs installed (D2 and D3) was deemed to prohibit the addition of the third ASIC (D1) to this prototype. Although the relatively large power consumption does not prevent operation of the monitor, thermal regulation (especially in the space environment) must be addressed.

Table 7.3: Overall power consumed by the HMRM from a 5 V supply, measured before and after the two ASICs were installed. Measurement uncertainties were 5 mW. These data have been used to calculate the mean power consumed per ASIC.

HMRM mode	Total power without ASICs installed / mW	Total power with ASICs D2 and D3 installed / mW	Mean power per ASIC / mW
Standby	375	1125	248
Radiation monitoring	390	1230	277
Transparent	390	1300	300

In addition to these normal operating values, it was discovered that an abnormally large current of over 300 mA (>1.5 W total power) was occasionally drawn. This was found to occur apparently randomly, with $\sim 3.5\%$ probability, when applying power to the monitor after a period without power. In these cases, the large current persists until the monitor is switched off.

This is likely to be caused by a latchup (due to formation of a parasitic thyristor within the CMOS circuitry) caused by the choice of power-up sequence. A modified power-up sequence to prevent this problem has been proposed; however, this cannot be implemented in the prototype unit without changes to the circuitry. Instead, the HMRM power supply will be monitored by the spacecraft on-board computer and will be power-cycled if the current exceeds a specified safe maximum.

7.5.3 ASIC temperature

As a result of the relatively large power consumption, thermal regulation of the monitor was identified as a potential problem. Higher temperatures will affect the dark noise by increasing leakage current shot noise and reset noise, and so should be avoided where possible. Measurements of the ASIC with a thermocouple indicated an operating temperature of approximately 60°C in air at room temperature. In addition to the poorer heat dissipation in space, the HMRM will be situated on a spacecraft panel with the predicted temperature profile shown in Fig. 7.7. Since the ASIC is designed to tolerate temperatures up to 120°C , correct monitor operation is expected although noise performance may suffer. While a temperature sensor is provided within the ASIC, its data is not collected by the present FPGA program implementation in order to simplify the timing of control and readout waveforms.

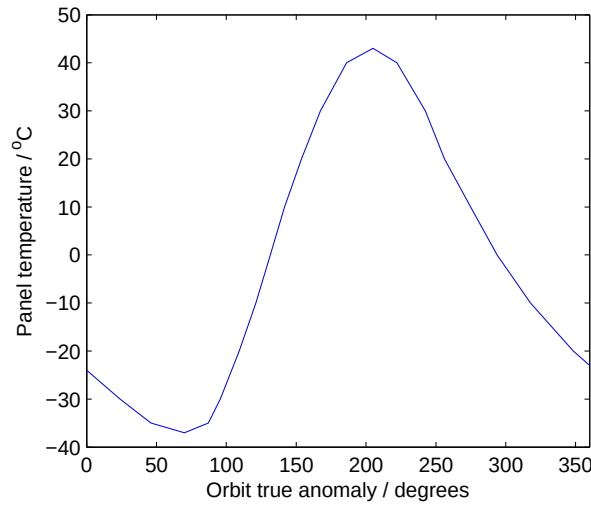


Figure 7.7: Predicted temperature profile of the spacecraft panel to which the HMRM will be attached (data provided by RAL Space). For an approximately circular orbit the true anomaly is proportional to time (the orbital period is ~ 100 minutes).

7.5.4 Pixel integration time and telemetry packet rate

During the monitor design process, the anticipated readout frame rate was 10 kHz, giving a frame time period of $100\ \mu\text{s}$. The maximum pixel exposure time (during which charge is integrated in the pixels) was thus $100\ \mu\text{s}$ also. This exposure time can be reduced by factors of two down to $3.125\ \mu\text{s}$ via the ‘electronic shutter’ system, described in Chapter 4.

Changes to the ASIC readout waveforms, however, meant that the frame time was increased to $127.7\ \mu\text{s}$, giving a frame rate of 7.83 kHz. This lengthened period will cause a slight increase in pile-up rates compared to the values predicted during the design process.

7.5.5 Pixel signal tests

A further test involved a visual check of pixel response by viewing ‘transparent frame’ read-outs. These frames are a simultaneous snapshot of all pixel values on all sensors of the monitor. By varying the trim and threshold settings, all eight pixel digital values may be obtained: an example is shown in Fig. 7.8. A large variation (FPN) is observed between individual pixel values because comparator trims were not applied at this stage. Note that offsets are correlated among half-columns as these pixels share a readout comparator.

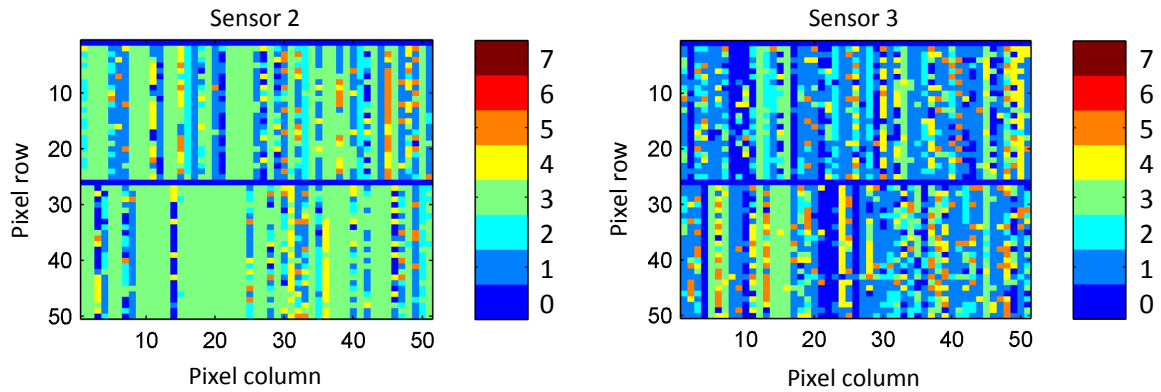


Figure 7.8: HMRM transparent readout for sensors 2 and 3 with default bias settings and no comparator trims applied

In all images acquired from sensors 2 and 3 there are two rows of pixels (rows 1 and 26) that are different from the others (having values consistently below the array mean, with little fluctuation). This observation was confirmed by the commercial evaluation testing, although it is not yet known if this may be rectified by altering the readout waveforms. The values of these pixels may be altered by changing the trim and threshold settings but do not show any other variation and so will be masked (rejected) during the following analysis.

7.5.6 Threshold scans

More detailed device testing is achieved by observing the pixel signals during a threshold scan; this is done using the lower comparator levels as these have higher resolution (but smaller range). Comparator level L1 was increased from a setting of zero to 127 while leaving all other levels at the same threshold value (L2–L7 all set to the maximum of 127). This procedure gives pixel values of 0 or 1 only, returned when pixel signals are less than or greater than L1, respectively. The scan is expected to be a smooth S-curve; however, as shown in Fig. 7.9, this was not observed.

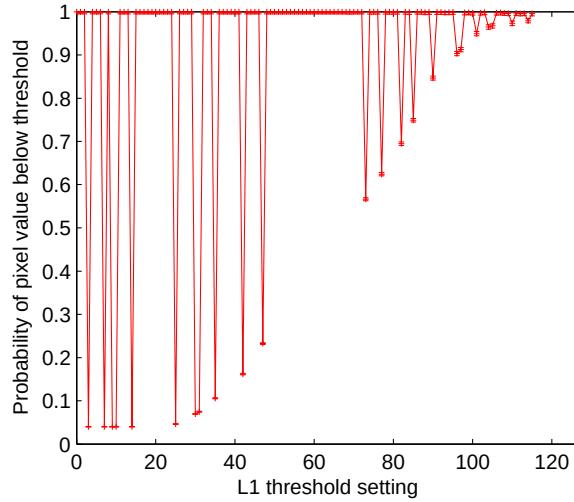


Figure 7.9: Threshold scan for comparator level L1 of sensor 2.

The fluctuations visible in Fig. 7.9 indicate that pixel 0 values occur more often than expected across much of the range. This problem may be due to pixel pickup/crosstalk between pixels and is discussed further in Section 7.6. Although thorough testing has not been completed, a similar effect has been observed with sensor 3. However, scans obtained using L2 on both sensors are apparently without errors and are capable of producing high resolution curves due to the small step size of this level. These have been used to make an initial calibration of sensors 2 and 3 in relative (DN) units.

7.6 Dark calibration

When in darkness, with no radiation present, the simplification $s_i = 0$ is made in Equation 7.4. In this case, if all trims p_j are set to zero, the mean signal measured for each pixel is:

$$m_i = c_j + d_i. \quad (7.12)$$

This value is referred to as the *total pixel offset*. Taking the mean over the 25 pixels of each comparator group gives:

$$\frac{1}{25} \sum_{i=1}^{25} m_i = c_j, \quad (7.13)$$

where, by definition:

$$\frac{1}{25} \sum_{i=1}^{25} d_i = 0. \quad (7.14)$$

In this situation the noise model (Equation 7.9) gives:

$$\sigma_{m_i} = \sigma_i. \quad (7.15)$$

The total pixel offsets (and hence the comparator offsets) and the pixel dark noise may be determined by fitting an Error function to the threshold scan data for each pixel and extracting the mean and standard deviation.

7.6.1 Sensor 2 calibration

All comparator trims were set to zero and the sensor placed in darkness with negligible radiation. Comparator level L2 was scanned across its full range: at each setting, the pixel values were read 33 times and the fraction below the threshold recorded, producing pixel S-curves. This result, for each individual pixel, is shown in Fig. 7.10; note that these plots are not images of the array but instead show the S-curve (value coded by colour) for each pixel simultaneously.

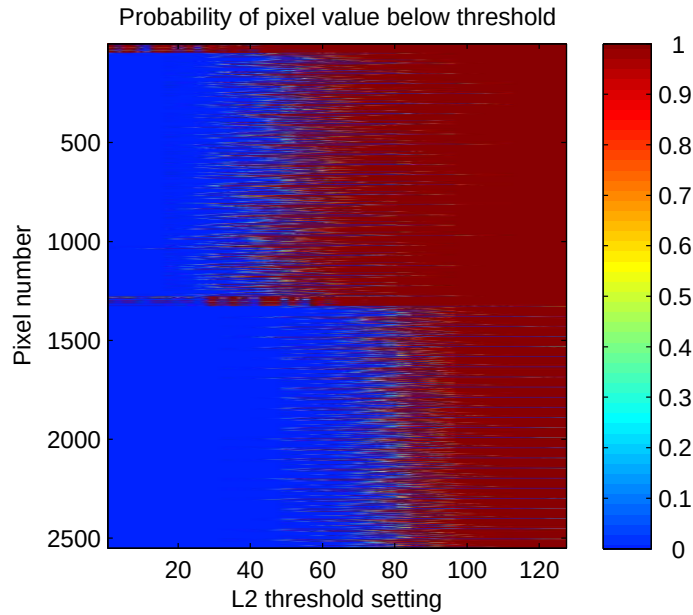


Figure 7.10: L2 threshold scan for sensor 2 calibration.

The pixel numbering used hereafter begins with pixel 1 at row 1, column 1, then increases along each row, finishing with pixel 2550 at row 50, column 51. Each comparator group is a half-column of 25 pixels: these are numbered in a similar pattern, so that comparators 1 to 51 include pixels 1 to 51, respectively, and comparators 52 to 102 include pixels 2500 to 2550. The two non-functioning rows of pixels, visible in these scan maps, therefore comprise pixel numbers 1 to 51 and 1276 to 1326.

Each individual pixel S-curve resulting from the L2 scan was fitted with an Error function, from which the mean and standard deviation were obtained. Fit examples are shown in Fig. 7.11 for pixel numbers 53 and 305.

Pixel offsets

Figure 7.12 shows the distribution of the extracted mean pixel signals; since the sensor is in darkness, with all trims set to zero, this is the total pixel offset (Equation 7.12), a measure of the offset dispersion of the pixels. This shows that, before trimming, a large offset range (of more than 80 DN) exists.

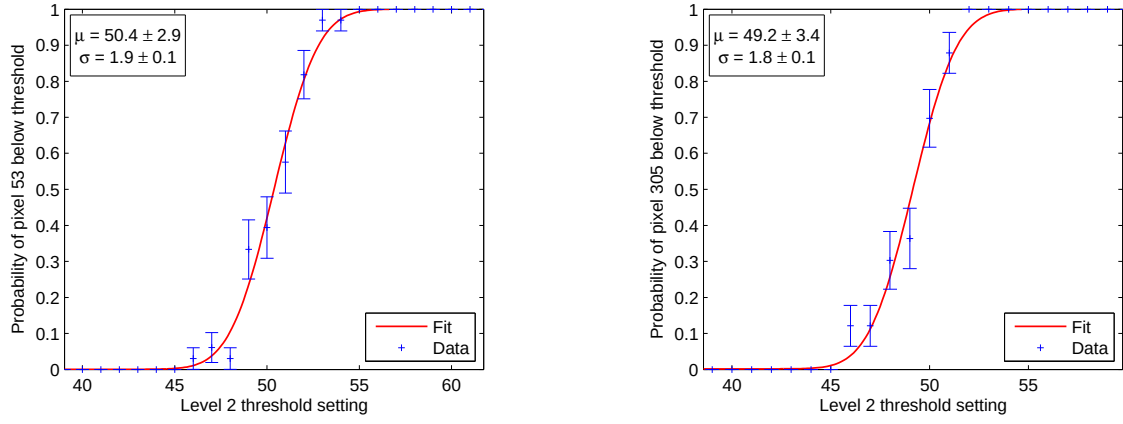


Figure 7.11: Error function fits to the L2 scan data for pixels 53 (left) and 305 (right).

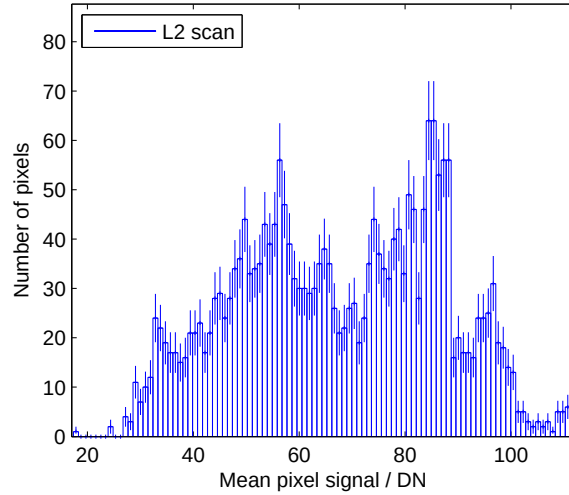


Figure 7.12: Distribution of mean pixel signals for sensor 2 measured in darkness with trims at zero. Under these conditions this quantity equals the total pixel offset.

Figure 7.13 shows the pixel mean (total pixel offset) in map form (corresponding to the pixel locations in the sensor array) and as a function of pixel number. Three important features are immediately apparent:

1. An average offset of approximately 30 DN exists between the upper and lower halves of the array (rows 1-25 and rows 26-50). This is understood to result from the significant difference in the track length of the readout circuitry between these two halves.
2. Pixel offset is correlated within each half-column of 25 pixels. This is also expected and results from these pixels sharing a comparator. This indicates that the offset originates primarily in the readout or digitisation circuitry, rather than within the pixel. This

correlation is also the cause of the periodicity visible in Fig. 7.13, since pixels sharing a comparator will occur every 51st pixel, using the chosen numbering system.

3. Certain rows and columns appear to be significantly different to the remainder (the majority of the array). These are: rows 1 and 26, and columns 1 and 51. Note that since each trim affects an entire half-column of pixels, the two rows (1 and 26) cannot be corrected by trimming, as this will adversely affect the remaining 24 pixels in each half-column. Instead, these pixels must be masked.

Although an attempt could be made to correct columns 1 and 51 by trim setting, it was decided to mask these pixels also (the adverse effects of masking are reduced because these pixels are at the edge of the array). This mask, which has been applied in all subsequent analyses, comprises a total of 198 pixels (7.8% of the array).

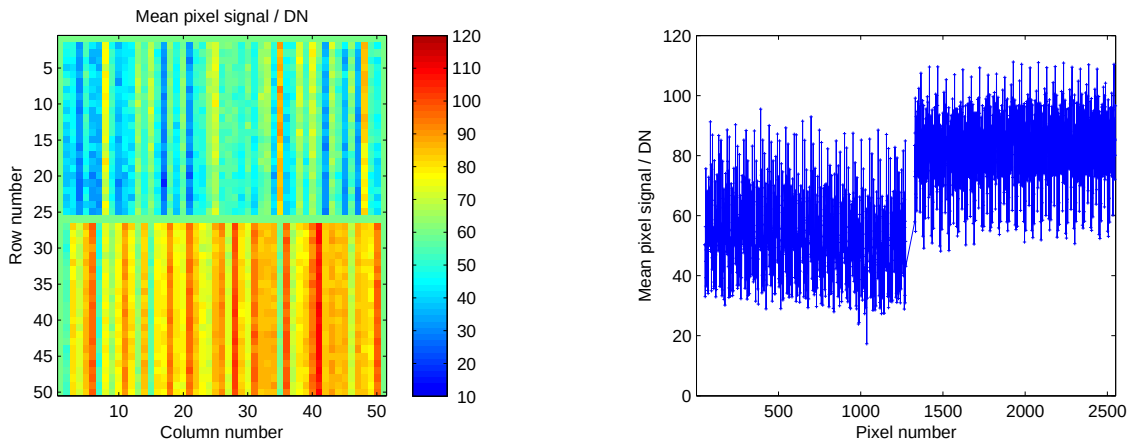


Figure 7.13: Mean pixel signals measured for sensor 2 in darkness with zero trims (the total pixel offset). These are shown as an array map (left) and as a function of pixel number (right).

Pixel dark noise

The width of the threshold scan S-curves, approximated as the standard deviation of the Error function fit, is a measure of the pixel noise in the absence of radiation. The distribution of these pixel noises is presented in Fig. 7.14, showing that the values appear approximately randomly distributed across the array. The most probable pixel noise for sensor 2 was measured to be 1.86 DN, with a distribution mean of 1.82 DN and standard deviation of 0.42 DN. Using a predicted conversion (from design considerations) of $1 \text{ DN} = 5.3 \text{ e}^-$, this gives a most probable pixel equivalent noise charge of $(9.9 \pm 0.2) \text{ e}^-$. Although a preliminary estimate only, this is consistent with design simulations.

Offset correction

The total pixel offset has been averaged over each comparator group of 25 pixels (excluding masked pixels as noted above), yielding 102 mean comparator offsets, c_j . This quantity is

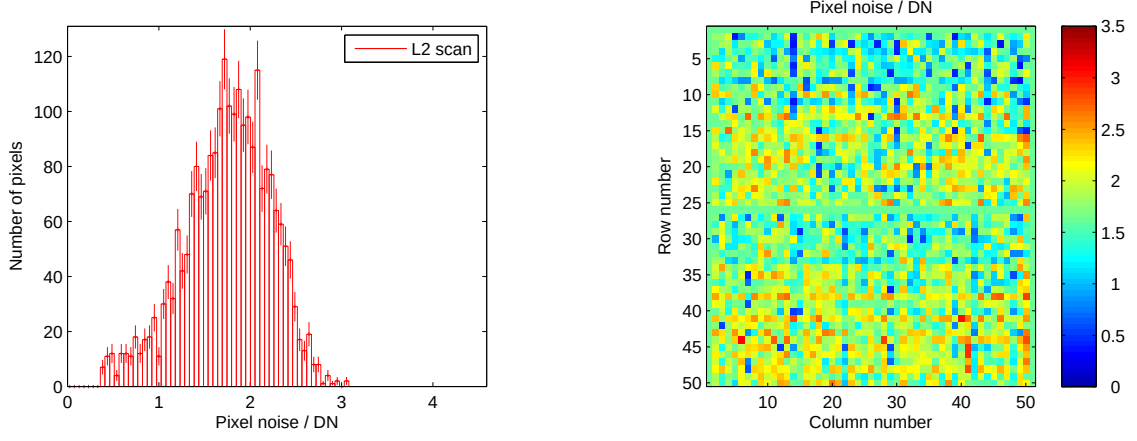


Figure 7.14: Pixel dark noise displayed as a histogram and in map form, showing the spatial distribution across the array. This is defined as the standard deviation of the pixel signal, measured via an Error function fit to the threshold scan data, when the sensor is in darkness.

referred to as the *effective* comparator offset since it will include the mean of any pedestal offset originating in the pixel, for each group of 25, as well as the (larger) offset at the comparator.

Referring to Equation 7.4, the aim of trimming is to null the comparator offsets, so that there is negligible difference between comparator groups. This corresponds to the condition that:

$$c_j + a_j p_j = 0. \quad (7.16)$$

In fact, Equation 7.16 is a special case since, in general, it is only necessary to remove the *difference* in offset between comparator groups:

$$c_j + a_j p_j = P, \quad (7.17)$$

where P is a constant for the whole array; in this case, the p_j are *relative trims* in that they may have an arbitrary constant value added. The value of P , together with the comparator threshold settings, must be chosen to put the mean dark signal sufficiently lower than the lowest comparator level for reliable noise rejection. This consideration will be addressed later: at present, the important quantities are the 102 relative p_j values that produce an equal (but arbitrary) offset for all comparator groups. This is:

$$p_j = -\text{round}\left(\frac{c_j}{a_j}\right) + z, \quad (7.18)$$

where z is simply chosen so that all p_j are brought into the allowed range of 0–255, and ‘round()’ indicates rounding to the nearest integer. First, however, the trim gain, a_j (in DN units), must be determined in order to allow the calculation of Equation 7.18.

Trim calibration

The trim gains are found by repeating the procedure described above after applying small shifts to the trims for all comparators. This causes a small shift in the location of each pixel’s

S-curve mean, as measured using threshold scans followed by Error function fits. Ten trim values were tested (0 to 9 in unit steps), the same values being applied to all comparators. After extracting the pixel mean in each case, a least-squares linear fit was used to obtain the formula for the S-curve mean as a function of trim setting, for each pixel. Two examples are shown in Fig. 7.15.

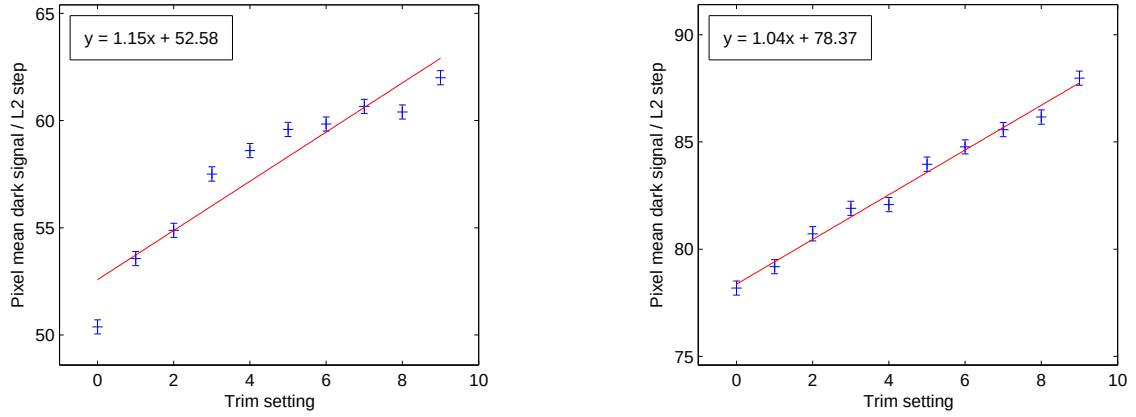


Figure 7.15: Linear fit to the pixel mean signal when expressed as a function of the comparator trim setting. The graphs show the result for pixel 53 (left) and pixel 365 (right). The line gradient and intercept are the trim gain and total pixel offset, respectively.

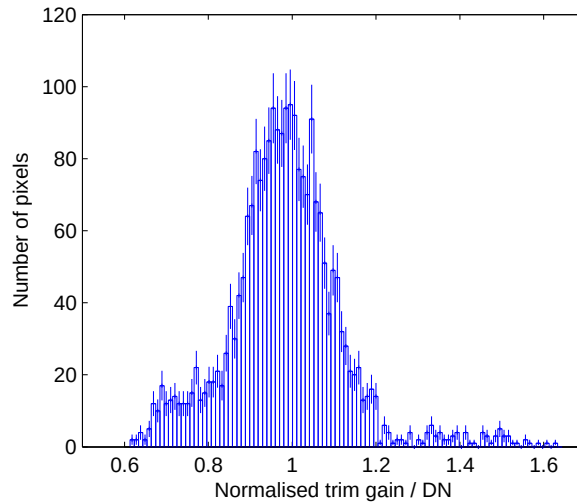


Figure 7.16: The distribution of trim gains, a_j , measured with each individual pixel, using a sequence of 10 global trim values.

Most pixels showed a satisfactorily linear response, with a minority displaying a less ideal response: examples of both are shown in Fig. 7.15. The gradient of the straight line for these graphs is the trim gain, as defined in Equation 7.4, while the intercept is the total pixel offset.

The result of extracting and histogramming the trim gain value for all pixels is shown in Fig. 7.16; the mean trim gain in the central peak was found to be 0.99 DN, with a standard deviation of 0.09 DN.

The total pixel offsets found by this method were consistent with the distribution found using a single trim setting (Fig. 7.12), although this new distribution is preferred since it is an average obtained over a range of trims.

Using the mean trim gain value, the relative trims were calculated using Equation 7.18 together with the comparator offsets derived from the total pixel offsets. These chosen trim values were then tested in a ‘software correction’ of the total pixel offsets. The resulting residual pedestal distribution is shown in Fig. 7.17, compared with the original distribution before trimming. The trimmed distribution has a standard deviation of 2.46 DN ($\sim 13.0 \text{ e}^-$); the residual pedestal dispersion is therefore slightly larger than the pixel dark noise.

All comparators were successfully trimmed by values well within the 0–255 range, having an overall p_j range of 78. This provides scope for shifting all comparator offsets by an additional amount, if necessary, as explained previously.

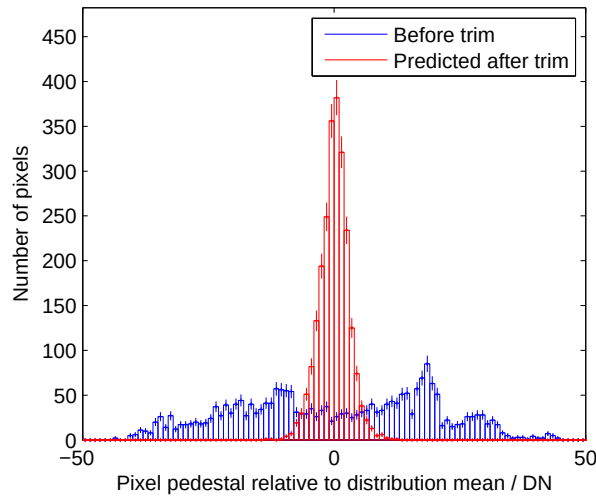


Figure 7.17: The distribution of the predicted pixel offset after trimming, compared to the case with all trims zero, for an L2 scan.

Combined dark pixel probability distribution

By using the scan data for each pixel and applying the chosen trim settings, the combined probability distribution for a dark pixel signal may be calculated. This treats all pixels as indistinguishable (each is equally likely to be chosen, after having applied the mask). The width of this distribution is a more relevant quantity than the single pixel noise, since this combined distribution includes pixel-to-pixel variability and thus is a closer approximation to the expected variance in a detected radiation signal. This distribution is shown in Fig. 7.18, together with the result of a Gaussian distribution fit. Note that this is a predicted result,

since in actual operation the trim values will be loaded into the ASIC. A true test of this process, where the trims are applied in hardware and the resulting pixel distributions measured directly, is presented below.

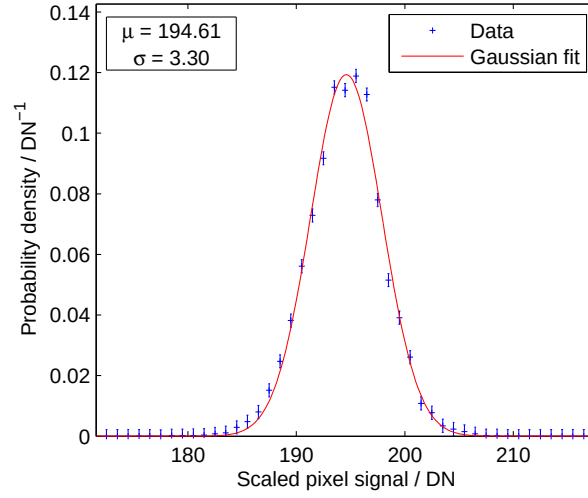


Figure 7.18: Combined pixel dark signal probability distribution for sensor 2, using a comparator trim software correction.

Verification of trim selection

The predicted effect of the chosen comparator trims has been obtained by simple subtraction of the trim value from the measured pixel signals (a ‘software correction’). In practice, these trims must be loaded into the ASIC where they are applied to the comparator in a ‘hardware correction’.

Table 7.4: Chosen trim settings for sensor 2 (an additional constant offset may be also be applied). X indicates comparators with all pixels masked.

Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer	Comparator	Trim integer
1	X	13	133	25	126	37	146	49	143	61	118	73	112	85	110	97	126
2	151	14	143	26	143	38	128	50	147	62	105	74	120	86	135	98	110
3	151	15	130	27	139	39	144	51	X	63	116	75	124	87	99	99	113
4	165	16	147	28	141	40	127	52	X	64	132	76	110	88	127	100	116
5	147	17	166	29	146	41	140	53	139	65	112	77	105	89	117	101	98
6	158	18	138	30	139	42	161	54	115	66	139	78	123	90	113	102	X
7	159	19	151	31	149	43	132	55	125	67	117	79	95	91	102		
8	122	20	136	32	136	44	141	56	111	68	118	80	114	92	88		
9	145	21	164	33	132	45	160	57	100	69	103	81	122	93	112		
10	159	22	146	34	153	46	137	58	143	70	116	82	101	94	108		
11	153	23	145	35	110	47	163	59	120	71	119	83	111	95	111		
12	151	24	134	36	153	48	114	60	131	72	100	84	110	96	110		

The success of the hardware correction can be tested by repeating the threshold scan process and measuring the remaining offset of each pixel. This was carried out for sensor 2, using the trim values given in Table 7.4, producing the scan map and combined pixel S-curve shown in Fig. 7.19. Comparing this map to that obtained with all trims zero (Fig. 7.10), it is immediately clear that the pixel offsets are now more consistent.

The mean pixel signal measured from Error function fits to each pixel scan is shown in Fig. 7.20 in map and histogram form. These results show a close agreement with the predicted distribution shown previously and exhibit a total offset range of approximately 10 DN. Significantly, correlation between pixels in rows can now be observed, whereas the previously strong correlation between comparator group pixels (half columns) is now negligible. The small differences between rows may be due to the presence of circuit tracks over some pixels.

Similar correlations may be seen in the noise distribution map measured with this scan, shown in Fig. 7.21. The noise distribution histogram shows that the most probable noise (approximately 1.9 DN) is the same as previously measured in Section 7.6.1, although the distribution appears slightly more dispersed.

The combined pixel probability density function, measured directly without making any software correction other than pixel masking, is shown in Fig. 7.22. This has a satisfactory full signal range of approximately 20 DN: this value incorporates the effects of single pixel noise (and the variation of this quantity between pixels) and residual pedestal dispersion. Note that the quadruple-peak appearance is also visible in Fig. 7.19 as three vertical bands in the scan map and as a central step and two outer fluctuations in the S-curve.

A Gaussian distribution (with S-curve integral) is not necessarily to be expected. This is because the pedestals are correlated and also because the trims are discrete, meaning that certain residual values may occur more often; this is the likely cause of the S-curve central step. However, the side peaks are more problematic, since the S-curve is expected to increase monotonically with threshold setting (neglecting statistical fluctuations).

These fluctuations are instead thought to be the result of electrical crosstalk (pickup) between readout circuits. This is distinct from charge diffusion crosstalk between adjacent pixels and could be caused by the sudden change in output state of several comparators when a large number of pixels change value simultaneously. This would affect a trimmed sensor more severely, since the actual aim of trimming is to achieve this simultaneity.

A pickup interference of this sort was observed in threshold scans of a similar sensor [88] – in that case, pixels could be disabled and only a small subset scanned at any time. A possible solution to allow calibration scans with the HMRM sensor would be to deliberately offset the majority of comparator trims by different amounts so as to ‘stagger’ the pixel transitions. A small number of comparators could be left with the chosen optimal trims for testing with a threshold scan. This process would then be repeated with a different set of comparators under test.

Radiation monitoring may not be adversely affected by this pickup because only a small number of pixels are expected to change value in any single readout frame. The optimised

trims may therefore be safe to use in this mode even if the existence of this problem is confirmed.

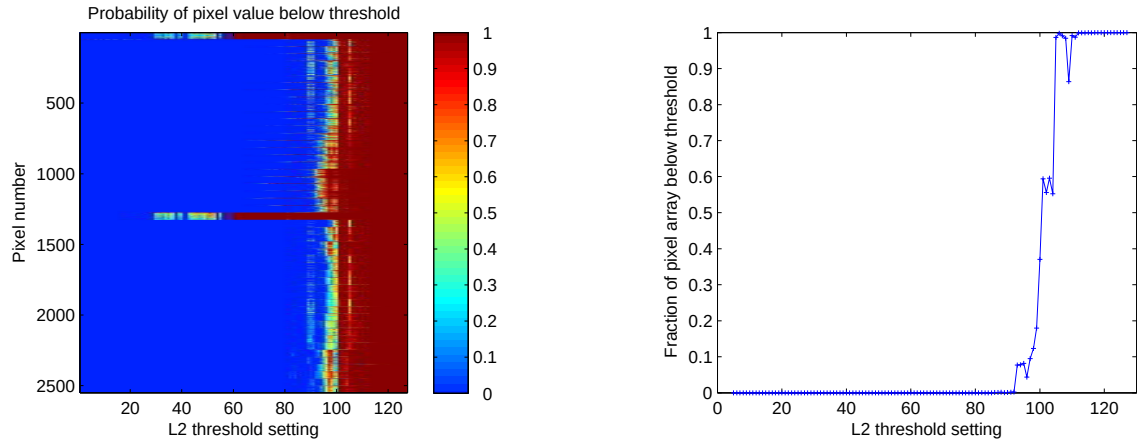


Figure 7.19: Results of the L2 threshold scan for sensor 2, after applying the chosen trims, showing data for each individual pixel (left) and as a combined S-curve for the whole array (right).

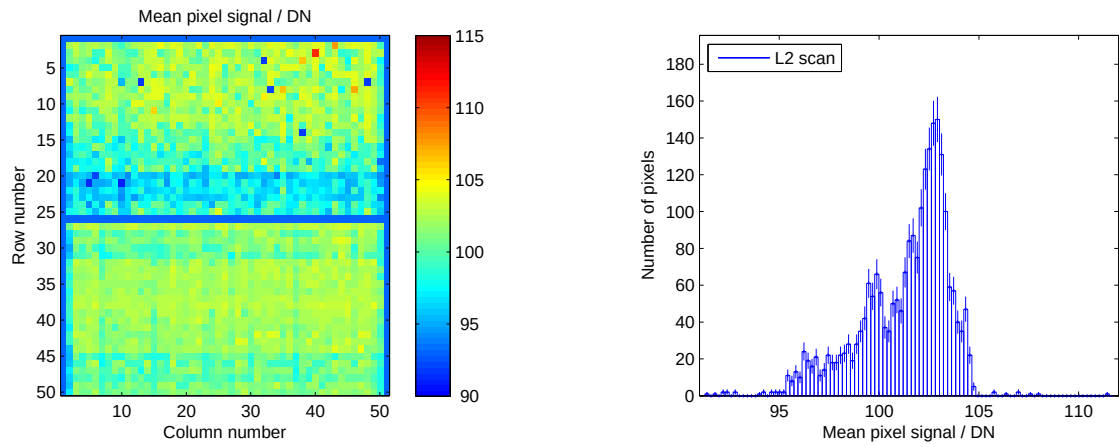


Figure 7.20: Mean dark pixel signals measured using a L2 threshold scan for sensor 2 after applying the initial trim selection, shown in array map and histogram form.

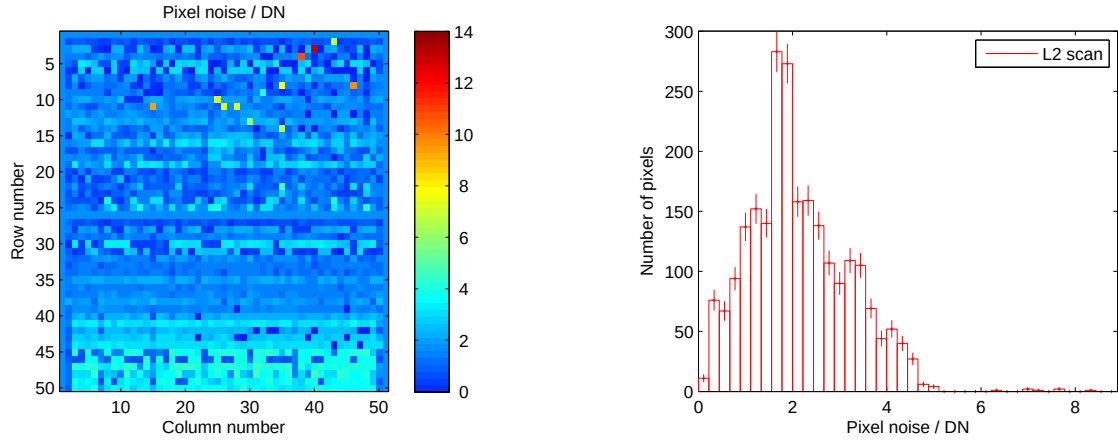


Figure 7.21: Pixel dark noise map (left) and histogram (right) for sensor 2, after applying the chosen trims.

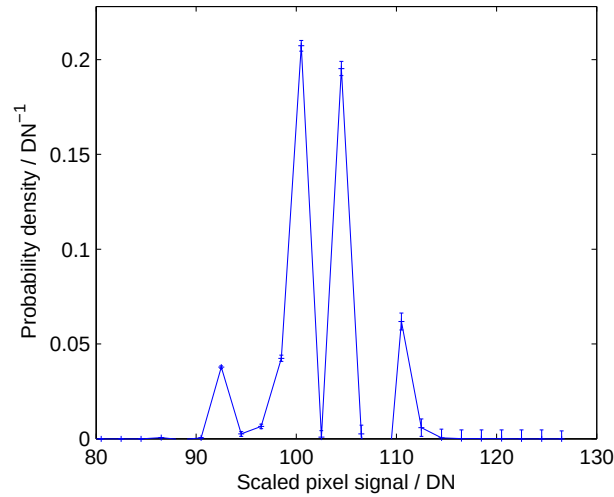


Figure 7.22: Combined pixel probability distribution measured for sensor 2 after applying the initial trims in hardware. No software corrections were applied to the data other than a pixel mask.

7.6.2 Sensor 3 calibration

The same dark calibration procedure was followed with sensor 3, the second of the two ASIC sensors installed in this HMRM prototype. The same two rows (1 and 26) of pixels were identified as anomalous, indicating that this is an error in the readout procedure or device design, rather than in an individual component. The same mask as had been selected for sensor 2 was therefore also used for sensor 3.

Broadly similar results were obtained for sensor 3 as for sensor 2; Fig. 7.23 shows the offset and noise distributions produced from an L2 scan. An overall offset range of more than 110 DN was found, which is greater than that measured for sensor 2 (approximately 80 DN).

An average offset difference of approximately 35 DN exists between the upper (rows 1-25) and lower (rows 26-50) halves of the array, similar to the value for sensor 2.

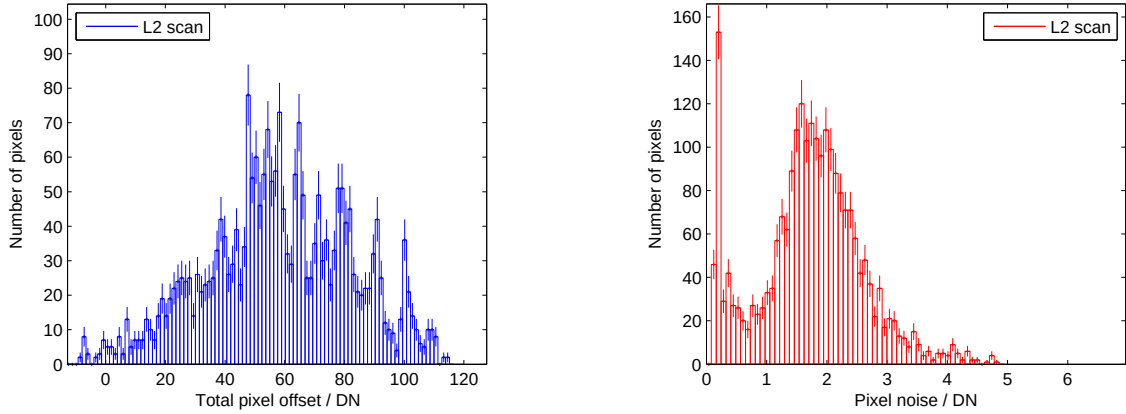


Figure 7.23: Total pixel offsets (left) and dark noise (right) for sensor 3.

The noise distribution central peak gives a mean noise of 1.90 DN ($\sim 10.1 \text{ e}^-$) and a standard deviation of 0.47 DN; both are close to the values measured for sensor 2. Operational trim settings were selected for sensor 3 using the same procedure as that described for sensor 2.

7.6.3 Ongoing calibration work

At the time of writing, the following procedures remain to be completed before the ASICs may be considered fully calibrated:

1. Perform threshold scans for each comparator level to confirm correct operation and to measure the inter-level gains (discussed in Section 7.2.5).
2. Use X-ray exposure to determine the system gain (by measurement of the full-energy X-ray photopeaks of ^{55}Fe).
3. Select comparator levels suitable for radiation monitor operation.
4. Repeat the calibration to assess noise and pedestal temperature dependence.
5. Characterise the charge sharing behaviour of the pixels in response to different particles (e.g. electron, proton, gamma and alpha).

Although not sufficient to allow full monitor functionality as designed, the present calibration data already allow a simplified system of operation based on particle counting. This is important if significant anomalies are identified in the device initially selected for flight. Due to the flexibility of the physics algorithm, this may be achieved without any modification of the monitor hardware or FPGA code. Instead of reconstructing the total energy deposit of each event, the monitor would count the number of pixels above a single threshold level (the cluster size). The algorithm would then operate on the cluster size data, in place of the energy deposit, providing particle count rates and cluster size histograms (the latter including sensor coincidence information).

7.7 Summary

A revised ASIC for the HMRM project has been manufactured and a two-sensor monitor prototype has been constructed successfully, validating the mechanical design and assembly procedure. The preliminary calibration has provided pixel dark noise measurements for sensors 2 and 3, with tentative equivalent noise charge values of approximately 10 e^- (using a predicted conversion factor) in accordance with design predictions. Additionally, the system of applying an on-chip trim correction to null the comparator offset errors has been verified, greatly reducing the fixed-pattern noise. Both of the input data shift registers have been shown to operate correctly, confirming that the error in the first ASIC version has been corrected in the latest version.

While some of the important sensor parameters have been measured, the ASICs have not yet been fully investigated or calibrated. This ongoing process must be completed before testing of the HMRM in radiation monitoring mode (for example, with particle beams or radioactive sources).

Chapter 8

Conclusion

The energetic particle environment in Earth orbit poses a significant risk to spacecraft operation. The fluxes are dominated by protons and electrons trapped in the Earth's magnetosphere together with unpredictable, transient increases at higher energies due to solar events. The interactions of these particles with spacecraft systems lead to cumulative degradation of components and single-event errors in electronics. Especially affected are semiconductor devices, including such important components as solar cells, imagers and computational electronics. Safety concerns for human space habitation and exploration pose even greater challenges from this perspective.

The use of radiation monitors on satellites allows improvements in design and mission planning (including estimated lifespan) and introduces the possibility of real-time alerting. Due to the dependence of damage capability on the particle species and energy, particle identification is a valuable advantage compared to simple dosimetry. This includes the identification of single particles and the longer-term reconstruction of the mean incident particle energy spectra. The development of a small, accurate instrument suitable for widespread use on satellites in Earth orbit could therefore open new prospects for radiation detection in space, with immediate application to commercial as well as scientific payloads.

The HMRM is a significant development step, with very low mass (~ 50 g) and modest engineering/integration costs, while providing a useful range of data products. These include detailed data characterising individual particle events in addition to simple dosimetry; Monte Carlo simulations have shown that certain particle types may be identified reliably from these event characterisations. More generally, particle energy spectra of useful resolution may be derived from HMRM data using a reconstruction method operating externally to the monitor. The HMRM is likely to be the first space radiation monitor to use CMOS APS devices for particle detection, benefiting from their excellent detection efficiency, radiation tolerance and low noise performance.

The monitor design was optimised for the detection of particles through multiple sampling of their energy deposits in a sensor telescope. This was undertaken with severe constraints on internal layout due to the small size of the instrument, leading necessarily to some compromises such as increased sensor separation. This did, however, result in a miniature device

with a volume of less than 10 cm^3 . Simulated performance was assessed in several important Earth orbits, from LEO to GEO, demonstrating suitability for operation in fluxes up to $\sim 10^8 \text{ cm}^{-2} \text{ s}^{-1}$.

On-board data analysis algorithms have been developed for real-time execution within the HMRM FPGA. These allow a fast response to rapidly changing environments, while retaining sufficient data for additional offline analyses. Dose rates and count rates are augmented by the provision of categorised event counts, some of which are strongly associated with a specific particle species and/or energy. These aspects of the design, together with the exposure time shuttering algorithm, have been tested using time-dependent orbit simulations, demonstrating good predicted performance in important Earth orbits.

The reliance on Geant4 Monte Carlo simulations throughout the development process has been assured through a validation of relevant particle interaction phenomena such as scattering and energy loss. In addition, charge diffusion simulations were used to provide insight into the response of pixel sensors to ionising energy deposits. Modelled output HMRM datasets have also been used as the inputs to reconstruction algorithms for particle energy spectra. A new procedure, developed for this project, has produced good reconstruction results even for short exposure times.

The project is now proceeding to the calibration and testing phase. Initial results, described in this thesis, indicate that the ASIC sensors achieve low noise performance as anticipated; further prototype monitor testing, involving particle beam and radioisotope exposure, is scheduled for 2013. Meanwhile, a prototype HMRM is being integrated onto the TechDemoSat-1 satellite which is due to be launched later in the same year.

Future phases of the programme will address design iteration and spaceflight qualification, with completion predicted for 2015–2016. A successful HMRM development and qualification process may pave the way for widespread use of this novel technology in many commercial and scientific settings.

Bibliography

- [1] European Space Research and Technology Centre. Invitation to Tender: Highly Miniaturised Active MEO Radiation Monitor Phase A-B. Technical Report AO/1-5978/09/NL/AT, ESA, 2009.
- [2] H. Araújo and E. F. Mitchell. HMRM simulation results and reconstruction algorithms. Technical Report HMRM-IC-REP-003, Imperial College London, 2010.
- [3] C. W. Carlson *et al.* The fast auroral snapshot (FAST) mission. *Geophys. Res. Lett.*, 25:2013–2016, 1998.
- [4] T. A. Farley and A. D. Tomassian. Source of high-energy protons in the Van Allen radiation belt. *Phys. Rev. Lett.*, 25:47–49, 1970.
- [5] J. R. Cummings *et al.* New evidence for geomagnetically trapped anomalous cosmic rays. *Geophys. Res. Lett.*, 20:2003–2006, 1993.
- [6] D. M. Sawyer and J. I. Vette. AP-8 Trapped Proton Environment for Solar Maximum and Solar Minimum. *SSDC/WDC-A-RS*, 1976.
- [7] J. I. Vette. The AE-8 Trapped Electron Model Environment. *SSDC/WDC-A-RS*, 1991.
- [8] ESA. SPENVIS: The space environment information system, www.spenvis.oma.be.
- [9] D. Heynderickx *et al.* New radiation environment and effects models in ESA’s space environment information system (SPENVIS). In *RADECS Proceedings*, pages 643–646, 2003.
- [10] H. Heckman and H. Nakano. East-west asymmetry in the flux of mirroring geomagnetically trapped protons. *J. Geophys. Res.*, 68:2117–2120, 1963.
- [11] G. D. Badhwar and A. Konradi. Conversion of omnidirectional proton fluxes into a pitch angle distribution. *J. Spacecraft Rockets*, 27:350–352, 1990.
- [12] C. Grimani *et al.* Cosmic-ray spectra near the LISA orbit. *Classical Quant. Grav.*, 21(5):S629, 2004.
- [13] Energetic particle environment in near-Earth orbit. *Adv. Space Res.*, 17:37–45, 1996.
- [14] J. Liliensten. *Space weather: research towards applications in Europe*. Springer, 2007.
- [15] R-G. Wang. Proton spectrum of the 2005 January 20 solar flare. *Chinese Phys. C*, 32(2), 2008.

- [16] J. Feynman *et al.* The JPL proton fluence model: an update. *J. Atmos. Sol.-Terr. Phy.*, 64:1679–1686, 2002.
- [17] J. Feynman *et al.* Interplanetary proton fluence model: JPL 1991. *J. Geophys. Res.*, 98:13281–13294, 1993.
- [18] L. Rosenqvist *et al.* Toolkit for updating interplanetary proton cumulated fluence models. *J. Spacecraft Rockets*, 42:1077–1090, 2005.
- [19] M. Ackermann *et al.* Detection of the characteristic pion-decay signature in supernova remnants. *Science*, 339:807–811, 2013.
- [20] J. A. Simpson. Elemental and isotopic composition of the galactic cosmic rays. *Annu. Rev. Nucl. Part. S.*, 33:323–382, 1983.
- [21] D. Reames *et al.* Energetic particle abundances in impulsive solar flare events. *Astrophys. J. Suppl. S.*, 90:649–667, 1994.
- [22] L. A. Fisk *et al.* An interpretation of the observed oxygen and nitrogen enhancements in low-energy cosmic rays. *Astrophys. J.*, 190:L35, 1974.
- [23] M. E. Pesses *et al.* Cosmic ray drift, shock wave acceleration, and the anomalous component of cosmic rays. *Astrophys. J.*, 246:L85–L88, 1981.
- [24] H. A. Bethe. Molière’s theory of multiple scattering. *Phys. Rev.*, 89:1256–1266, 1953.
- [25] J. Beringer *et al.* (Particle Data Group). Review of particle physics. *Phys. Rev. D*, 86:010001, 2012.
- [26] J. V. Jelley. Cerenkov radiation. *Progress in Nuclear Physics*, 3:84, 1953.
- [27] H. Bethe. Zur theorie des durchgangs schneller korpuskularstrahlen durch materie. *Ann. Physik*, 397:325–400, 1930.
- [28] N. J. Carron. *An introduction to the passage of energetic particles through matter*. Taylor and Francis, 2006.
- [29] R. M. Sternheimer. The density effect for the ionization loss in various materials. *Phys. Rev.*, 88:851–859, 1952.
- [30] M. J. Berger *et al.* ESTAR: Electron stopping power and range. *National Institute of Standards and Technology*, <http://physics.nist.gov/PhysRefData/Star/Text/ESTAR.html>.
- [31] W. H. Barkas *et al.* Mass-ratio method applied to the measurement of l-meson masses and the energy balance in pion decay. *Phys. Rev.*, 101:778, 1956.
- [32] F. Bloch. Zur bremsung rasch bewegter teilchen beim durchgang durch materie. *Ann. Physik*, 408:285–320, 1933.
- [33] J. F. Ziegler. Stopping of energetic light ions in elemental matter. *J. appl. phys.*, 85:1249–1272, 1999.
- [34] L. D. Landau. On the energy loss of fast particles by ionization. *J. Exp. Phys. (USSR)*, 8:201, 1944.

- [35] P. V. Vavilov. Ionization losses of high-energy heavy particles. *Sov. Phys. JETP*, 5:749, 1957.
- [36] G. Hall. Ionisation energy losses of highly relativistic charged particles in thin silicon layers. *Nucl. Instrum. Methods*, 220:356–362, 1984.
- [37] M. W. Thompson. *Defects and radiation damage in metals*. Cambridge University Press, 1969.
- [38] J. W. Corbett. *Electron radiation damage in semiconductors and metals*. Academic Press, 1966.
- [39] G. F. Knoll. *Radiation detection and measurement (4th. edition)*. John Wiley and Sons, Inc., 2010.
- [40] H. Spieler. *Semiconductor detector systems*. Oxford Science Publications, 2005.
- [41] A. Holmes-Siedle and L. Adams. *Handbook of radiation effects*. Oxford Science Publications, 1993.
- [42] L. Rossi *et al.* *Pixel detectors from fundamentals to applications*. Springer, 2006.
- [43] H. L. Hughes and J. M. Benedetto. Radiation effects and hardening of MOS technology: devices and circuits. *IEEE Trans. Nuc. Sci.*, 50:500–521, 2003.
- [44] G. Lutz. *Semiconductor radiation detectors*. Springer, 1999.
- [45] ICRP Publication 103. *The 2007 Recommendations of the International Commission on Radiological Protection, Annals of the ICRP*. Elsevier, 2007.
- [46] United Nations. *Report of the United Nations Scientific Committee on the Effect of Atomic Radiation, General Assembly*. Official Records A/63/46, 2008.
- [47] G. D. Badhwar and F. A. Cucinotta. Depth dependence of absorbed dose, dose equivalent and linear energy transfer spectra of galactic and trapped particles in polyethylene and comparison with calculations of models. *Radiat. Res.*, 149:209–218, 1998.
- [48] E. R. Benton and E. V. Benton. Space radiation dosimetry in low-Earth orbit and beyond. *Nucl. Instrum. Meth. B*, 184:255–294, 2001.
- [49] T. Ersmark *et al.* Influence of geometry model approximations on Geant4 simulation results of the Columbus/ISS radiation environment. *Radiat. Meas.*, 42:1342–1350, 2007.
- [50] G. Reitz *et al.* Space radiation measurements on-board ISS — the DOSMAP experiment. *Radiat. Prot. Dosim.*, 116:374–379, 2005.
- [51] G. D. Badhwar. Radiation measurements in low Earth orbit: US and Russian results. *Health Phys.*, 79:507–514, 2000.
- [52] P. Spillantini *et al.* Shielding from cosmic radiation for interplanetary missions: active and passive methods. *Radiat. Meas.*, 42:14–23, 2007.
- [53] F. A. Cucinotta and M. Durante. Cancer risk from exposure to galactic cosmic rays: implications for space exploration by human beings. *Lancet Oncol.*, 7:431–435, 2006.

- [54] J. A. Van Allen *et al.* Radiation observations with satellite 1958 ϵ . *J. Geophys. Res.*, 64:271–286, 1959.
- [55] J. A. Van Allen. The geomagnetically trapped corpuscular radiation. *J. Geophys. Res.*, 64:1683–1689, 1959.
- [56] M. Cyamukungu *et al.* Magnetic storm acceleration of radiation belt electrons observed by the Scintillating Fibre Detector (SFD) onboard EQUATOR-S. In *Ann. Geophys.*, volume 17, pages 1622–1625, 1999.
- [57] A. D. Johnstone. The geometric factor of a cylindrical plate electrostatic analyzer. *Rev. Sci. Instrum.*, 43:1030–1040, 1972.
- [58] D. McCann *et al.* Miniature ion mass analyzer. *Planet. Space Sci.*, 55:1190–1196, 2007.
- [59] S. Barabash *et al.* The analyser of space plasmas and energetic atoms (ASPERA-4) for the Venus Express mission. *Planet. Space Sci.*, 55:1772–1792, 2007.
- [60] F. Barao. AMS—Alpha Magnetic Spectrometer on the International Space Station. *Nucl. Instrum. Meth. A*, 535:134–138, 2004.
- [61] M. Aguilar *et al.* First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5 – 350 GeV. *Phys. Rev. Lett.*, 110:141102, 2013.
- [62] G. Bertolini and A. Coche. *Semiconductor Detectors*. Elsevier Science, 1968.
- [63] U. Fano. Ionization Yield of Radiations. II. The Fluctuations of the Number of Ions. *Phys. Rev.*, 72:26–29, 1947.
- [64] B. Taylor *et al.* Results from the Galileo GioveA radiation monitors and comparison with existing radiation belt models. *IEEE Trans. Nuc. Sci.*, 54:1076–1081, 2007.
- [65] S-J. Kim *et al.* Use of a MOSFET for radiation monitoring in space and comparison with the NASA trapped particle model. *J. Korean Phys. Soc.*, 48:865–869, 2006.
- [66] I. Thomson *et al.* Dose-depth and SEU monitors for the STRV-1c satellite. *IEEE Trans. Nuc. Sci.*, 45:2765–2770, 1998.
- [67] A. Mohammadzadeh *et al.* The ESA Standard Radiation Environment Monitor program first results from PROBA-I and INTEGRAL. *IEEE Trans. Nuc. Sci.*, 50:2272–2277, 2003.
- [68] H. D. R. Evans *et al.* Results from the ESA SREM monitors and comparison with existing radiation belt models. *Adv. Space Res.*, 42:1527–1537, 2008.
- [69] M. Siegl *et al.* Inner Belt Anisotropy Investigations Based on the Standard Radiation Environment Monitor (SREM). *IEEE Trans. Nuc. Sci.*, 57:2017–2023, 2010.
- [70] P. McNamara *et al.* (LISA Pathfinder Science Working Team). LISA Pathfinder. *Classical Quant. Grav.*, 25, 2008.
- [71] H. Araújo *et al.* Detailed calculation of test-mass charging in the LISA mission. *As-tropart. Phys.*, 22:451–469, 2005.

- [72] S. McKenna-Lawlor *et al.* A high-energy (35–500MeV) proton monitor for the Gravity Probe-B Mission. *Nucl. Instrum. Meth. A*, 498:220–230, 2003.
- [73] T. S. Lomheim *et al.* Imaging charge-coupled device (CCD) transient response to 17 and 50 MeV proton and heavy-ion irradiation. *IEEE Trans. Nuc. Sci.*, 37:1876–1885, 1990.
- [74] A. M. Chugg *et al.* A CCD miniature radiation monitor. *IEEE Trans. Nuc. Sci.*, 49:1327–1332, 2002.
- [75] X. Llopart *et al.* Medipix2: A 64-k pixel readout chip with 55 μm square elements working in single photon counting mode. *IEEE Trans. Nuc. Sci.*, 49:2279–2283, 2002.
- [76] X. Llopart *et al.* Timepix, a 65k programmable pixel readout chip for arrival time, energy and/or photon counting measurements. *Nucl. Instrum. Meth. A*, 581:485–494, 2007.
- [77] L. S. Pinsky *et al.* Penetrating heavy ion charge and velocity discrimination with a TimePix-based Si detector (for space radiation applications). *Nucl. Instrum. Meth. A*, 633:S190–S193, 2011.
- [78] L. S. Pinsky *et al.* Preparing for the first Medipix detectors in space. In *Aerosp. Conf. Proc.*, pages 1–6, 2012.
- [79] S. Mendis *et al.* CMOS active pixel image sensor. *IEEE Trans. Electron. Dev.*, 41:452–453, 1994.
- [80] E. El-Sayed *et al.* Design and characterization of ionizing radiation-tolerant CMOS APS image sensors up to 30 Mrd (Si) total dose. *IEEE Trans. Nuc. Sci.*, 48:1796–1806, 2001.
- [81] C. C. Liebe *et al.* Micro APS based star tracker. In *Aerosp. Conf. Proc.*, volume 5, pages 2285–2299, 2002.
- [82] M. Stanitski. Advanced monolithic active pixel sensors for tracking, vertexing and calorimetry with full CMOS capability. *Nucl. Instrum. Meth. A*, 650:178–183, 2011.
- [83] Y. Zhou *et al.* CMOS pixel sensor for a space radiation monitor with very low cost, power and mass. *JINST*, 7:C12003, 2012.
- [84] S. Agostinelli *et al.* Geant4 - A simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250–303, 2003.
- [85] J. Allison *et al.* Geant4 developments and applications. *IEEE Trans. Nuc. Sci.*, 53(1):270–278, 2006.
- [86] E. F. Mitchell and H. Araújo. HMRM design recommendations. Technical Report HMRM-IC-TN-041, Imperial College London, 2011.
- [87] ESA directorate of technical and quality management. *Statement of work: Highly miniaturised radiation monitoring and highly miniaturised active MEO radiation monitor Ph. A-B*. T604-001EE and T704-027EE, 2009.

- [88] J. A. Ballin *et al.* Design and performance of a CMOS study sensor for a binary readout electromagnetic calorimeter. *JINST*, 6:5009, 2011.
- [89] R. E. Coath *et al.* A low noise pixel architecture for scientific CMOS monolithic active pixel sensors. *IEEE Trans. Nuc. Sci.*, 57:2490–2496, 2010.
- [90] W. Dulinski *et al.* Radiation hardness study of an APS CMOS particle tracker. In *IEEE Nucl. Sci. Conf. R.*, volume 1, pages 100–103, 2001.
- [91] E. S. Eid *et al.* Design and characterization of ionizing radiation-tolerant CMOS APS image sensors up to 30 Mrd (Si) total dose. *IEEE Trans. Nuc. Sci.*, 48:1796–1806, 2001.
- [92] J. A. Ballin *et al.* Monolithic Active Pixel Sensors (MAPS) in a Quadruple Well Technology for Nearly 100% Fill Factor and Full CMOS Pixels. *Sensors*, 8:5336–5351, 2008.
- [93] Sheldahl, Multek corporation. The Red Book: Technical Product data. <http://www.sheldahl.com/Products/WebPages/RedBook.aspx>.
- [94] P. W. Fortescue *et al.* *Spacecraft Systems Engineering (3rd. edition)*. Wiley-Blackwell, 2003.
- [95] R. Li *et al.* Damage effect of keV proton irradiation on aluminized Kapton film. *Radiat. Phys. Chem.*, 77:482–489, 2008.
- [96] R. K. Hart. The oxidation of aluminium in dry and humid oxygen atmospheres. *P. Roy. Soc. Lond. A. Mat.*, 236:68–88, 1956.
- [97] M. Born and E. Wolf. *Principles of Optics (6th edition)*. Cambridge University Press, 2006.
- [98] G. Hass and J. E. Waylonis. Optical constants and reflectance and transmittance of evaporated aluminum in the visible and ultraviolet. *J. Opt. Soc. Am.*, 51:719–722, 1961.
- [99] E. D. Palik. *Handbook of Optical Constants of Solids 3*. Academic Press, 1998.
- [100] A. D. Rakic. Algorithm for the determination of intrinsic optical constants of metal films: application to aluminum. *Appl. Opt.*, 34:4755–4767, 1995.
- [101] N. Baier (Sheldahl Product Engineer). Private communication. September 2012.
- [102] C. W. Allen. *Astrophysical Quantities (3rd. edition)*. The Athlone Press, 1973.
- [103] C. D. Arvanitis *et al.* Empirical electro-optical and x-ray performance evaluation of CMOS active pixels sensor for low dose, high resolution x-ray medical imaging. *Med. Phys.*, 34:4612–4625, 2007.
- [104] A. Blue *et al.* Optical and electrical characterization of a back-thinned CMOS active pixel sensor. *Nucl. Instrum. Meth. A*, 604:215–217, 2009.
- [105] Microsemi Corp. *IGLOO Low Power Flash FPGAs with Flash*Freeze Technology*. Component Datasheet, 2012.
- [106] Actel Corp. *Radiation-Hardened FPGAs v3.1*. Component Datasheet, 2005.

- [107] Bosch. *CAN Specification Version 2.0*. Robert Bosch GmbH, Postfach 50, D-7000 Stuttgart 1, 1991.
- [108] E. F. Mitchell *et al.* HMRM physics algorithm. Technical Report HMRM-IC-TN-063, Imperial College London, 2011.
- [109] Consultative Committee for Space Data Systems. *Recommendation for space data system standards: Packet telemetry*. CCSDS 102.0-B-5, 2000.
- [110] C. Teyssier *et al.* Performance of the Medipix and Timepix devices for the recognition of electron-gamma radiation fields. *Nucl. Instrum. Meth. A*, 650:92–97, 2011.
- [111] V. Ivanchenko *et al.* Recent improvements in Geant4 electromagnetic physics models and interfaces. *Prog. Nucl. Sci. Tech.*, 2:898–903, 2011.
- [112] J. Apostolakis *et al.* Validation and verification of Geant4 standard electromagnetic physics. *J. Phys.: Conf. Ser.*, 219:1742–6596, 2010.
- [113] J. Allison *et al.* New Geant4 model and interface developments for improved space electron transport simulations: first results. *RADECS proceedings*, PA-19:115–118, 2011.
- [114] Geant4 collaboration. Geant4 User’s Guide for Application Developers. *Version: Geant4 9.5.0*, 2011.
- [115] H. Araújo *et al.* Space science simulations using Geant4 — LISA application. Technical report, Imperial College, 2004.
- [116] J. Apostolakis *et al.* Geometry and physics of the Geant4 toolkit for high and medium energy applications. *Radiat. Phys. Chem.*, 78:859–873, 2009.
- [117] Geant4 collaboration. Physics reference manual. *Version: Geant4 9.5.0*, 2011.
- [118] A. Darlinski. *Phys. Stat. Sol.*, 63:663, 1981.
- [119] G. R. Massoumi *et al.* Positron and electron backscattering from solids. *Phys. Rev. Lett.*, 68:3873–3876, 1992.
- [120] G. R. Massoumi *et al.* Electron and positron backscattering in the medium-energy range. *Phys. Rev. B*, 47:11007–11018, 1993.
- [121] W. N. Lennard *et al.* Stopping powers for positrons and electrons. *Phys. Rev. Lett.*, 74:3947–3950, 1995.
- [122] K. O. Al-Ahmad and D. E. Watt. Stopping powers and extrapolated ranges for electrons (1–10 keV) in metals. *J. Phys. D: Appl. Phys.*, 16:2257–2267, 1983.
- [123] C. J. Sofield *et al.* Energy straggling of 5.486 MeV alpha particles in Al. *Phys. Rev. A*, 15:2221–2226, 1977.
- [124] D. I. Ibrahim and M. B. Al-Bedri. Energy loss straggling of 5.486 MeV alpha particles in Melinex, Al, Ni and Cu. *Appl. Radiat. Isotopes*, 70:617–619, 2012.
- [125] B. Gottschalk *et al.* Multiple Coulomb scattering of 160 MeV protons. *Nucl. Instrum. Meth. B*, 74:467–490, 1993.

- [126] G. J. Lockwood *et al.* SAND79-0414.UC-34a: Calorimetric Measurement of Electron Energy Deposition in Extended Media - Theory vs Experiment. Technical report, Sandia National Laboratories, 1987.
- [127] NSRL. Bragg peak measurements at NSRL. *BNL*, <http://www.bnl.gov/medical/NASA/CAD/Bragg/Bragg.asp>.
- [128] S. Meroli *et al.* Energy loss measurement for charged particles in very thin silicon layers. *JINST*, 6:P06013, 2011.
- [129] K. Nagata *et al.* Energy loss and straggling of high-energy electrons in silicon detectors. *Jap. J. Appl. Phys.*, 14:697, 1975.
- [130] H. Bichsel. Straggling in thin silicon detectors. *Rev. Mod. Phys.*, 60:663, 1988.
- [131] S. Hancock *et al.* Energy loss and energy straggling of protons and pions in the momentum range 0.7 to 115 GeV/c. *Phys. Rev. A*, 28:615–620, 1983.
- [132] J. J. Kolata *et al.* Energy-Loss Straggling of Protons in Silicon. *Phys. Rev.*, 176:484, 1968.
- [133] D. W. Aitken *et al.* Energy loss and straggling in silicon by high-energy electrons, positive pions, and protons. *Phys. Rev.*, 179:393–398, 1969.
- [134] H. D. Maccabee *et al.* Fluctuations of energy loss by heavy charged particles in thin absorbers. *Phys. Rev.*, 165:469, 1968.
- [135] H. Esbensen *et al.* Random and channeled energy loss in thin germanium and silicon crystals for positive and negative 2–15 GeV/c pions, kaons, and protons. *Phys. Rev. B*, 18:1039, 1978.
- [136] G. De Geronimo *et al.* A novel position and time sensing active pixel sensor with field-assisted electron collection for charged particle tracking and electron microscopy. *Nucl. Instrum. Meth. A*, 568:167–175, 2006.
- [137] R. Turchetta *et al.* Accelerated particle and high energy radiation sensor. *US patent application 10/556,028*, May 2004.
- [138] M. D. Mikhailov and M. N. Ozisik. *Unified analysis and solutions of heat and mass diffusion*. Dover Publications, Inc., 1984.
- [139] MATLAB. *version 7.12.0.635 (R2011a)*. The MathWorks Inc., Massachusetts, USA, 2011.
- [140] G. Deptuch *et al.* Simulation and measurements of charge collection in monolithic active pixel sensors. *Nucl. Instrum. Meth. A*, 465:92–100, 2001.
- [141] C. Jacoboni *et al.* A review of some charge transport properties of silicon. *Solid state electron.*, 20:77–89, 1977.
- [142] M. A. Green. Intrinsic concentration, effective densities of states, and effective mass in silicon. *J. Appl. Phys.*, 67:2944–2954, 1990.

- [143] R. Turchetta *et al.* A monolithic active pixel sensor for charged particle tracking and imaging using standard VLSI CMOS technology. *Nucl. Instrum. Meth. A*, 458:677–689, 2001.
- [144] L. Servoli *et al.* Use of a standard CMOS imager as position detector for charged particles. *Nucl. Instrum. Meth. B*, 215:228–231, 2011.
- [145] L. Jungermann. *Space-Qualified Electronics for the AMS02 Experiment and Medical Radiation Imaging*. PhD thesis, University of Karlsruhe, Germany, 2005.
- [146] Synopsys. Sentaurus device simulator. <http://www.synopsys.com/Tools/T-CAD/DeviceSimulation/Pages/SentaurusDevice.aspx>.
- [147] Silvaco. Atlas device simulator. <http://www.silvaco.com/products/device-simulation/atlas.html>.
- [148] P. Buhler. The Radiation Environment Monitor: scientific data extraction. Technical Report 11108/94/N1/JG(SC), Paul Scherrer Institut, 1996.
- [149] T. Michel *et al.* Reconstruction of X-ray spectra with the energy sensitive photon counting detector Medipix2. *Nucl. Instrum. Meth. A*, 598:510–514, 2009.
- [150] C. Ponchut. Correction of the charge sharing in photon-counting pixel detector data. *Nucl. Instrum. Meth. A*, 591:311–313, 2008.
- [151] E. F. Mitchell and H. Araújo. HMRM spectral reconstruction using matrix method. Technical Report HMRM-IC-TN-051, Imperial College London, 2010.
- [152] E. F. Mitchell and H. Araújo. HMRM maximum likelihood spectral reconstruction. Technical Report HMRM-IC-TN-057, Imperial College London, 2010.
- [153] J. A. Nelder and R. Mead. A simplex method for function minimization. *Comput. J.*, 7:308–313, 1965.
- [154] W. S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Am. Stat. Assoc.*, 74:829–836, 1979.
- [155] G. Cowan. *Statistical data analysis*. Oxford science publications, 1998.
- [156] M. Siegl *et al.* Inner Belt Anisotropy Investigations Based on the Standard Radiation Environment Monitor (SREM). *IEEE Trans. Nuc. Sci.*, 57:2017–2023, 2010.
- [157] N. Guerrini. ESA HMRM User Manual v2.0. *STFC - RAL*, 2012.
- [158] S. Kleinfelder and M. Ahooie. Integrated sensors for charged-particle imaging using per-pixel correlated double sampling. *IEEE Trans. Nuc. Sci.*, 56:1056–1061, 2009.
- [159] J. R. Janesik. *Photon Transfer*. SPIE Press, 2007.
- [160] K. K. Hung *et al.* Flicker noise characteristics of advanced MOS technologies. *IEEE Dev. Mod.*, pages 34–37, 1998.
- [161] E. F. Mitchell. HMRM ASIC calibration plan. Technical Report HMRM-IC-TN-092, Imperial College London, 2011.

Appendix A

List of acronyms

ADC	Analogue to digital converter
ASIC	Application specific integrated circuit
CAD	Computer assisted design
CAN	Controller area network
CCD	Charge coupled device
CCSDS	Consultative Committee for Space Data Systems
CDS	Correlated double sampling
CME	Coronal mass ejection
CMOS	Complementary metal-oxide-semiconductor
CSDA	Continuous slowing-down approximation
CTE	Charge transfer efficiency
DAC	Digital to analogue converter
DN	Digital number
DNA	Deoxyribonucleic Acid
EA	Electrostatic analyser
EM	Electromagnetic
ENC	Equivalent noise charge
ESA	European Space Agency
FDN	Floating diffusion node
FPGA	Field-programmable gate array
FPN	Fixed pattern noise
FWHM	Full-width at half-maximum
GCR	Galactic cosmic ray
GEO	Geostationary orbit
GM	Geiger-Muller
HMRM	Highly Miniaturised Radiation Monitor
ID	Identification
ISS	International Space Station
LEO	Low Earth orbit
LET	Linear energy transfer

(M)APS	(Monolithic) active pixel sensor
MCP	Microchannel plate
MEO	Medium Earth orbit
MIP	Minimum ionising particle
MOSFET	Metal-oxide-semiconductor field effect transistor
NASA	National Aeronautics and Space Administration
NIEL	Non-ionising energy loss
NMOS	N-channel MOSFET
PCB	Printed circuit board
PMOS	P-channel MOSFET
PMT	Photomultiplier tube
PPS	Passive pixel sensor
RAL	Rutherford Appleton Laboratory
RAM	Random access memory
RBE	Relative biological effectiveness
RICH	Ring-imaging Cherenkov
RMS	Root mean square
SAA	South Atlantic Anomaly
SEP	Solar energetic particle
SEU	Single event upset
SSO	Sun-synchronous orbit
STFC	Science and Technology Facilities Council
TID	Total ionising dose
TLD	Thermoluminescent dosimeter
TOF	Time-of-flight
TRD	Transition radiation detector
UV	Ultraviolet
VHDL	Very high speed hardware description language

Appendix B

HMRM ASIC high-level design

Figure B.1 shows the high-level design of the HMRM sensor ASIC.

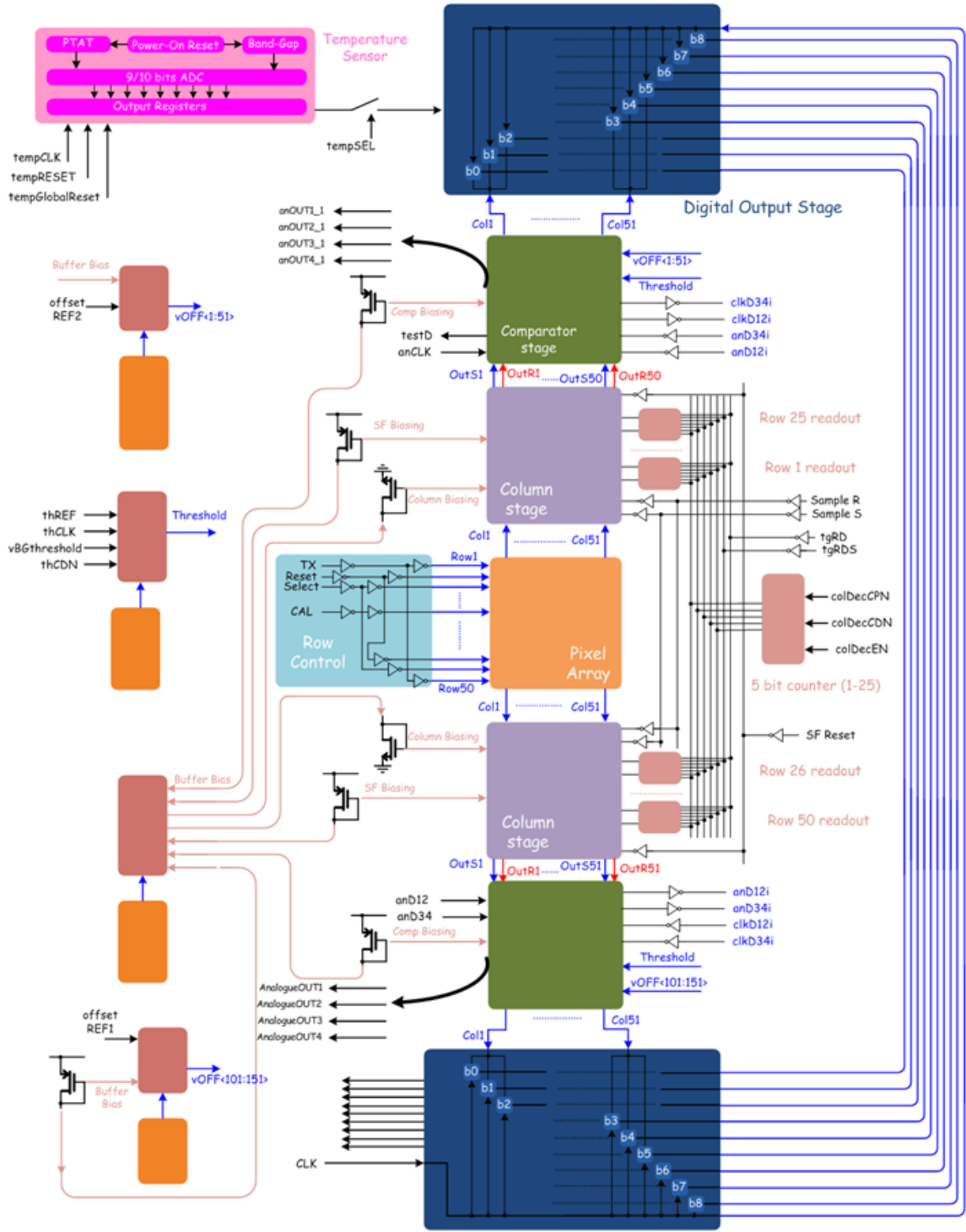


Figure B.1: HMRM ASIC high-level design (courtesy N. Guerrini, RAL Technology).

Appendix C

Geometry optimisation results

Table C.1 shows the changes to the baseline geometry (run 1) which were tested in each optimisation run discussed in Chapter 3. Tables C.2 and C.3 give the efficiency and purity results for protons and electrons, respectively.

Table C.1: Details of the geometry variations tested in each geometry optimisation run.

Run	Casing			Degradar (D3 — D4)		Back shield		Aperture	
	Material	Thickness / μm		Material	Thickness / μm	Material	Thickness / μm	Definition	Shape
		Front	Lateral						
1	Ti	2000	2000	Ti	2000	Ti	3000	C2	Pyramidal
2	Al	3360	3360	Al	3360	Al	5040	C2	Pyramidal
3	Ti	2000	2000	Mo	890	Ti	3000	C2	Pyramidal
4	Ti	2000	2000	Ti	2000	Ti	3000	C2	Conical
5	Ti	2000	2000	Ti	2000	Ti	3000	C3	Pyramidal
6	Ti	2000	1000	Ti	2000	Ti	3000	C2	Pyramidal
7	Ti	2000	2000	Ti	2000	Al + Ti	1000 + 2000	C2	Pyramidal

Table C.2: The identification purity (red) and efficiency (blue) percentages for each proton primary energy range as a result of the optimisation geometry variations.

Run	Proton energy range / MeV													
	1.2 — 2.8		2.8 — 5.1		5.1 — 9.1		9.1 — 17		17 — 33		33 — 100		100 — 300	
1	58.0	80.3	71.9	35.9	20.3	6.4	72.9	10.9	33.4	10.8	52.7	10.7	49.7	15.6
2	57.9	81.5	65.1	34.2	22.3	38.7	70.2	11.5	37.6	11.0	54.8	11.1	51.5	15.4
3	73.2	73.5	89.8	30.7	46.3	75.3	94.4	10.2	63.3	17.0	22.9	5.5	23.3	7.6
4	60.0	87.5	66.0	38.5	21.9	43.2	75.3	13.2	34.1	12.4	52.3	11.2	50.8	16.2
5	83.9	66.9	90.3	39.1	44.5	78.6	91.9	16.1	61.6	20.5	1.8	6.4	24.9	7.2
6	73.5	74.0	89.8	30.5	44.1	74.7	95.5	10.3	65.4	16.3	24.3	5.5	20.5	6.7
7	74.4	73.6	88.8	30.0	46.3	74.5	96.8	9.8	63.8	15.4	25.4	6.3	23.4	7.9
σ	0.27	0.34	0.41	0.23	0.21	0.34	0.45	0.17	0.71	0.31	0.45	0.14	0.35	0.13

Table C.3: The identification purity (red) and efficiency (blue) percentages for each electron primary energy range as a result of the optimisation geometry variations.

Run	Electron energy range / MeV													
	0.05 — 0.09		0.09 — 0.15		0.15 — 0.28		0.28 — 0.51		0.51 — 1.3		1.3 — 3.3		3.3 — 6.0	
1	6.9	47.6	2.9	5.7	6.9	18.6	11.6	11.9	9.8	2.0	28.7	7.9	13.5	21.2
2	7.0	48.5	2.9	5.8	6.2	17.7	11.5	11.9	7.5	1.3	27.9	9.1	13.4	20.9
3	8.0	36.3	9.7	14.6	8.3	20.1	3.3	11.3	24.8	7.1	31.2	21.7	10.3	12.1
4	6.7	46.9	3.0	6.2	7.1	20.7	11.5	11.0	13.5	1.9	31.9	8.6	12.9	19.7
5	6.7	37.6	11.1	18.3	5.6	18.2	11.7	12.2	23.2	8.2	30.0	24.5	11.3	12.5
6	7.3	36.1	9.2	15.3	8.0	20.1	13.2	11.2	22.7	6.6	31.8	20.7	9.7	11.1
7	7.4	34.3	8.8	13.8	8.4	20.1	14.5	1.2	22.8	6.5	30.3	20.8	9.4	12.0
σ	0.09	0.44	0.24	0.42	0.11	0.29	0.39	0.33	0.49	0.15	0.29	0.25	0.18	0.22

Appendix D

HMRM physics algorithm flowchart

The HMRM physics algorithm is presented in flowchart form in Figs. D.1 to D.5, abstracted from the implementation in the FPGA. All variables and parameters are listed and described in Tables D.1 and D.2. The following assumptions concerning the organisation of the system have been made:

1. All processes shown are executed within the HMRM FPGA; spacecraft or ground segment processes are not included.
2. Although the method of implementation in the FPGA is unspecified, the equations have been simplified where necessary and were successfully translated into VHDL code without significant alteration.
3. All input data tables are treated symbolically and specific values for these may be altered and uploaded via telecommand.

Figure D.1 shows the algorithm main function from which all analysis subroutines are called. This illustrates how the readout and algorithm execution operate on a continuous loop, with a fixed time period (nominally 100 μ s). The analysis functions have been split into four subroutines concerned with pixel processing, count rate measurements, dose measurements, and event identification; these are shown in Figs. D.2, D.3, D.4 and D.5, respectively.

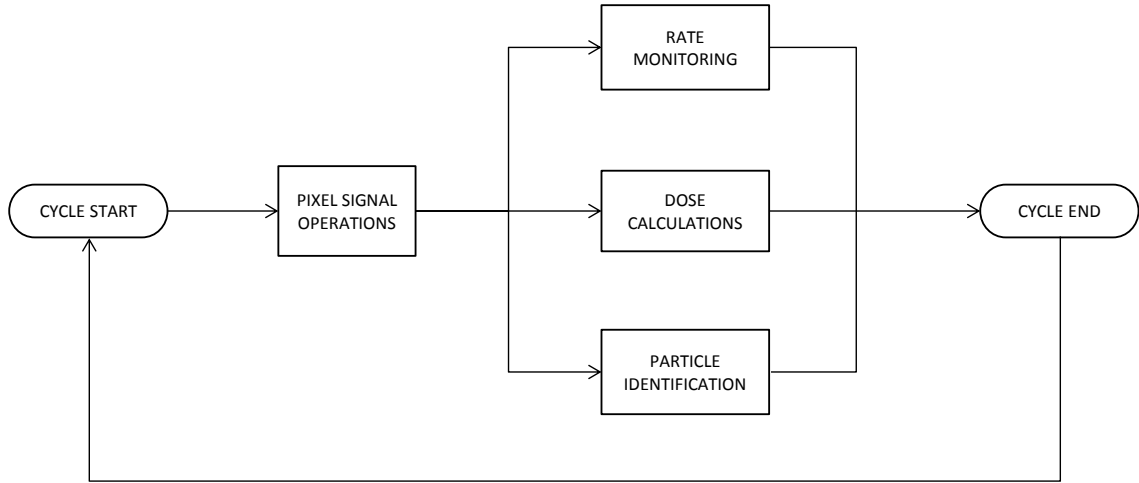


Figure D.1: FPGA physics algorithm ‘main function’ showing the ordering of the various analysis subroutines. The cycle period is 100 μ s.

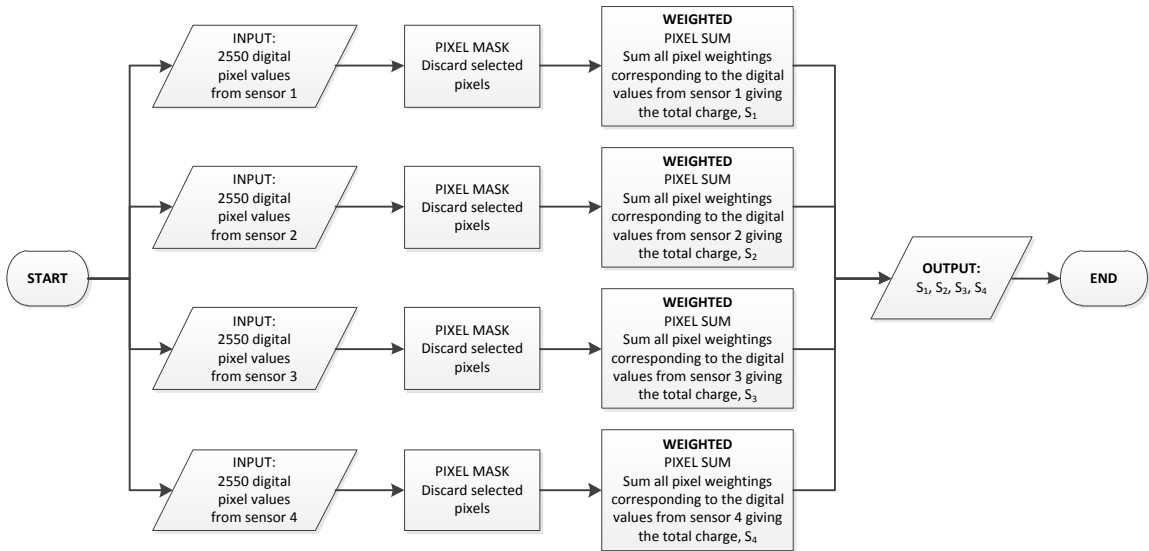


Figure D.2: FPGA physics algorithm: pixel signal operations.

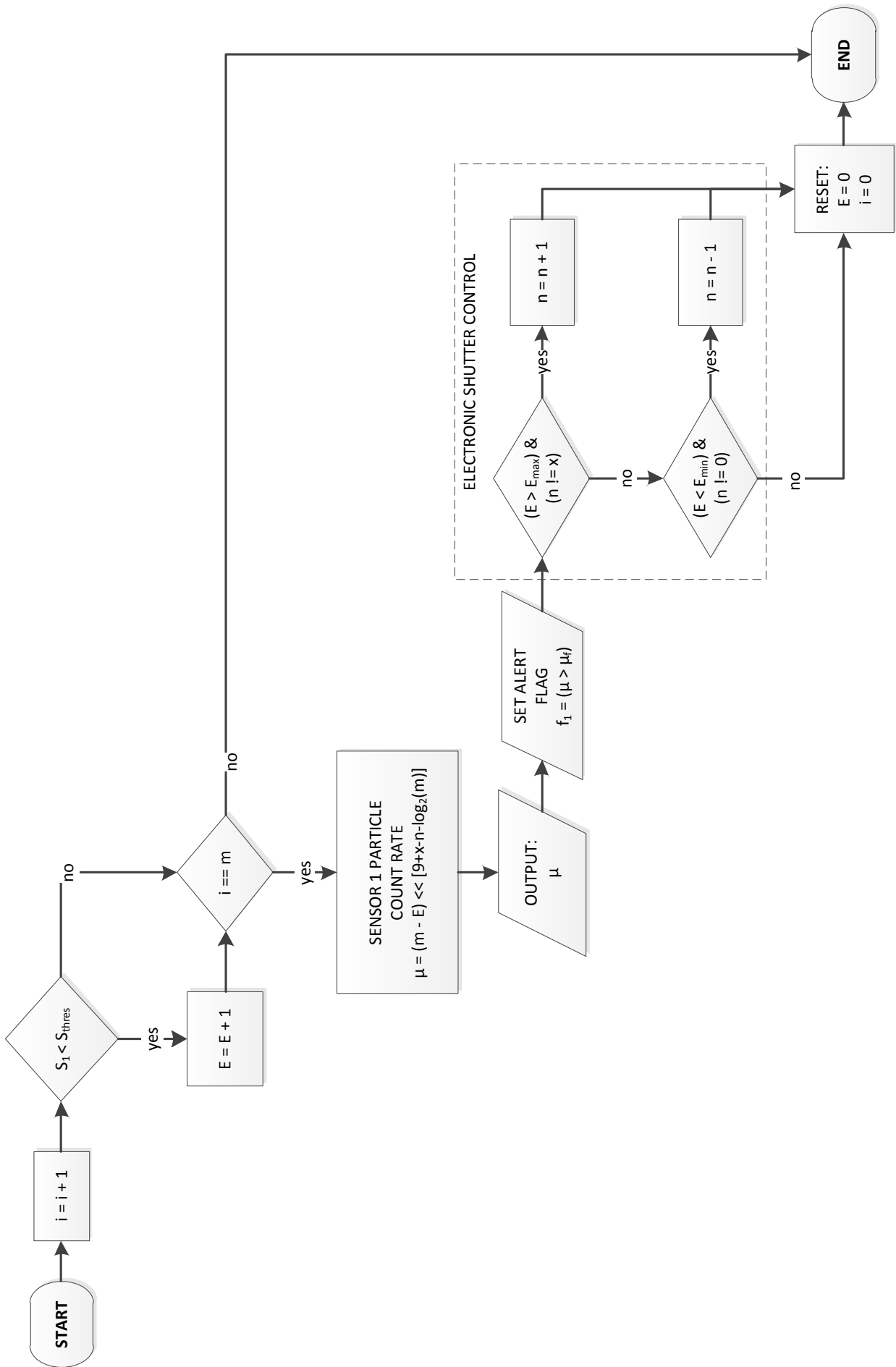


Figure D.3: FPGA physics algorithm: rate monitoring subroutine.

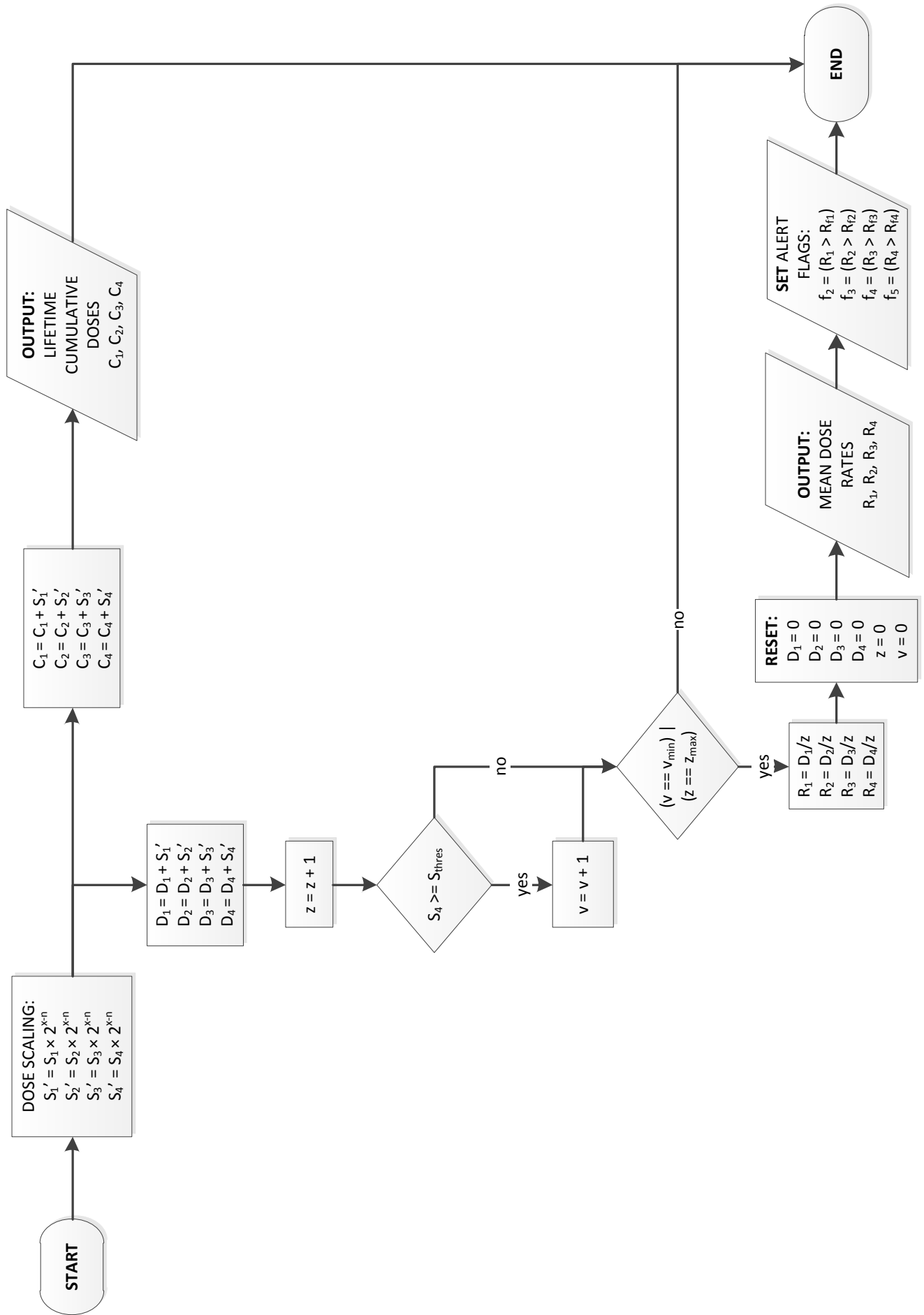


Figure D.4: FPGA physics algorithm: dose calculation subroutine.

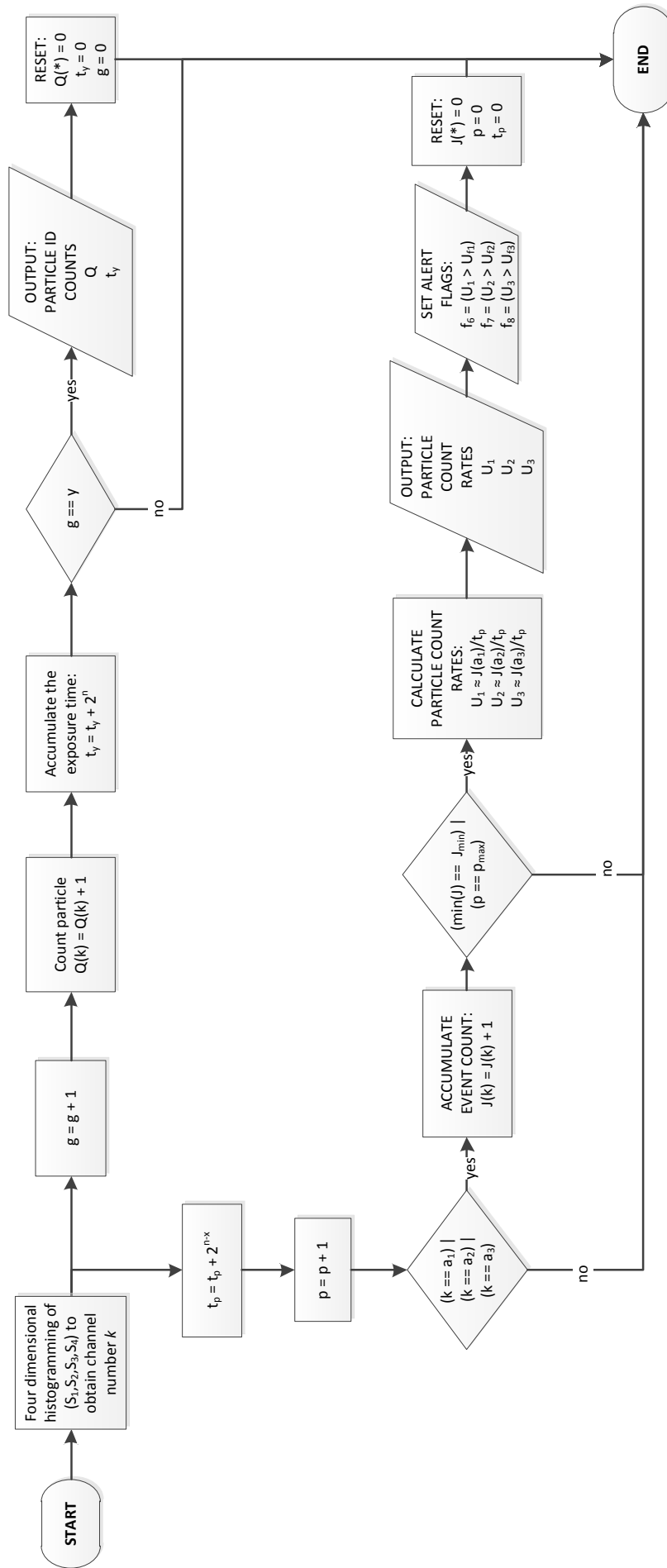


Figure D.5: FPGA physics algorithm: event identification subroutine.

Table D.1: Externally specified parameters required during the physics algorithm; these are uploaded by telecommand.

Label	Description
m	The shuttering control and particle count rate integration period (integer number of cycles)
y	The event identification count integration period (integer number of cycles)
H	The ID table, a 31×8 array of values giving the 4-D histogram bin edges
S_{thres}	Threshold sensor energy deposit value for indicating a particle hit
a_1, a_2, a_3	Three integers indicating the channel numbers for which particle count rates are calculated
E_{min}, E_{max}	The minimum and maximum number of empty readout frames for sensor 1, out of a series of m
p_{max}	The maximum allowable identified particle rate integration period (integer number of cycles)
J_{min}	The minimum preferred number of identified particle counts for rate calculation
z_{max}	The maximum allowable dose rate integration period (integer number of cycles)
v_{min}	The minimum preferred number of events for dose rate calculation
x	The maximum shutter index (nominally, $x=5$)
μ_f	Alert flag threshold value corresponding to the particle count rate
R_{f1-4}	Alert flag threshold value corresponding to the sensor dose rates
U_{f1-4}	Alert flag threshold value corresponding to the identified particle count rates

Table D.2: Internally generated variables used during the physics algorithm.

Label	Description
i	Loop counter: incremented on each sensor readout (counts up to m)
g	Loop counter: incremented on each sensor readout (counts up to y)
S_{1-4}	Reconstructed energy deposit for each sensor 1-4
E	The number of empty frames for sensor 1 (in a series of m frames)
n	Integer shutter index, controlling charge integration time: $0 \leq n \leq x$
S'_{1-4}	Reconstructed sensor energy deposit, scaled to that expected when $n=x$
C_{1-4}	Lifetime cumulative dose for each sensor 1-4
z	Dose rate integration period (integer number of cycles)
v	The number of particle counts on sensor 4 in period z
D_{1-4}	Scaled dose, accumulated over z cycles, for each sensor 1-4
R_{1-4}	Mean dose rate for each sensor 1-4
f_{1-4}	Eight alert flag bits, derived from the eight measured rates
μ	Mean particle count rate for sensor 1
Q	Array of 32 event identification counts, accumulated over y cycles
p	Identified particle rate integration period (integer number of cycles)
$J(a)$	Array of identified particle counts, accumulated over p cycles (for channels a_1, a_2, a_3)
U_{1-4}	Identified particle count rates for the three selected channels: a_1, a_2, a_3
k	Channel number obtained in the 4-D histogramming of (S_1, S_2, S_3, S_4) : $1 \leq k \leq 32$
t_y	Cumulative sum of the exposure time during period y
t_p	Cumulative sum of the exposure time during period p

Appendix E

Orbit simulation ID table

Table E.1: ID table used for the simulation of reference orbit A presented in Chapter 5.

Channel	ID table entry / keV								Mode
	Sensor 1		Sensor 2		Sensor 3		Sensor 4		
	L1	U1	L2	U2	L3	U3	L4	U4	
1	0.0	0.2	0.0	0.2	0.0	0.2	0.0	0.2	—
2	361.0	624.2	23.4	208.8	0.0	0.2	0.0	0.2	2
3	69.8	208.8	69.8	208.7	208.8	624.2	0.0	0.2	3
4	40.4	120.7	40.4	120.7	69.8	120.7	0.0	0.2	3
5	0.0	0.2	69.8	120.7	69.8	120.7	0.0	0.2	3
6	23.4	69.8	23.4	69.8	23.4	40.4	0.0	0.2	3
7	0.0	0.2	0.0	0.2	7.3	21.4	13.5	69.8	4
8	4.2	6.6	4.2	6.6	4.1	11.9	1.5	13.5	4
9	4.5	13.5	1.5	2.6	0.0	0.2	0.0	0.2	2
10	1.5	2.6	1.5	4.5	0.0	0.2	0.0	0.2	2
11	2.6	4.5	1.5	2.6	0.0	0.2	0.0	0.2	2
12	1.5	2.6	1.5	2.6	2.6	4.5	0.0	0.2	3
13	0.5	1.5	1.5	2.6	1.5	2.6	0.0	0.2	3
14	1.5	7.8	1.5	2.6	0.9	1.5	0.0	0.2	3
15	361.0	624.2	0.0	0.2	0.0	0.2	0.0	0.2	1
16	13.5	23.4	0.0	0.2	0.0	0.2	0.0	0.2	1
17	120.7	208.8	120.8	208.8	0.0	0.2	0.0	0.2	2
18	7.8	13.5	0.0	0.2	0.0	0.2	0.0	0.2	1
19	4.5	7.8	0.0	0.2	0.0	0.2	0.0	0.2	1
20	23.4	40.4	0.0	0.2	0.0	0.2	0.0	0.2	1
21	2.6	4.5	0.0	0.2	0.0	0.2	0.0	0.2	1
22	1.5	2.6	0.0	0.2	0.0	0.2	0.0	0.2	1
23	0.9	1.5	0.0	0.2	0.0	0.2	0.0	0.2	1
24	0.5	0.9	0.0	0.2	0.0	0.2	0.0	0.2	1
25	0.3	0.5	0.0	0.2	0.0	0.2	0.0	0.2	1
26	40.4	69.8	0.0	0.2	0.0	0.2	0.0	0.2	1
27	0.0	0.2	4.5	7.8	0.0	0.2	0.0	0.2	2
28	208.8	361.0	361.1	624.3	0.0	0.2	0.0	0.2	2
29	2.6	4.5	2.6	4.5	0.0	0.2	0.0	0.2	2
30	40.4	69.8	40.4	69.8	0.0	0.2	0.0	0.2	2
31	69.8	120.7	69.8	120.8	0.0	0.2	0.0	0.2	2
32	Other events								

Appendix F

Reconstructions of mean particle spectra in orbits B, C and E

Figures F.3 to F.6 show reconstructed particle energy spectra for the mean AP-8/AE-8 modelled fluxes in reference orbits B, C and E over simulated periods of order 100 s. Orbits A and D are presented in Chapter 6.

Orbit B: 400–4,000 km, 83° inclination, highly elliptical orbit

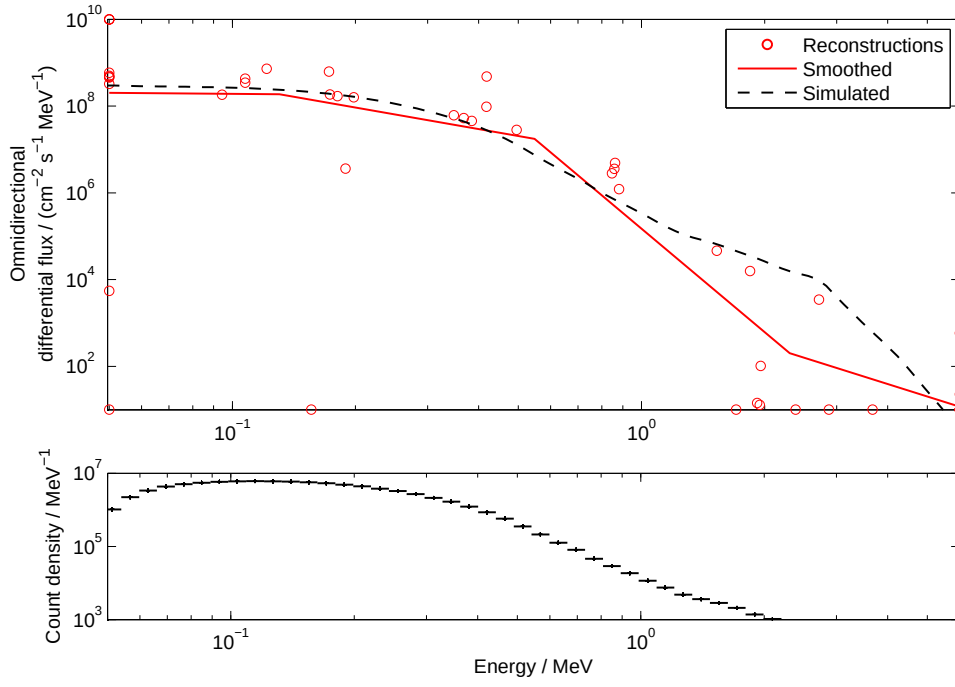


Figure F.1: Electron spectrum reconstruction for a 592 s simulated exposure in reference orbit B with 5 spectrum nodes per species.

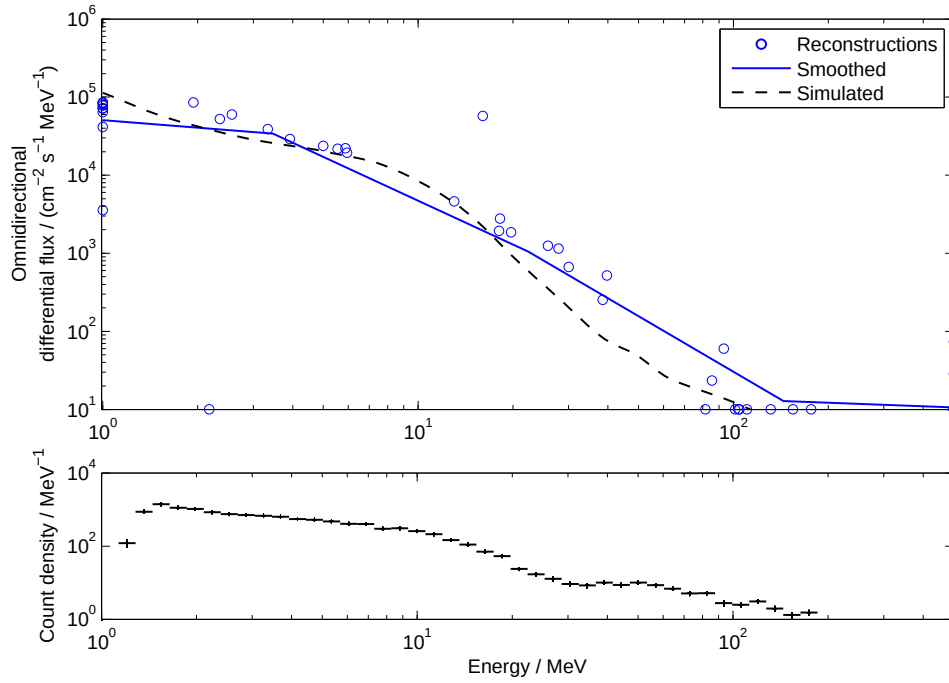


Figure F.2: Proton spectrum reconstruction for a 592 s simulated exposure in reference orbit B with 5 spectrum nodes per species.

Orbit C: 10,000 km, 0° inclination, medium Earth orbit

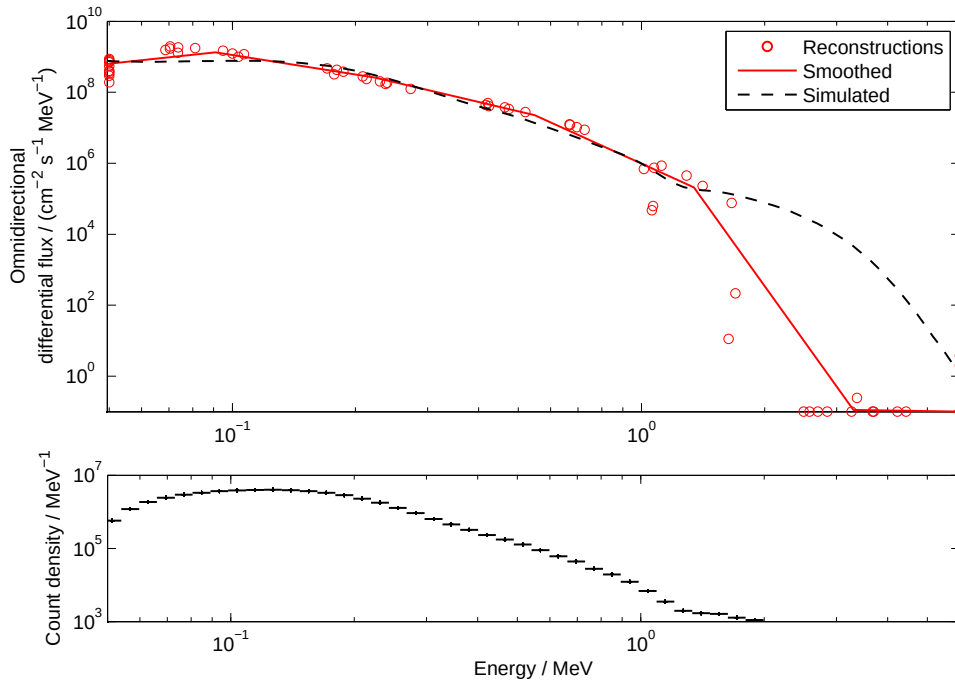


Figure F.3: Electron spectrum reconstruction for a 129 s simulated exposure in reference orbit C with seven spectrum nodes per species.

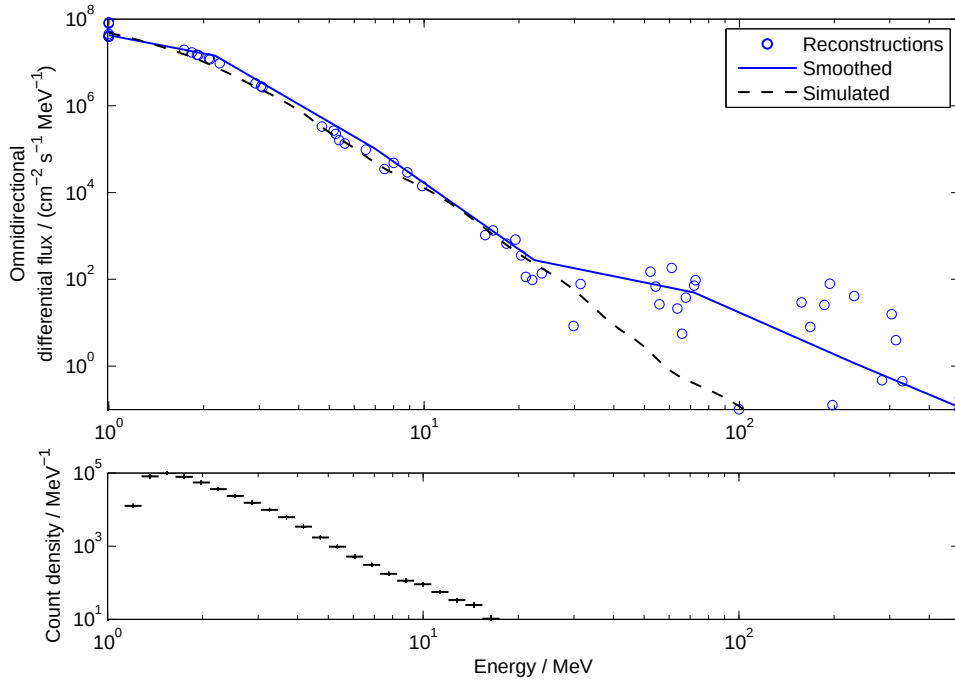


Figure F.4: Proton spectrum reconstruction for a 129 s simulated exposure in reference orbit C with seven spectrum nodes per species.

Orbit E: 35,786 km geostationary orbit

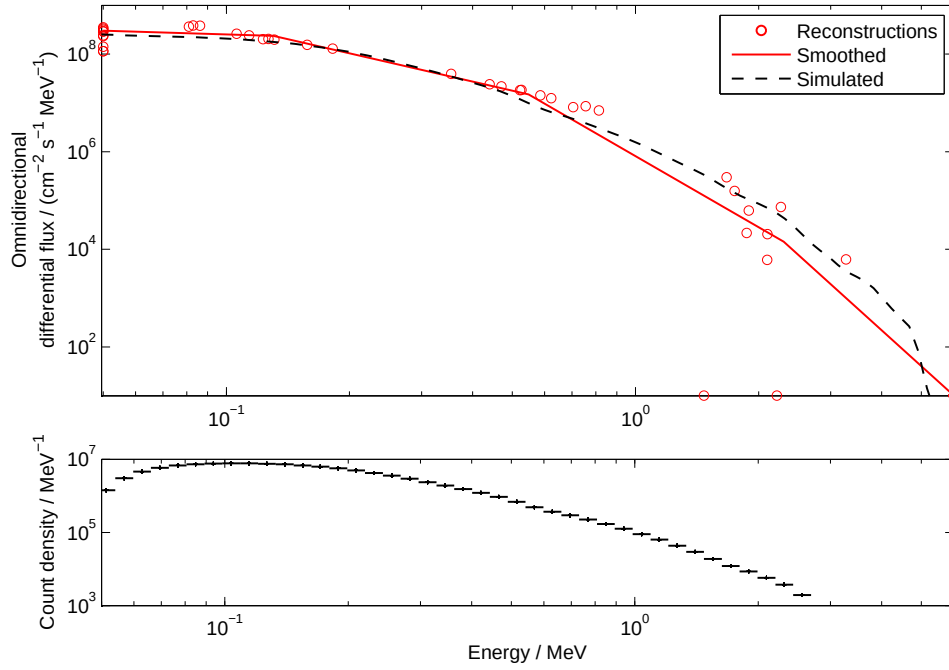


Figure F.5: Electron spectrum reconstruction for a 981 s simulated exposure in reference orbit E with five spectrum nodes per species.

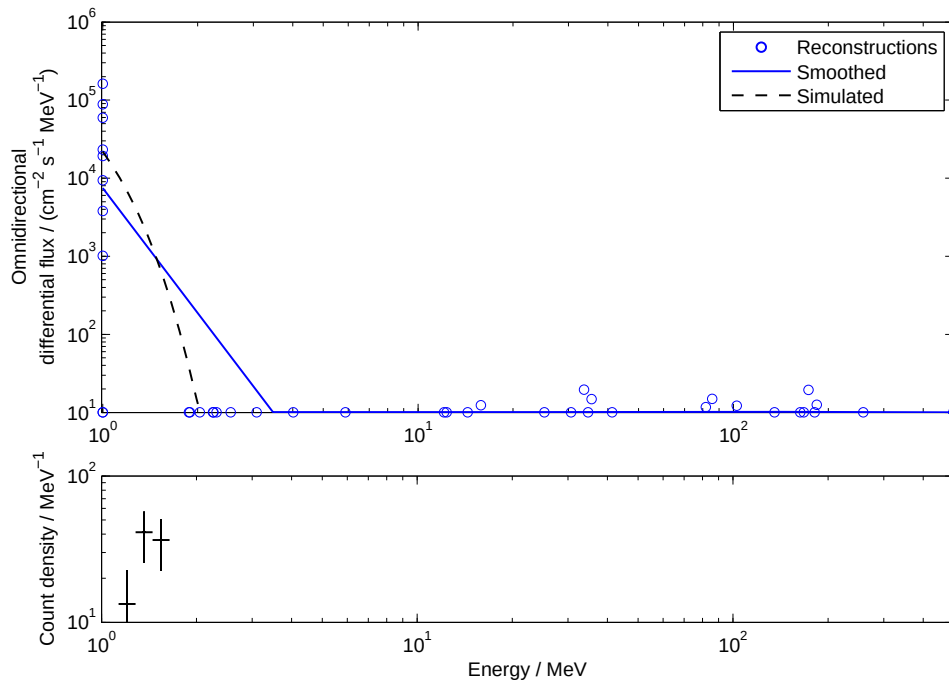


Figure F.6: Proton spectrum reconstruction for a 981 s simulated exposure in reference orbit E with five spectrum nodes per species.