8 July 2026
casey.niel.farren-colloty@cern.ch

Fast Monte Carlo Simulations for the MALTA Pixel Sensor

Casey Farren-Colloty
CERN, CH-1211 Geneva, Switzerland

Keywords: DMAPS, Monte Carlo simulations, MALTA

Summary

This document presents the development and implementation of Malta2FastSim, a fast Monte Carlo simulation module for MALTA (Monolithic Active Pixel Sensor for ATLAS) pixel detectors within the Allpix^2 framework. MALTA sensors represent a significant advancement in monolithic CMOS detector technology, featuring small collection electrodes (3 μm diameter) and asynchronous readout architectures designed for operation in the harsh radiation environment like those of the High Luminosity Large Hadron Collider (HL-LHC).

The Malta2FastSim module addresses a gap in MALTA detector simulation capabilities by preserving the validated physics models established by the standalone PyMaltaSim implementation while enabling seamless integration within modern detector simulation workflows. The module implements comprehensive physics models including charge-dependent timing delays, position-dependent charge collection efficiency, and realistic threshold calculations based on Digital-to-Analog Converter (DAC) settings. Two key algorithmic innovations distinguish this work: a pixel-grouped accumulation algorithm that solves fundamental efficiency problems in high-resolution trajectory sampling, and selective Highland multiple scattering implementation that ensures realistic cluster modelling for perpendicular particle trajectories.

Rather than applying threshold tests to individual trajectory steps, the pixel-grouped accumulation algorithm accumulates charge within each pixel before threshold comparison, maintaining detection efficiency while enabling arbitrary trajectory resolution. The selective Highland scattering implementation addresses cluster modelling limitations in test beam configurations, where perpendicular particle trajectories would otherwise produce unrealistic single-pixel hits.

The mathematical foundations underlying MALTA detector operation are comprehensively explored, including detailed derivations of the drift-diffusion equations governing charge transport in depleted silicon, electric field distributions for small-electrode geometries, and Ramo's theorem for signal induction. The appendix provides complete derivations of the Bethe-Bloch formula from first principles, the Landau distribution for energy loss fluctuations, and the Highland formula for multiple scattering, establishing the theoretical framework necessary for accurate detector simulation.

Key technical points include complete physics model preservation from PyMaltaSim, native Allpix^2 framework integration, algorithmic solutions to high-resolution sampling efficiency challenges, and comprehensive implementation of the complete detector readout chain from charge

deposition through digital output. The module represents a significant step forward in fast simulation capabilities for monolithic pixel detectors, enabling the detailed studies necessary for successful deployment in next-generation particle physics experiments.

Contents

1 Theoretical Foundation and Background

1.1 Semiconductor Physics for Particle Detection

To understand the operation of MALTA detectors, we must first establish the fundamental physics governing particle detection in silicon semiconductors. When high-energy charged particles traverse silicon, they lose energy through ionisation processes, creating electron-hole pairs that can be collected as measurable electrical signals.

1.1.1 Energy Loss Mechanisms in Silicon

High-energy charged particles interact with silicon atoms primarily through electromagnetic processes. For particles with sufficient rest mass-such as those encountered at the Super Proton Synchrotron (SPS) where MALTA operates-the energy loss can be described by the Bethe-Bloch formula:

$$-\frac{dE}{dx} = Kz^2\frac{Z}{A}\frac{1}{\beta^2}\left[\frac{1}{2}\ln\frac{2m_e c^2\beta^2\gamma^2 T_{max}}{I^2} - \beta^2\right] \tag{1}$$

where K is a constant, z is the charge of the incident particle, Z and A are the atomic number and mass of the target material, $\beta = v/c$, and I is the mean excitation energy of the target atoms.

In silicon, this energy loss results in the creation of electron-hole pairs, with approximately 3.6 eV required per pair. For minimum ionising particles, this translates to roughly 60 electron-hole pairs per micrometre of silicon traversed. The energy deposition process involves both primary ionisation-where the incident particle directly knocks electrons from silicon atoms-and secondary ionisation from high-energy δ-rays produced in the primary interactions.

1.1.2 Charge Transport in Depleted Silicon

Once created, the electron-hole pairs must be collected before they can recombine. In MALTA detectors, this collection occurs through drift-based mechanisms in a depleted silicon region. The fundamental relationship governing charge transport is:

$$v = \mu E \tag{2}$$

where v is the drift velocity, μ is the mobility, and E is the electric field strength. In silicon at room temperature, electron mobility is approximately $\mu_e \approx 1400$ cm^2/(V·s), while hole mobility is significantly lower at $\mu_h \approx 450$ cm^2/(V·s).

The creation of a depletion region is crucial for efficient charge collection. Through the application of reverse bias, we extend the space charge region where electric fields are present. This is particularly important given that ATLAS has stringent timing requirements-particle hits must be associated with bunch crossings that are only 25 ns apart. In standard sensors with partial depletion, charge generated outside the depletion region is collected by slow diffusion processes with collection times of approximately 100 ns, which would be inadequate for HL-LHC operation.

1.1.3 Signal Formation Theory

The process by which moving charges generate measurable electrical signals is described by Ramo's theorem:

$$i(t) = q\vec{E}_w(\vec{r}(t)) \cdot \vec{v}(t) \tag{3}$$

where $i(t)$ is the induced current, q is the charge, $\vec{E}_w$ is the weighting field, and $\vec{v}(t)$ is the velocity of the charge carrier.

The weighting field $\vec{E}_w$ is a purely geometric quantity that determines how effectively charge motion at a given position influences the electrode. Understanding this concept is essential for MALTA's small-electrode design:

- **High weighting**: Charge motion strongly influences the electrode signal
- **Low weighting**: Charge motion produces minimal electrode response
- **Zero weighting**: Charge motion produces no signal

For MALTA's small collection electrodes, the final signal amplitude is given by $V = Q/C$, where the capacitance C is kept below 5 fF. This low capacitance is crucial for achieving high signal-to-noise ratios, as it maximises the voltage step produced by collected charge.

1.2 DMAPS Technology and MALTA Architecture

Central to the innovative nature of MALTA is its implementation as a DMAPS-a Depleted Monolithic Active Pixel Sensor. To minimise technical jargon, let us examine this designation systematically:

- **Depleted**: The sensor operates with very few free charge carriers in the bulk silicon. This is achieved through high-resistivity substrates and appropriate biasing. From Ohm's law, $\vec{E} = \rho\vec{J}$, high resistivity ρ reduces current density $\vec{J}$ for a given electric field, minimising leakage current and noise.
- **Monolithic**: The entire detector is fabricated from a single piece of silicon. Both the sensor and front-end electronics exist on the same substrate, unlike hybrid sensors where these components are manufactured separately and joined through bump-bonding. This eliminates the inhomogeneities created at bump-bond interfaces that can disrupt charge flow and cause information loss.
- **Active**: Each pixel element contains amplification and signal processing circuitry, enabling local signal conditioning before readout.
- **Pixel**: The detector is segmented into discrete detection elements-MALTA contains a 512×512 matrix of such pixels, totalling over 260,000 individual sensors.

1.2.1 CMOS Process Technology

MALTA employs Complementary Metal-Oxide-Semiconductor (CMOS) technology, similar to that used in consumer electronics such as cameras and computers. However, rather than detecting visible photons, these sensors are optimised for charged particle detection. The use of commercial CMOS processes provides significant advantages in terms of cost, availability, and integration density.

The Tower Semiconductor 180 nm CMOS imaging process used for MALTA provides the foundation for implementing both sensor and readout electronics on the same substrate[2, 8]. This process has been modified specifically for particle detection applications, incorporating specialised implant geometries and doping profiles optimised for radiation hardness and charge collection efficiency.

1.2.2 Small Collection Electrode Philosophy

A fundamental design principle of MALTA is the use of small collection electrodes-hexagonal structures with diameters of approximately 3 μm positioned at the centre of each 36.4 $\times$ 36.4 μm^2 pixel. This design choice is driven by the relationship between electrode capacitance and signal amplitude.

The capacitance of the collection electrode determines the voltage step produced by collected charge:

$$V = \frac{Q}{C} \tag{4}$$

By minimising the electrode area, the capacitance is reduced to approximately 5 fF, maximising the signal amplitude for a given amount of collected charge. This approach enables the use of open-loop amplification schemes in the front-end electronics, simplifying the circuit design and reducing power consumption to approximately 1 μW per pixel.

1.2.3 Substrate Technologies

MALTA sensors are produced on two distinct substrate technologies, each offering specific advantages:

- **Epitaxial Substrates**: These consist of a high-resistivity epitaxial layer (typically 25-30 μm thick) grown on a heavily-doped substrate. For minimum ionising particles, these substrates generate approximately 1500 e$^-$ of signal charge.
- **Czochralski Substrates**: These employ thick (100-300 μm), high-resistivity p-type silicon substrates produced by the Czochralski crystal growth method. The increased thickness results in significantly enhanced signal generation, with collected charge exceeding 6000 e$^-$ for minimum ionising particles-more than a factor of four improvement over epitaxial substrates.

The enhanced signal amplitude from Czochralski substrates enables more robust operation, particularly in radiation-damaged environments where charge trapping reduces collection efficiency.

2 Introduction to MALTA Detectors

2.1 DMAPS in the Context of High-Energy Physics

Monolithic CMOS pixel sensors represent a significant departure from traditional hybrid pixel detector architectures. In hybrid systems, the sensor element and readout electronics are fabricated separately using optimised processes for each function, then joined through fine-pitch bump-bonding. While this approach has proven successful in applications such as the ATLAS Inner Tracker (ITk)[1], it comes with inherent limitations including high production costs, increased material budget, and constraints on pixel pitch.

DMAPS technology addresses these limitations by integrating both functions on a single silicon substrate. This integration offers several key advantages:

- **Reduced material budget**: Elimination of bump-bonds and separate readout chips reduces the probability of multiple scattering, improving track momentum resolution
- **Lower production costs**: Commercial CMOS processes reduce fabrication expenses compared to specialised hybrid approaches
- **Enhanced granularity**: Monolithic integration enables smaller pixel pitches, improving spatial resolution
- **Simplified assembly**: Direct integration eliminates complex bump-bonding procedures

2.2 MALTA Detector Architecture

The MALTA detector implements a sophisticated architecture designed to meet the demanding requirements of HL-LHC experiments. The fundamental specifications include operation as a tracking sensor with spatial resolution of approximately 10 μm while maintaining the ability to process hit rates exceeding 100 MHz/cm^2 at the LHC bunch crossing frequency of 40 MHz.

2.2.1 Physical Structure

The MALTA sensor consists of a 512×512 pixel matrix covering an active area of 18.3×18.3 mm^2. Each pixel contains a 3 μm diameter hexagonal collection electrode positioned at the pixel centre, surrounded by the necessary amplification and discrimination circuitry.

The readout architecture employs an asynchronous design that avoids distributing high-frequency clock signals across the pixel matrix. This approach minimises analog-digital crosstalk-a critical consideration for maintaining signal integrity in high-density pixel arrays. Pixels are organised in groups of 2×8, with hits from individual pixels sent to a common reference pulse generator within each group.

2.2.2 Timing and Spatial Organisation

The temporal structure of MALTA readout is designed around the 25 ns bunch crossing interval of the LHC. The asynchronous architecture generates reference pulses that are appended with pixel and group address information (16-bit and 5-bit respectively). These hits are distributed across two parallel 22-bit wide buses-one for even groups and another for odd groups. This separation ensures that adjacent groups cannot share the same bus, reducing crosstalk on the hit address bus.

The spatial grouping of pixels into 2×8 clusters reflects the underlying charge sharing characteristics expected in silicon sensors. When particles traverse the sensor at non-normal incidence or undergo multiple scattering, the deposited charge is often shared among neighbouring pixels, creating cluster patterns that provide information about the particle trajectory and energy deposition.

2.2.3 Process Modifications for Enhanced Performance

Standard CMOS imaging processes require modification to achieve the radiation hardness and charge collection efficiency needed for particle physics applications. MALTA incorporates two key process modifications developed through extensive TCAD simulation studies[7, 5]:

- **N-layer Gap (NGAP)**: This modification involves creating a gap in the low-dose n-type blanket implant at pixel edges. TCAD simulations demonstrate that this approach enhances lateral electric fields in pixel corners, regions that are typically problematic for charge collection efficiency.
- **Extra Deep P-Well (XDPW)**: An additional deep p-well implant is introduced around pixel edges to further enhance charge collection. This 4 μm wide implant creates stronger lateral fields that improve the drift of ionisation charge toward collection electrodes, particularly important for maintaining efficiency after radiation damage.

The electric field configurations resulting from these modifications are illustrated through TCAD simulations. In standard configurations with continuous n^- layers, electric field lines point predominantly vertical at pixel corners, creating potential minima that can trap charge carriers. The XDPW modification generates stronger lateral fields that enhance charge collection from these challenging regions.

2.3 Performance Characteristics and Radiation Hardness

MALTA detectors are designed to operate in the harsh radiation environment of HL-LHC experiments, where sensors must maintain performance after exposure to fluences exceeding 10^{15} 1-MeV neutron equivalent per cm^2 and total ionising doses of 100 Mrad.

2.3.1 Detection Efficiency and Signal Characteristics

Unirradiated MALTA sensors demonstrate excellent detection efficiency, with more than 98% of hits collected within the 25 ns time window required for bunch crossing association. The

timing resolution, measured using scintillator-based reference systems, achieves $\sigma_t \approx 1.7-2.1$ ns-well within the requirements for HL-LHC operation.

The signal amplitude varies significantly between substrate technologies. Epitaxial substrates generate approximately 1500 e^- for minimum ionising particles, while Czochralski substrates produce more than 6000 e^-. This enhanced signal amplitude provides improved robustness against radiation-induced charge trapping effects.

2.3.2 Radiation Damage Mechanisms

Silicon particle detectors face two primary forms of radiation damage:

- **Non-Ionising Energy Loss (NIEL)**: When incident particles transfer more than approximately 25 eV to silicon atoms, they can displace atoms from lattice sites, creating vacancy and interstitial defects. These defects act as charge traps, reducing collection efficiency and increasing leakage current proportional to radiation fluence.
- **Total Ionising Dose (TID)**: Ionising radiation creates positive charges at silicon-oxide interfaces, affecting the operation of CMOS electronics through threshold voltage shifts and increased surface leakage currents.

For MALTA applications, displacement damage represents the primary limitation. At HL-LHC fluence levels, charge carriers must drift hundreds of micrometers from generation sites to collection electrodes, with significant probability of encountering traps along their path.

2.3.3 Radiation Hardness Solutions

The combination of process modifications (NGAP and XDPW), Czochralski substrates, and optimised front-end electronics provides MALTA with enhanced radiation tolerance. Test beam measurements on irradiated samples demonstrate maintained detection efficiency at fluences up to 3×10^{15} 1-MeV n_{eq}/cm^2[8], approaching the requirements for the innermost tracking layers at HL-LHC.

The use of Czochralski substrates provides particular advantages in radiation-damaged environments. The enhanced signal amplitude helps overcome the effects of charge trapping, while the thick substrate geometry provides improved charge collection efficiency even when surface regions become heavily damaged.

Furthermore, the asynchronous readout architecture provides resilience against radiation-induced timing variations, as the system can accommodate hit time variations without requiring precise clock distribution across the pixel matrix.

2.4 Readout Architecture and Performance Characteristics

The MALTA detector implements a sophisticated readout system designed to handle the demanding requirements of HL-LHC operation. Let us examine the key architectural elements that enable high-rate operation while maintaining excellent timing performance.

2.4.1 Asynchronous Readout Design

Traditional pixel detectors often employ synchronous readout schemes that distribute clock signals across the pixel matrix. However, at the frequencies required for HL-LHC operation, this approach introduces significant challenges related to clock distribution, power consumption, and analog-digital crosstalk. MALTA addresses these limitations through an asynchronous architecture that eliminates the need for high-frequency clock distribution.

In the asynchronous approach, each pixel generates its own timing reference when a hit occurs. This local timing generation eliminates the complex clock distribution networks required in synchronous systems, while simultaneously reducing power consumption and minimising crosstalk between analog and digital circuits.

2.4.2 Spatial Organisation: The 2×8 Pixel Grouping

The physical organisation of MALTA pixels reflects both the underlying physics of charge sharing and practical considerations for high-rate readout. Pixels are arranged in groups of 2×8, with each group sharing common readout resources. This grouping scheme serves several important functions.

When charged particles traverse the detector at non-normal incidence or undergo scattering within the silicon substrate, the deposited charge is often shared among neighbouring pixels. The 2×8 grouping captures the typical cluster patterns expected from such charge sharing, ensuring that related hits are processed together efficiently.

From a readout perspective, hits within a group are sent to a common reference pulse generator. This pulse generator creates a timing reference that is appended with both pixel address information (16-bit) and group address information (5-bit), creating a complete spatial and temporal record of each hit.

2.4.3 Temporal Processing and Column Merging

The temporal structure of MALTA readout operates on multiple time scales to accommodate both the physics of signal formation and the practical constraints of data transmission. The fundamental time resolution is set at 25 ns, matching the LHC bunch crossing interval.

Within this framework, the system implements column merging with a characteristic spacing of 1.6 ns. This fine temporal resolution accounts for the time required for signals to propagate along the pixel columns, ensuring that hits occurring in rapid succession can be properly separated and identified.

The time-end-of-column (TEOC) calculation incorporates both the hit occurrence time and a position-dependent propagation delay:

$$t_{\text{TEOC}} = t_{\text{hit}} + \text{row} \times 0.0125 \text{ ns} \tag{5}$$

This calculation ensures that the spatial position of hits within columns is properly accounted for in the temporal processing.

2.4.4 Measured Performance Characteristics

Extensive test beam measurements have demonstrated MALTA's excellent performance characteristics. The timing resolution, measured using scintillator-based reference systems,

achieves $\sigma_t \approx 1.7 - 2.1$ ns. This resolution is well within the requirements for HL-LHC operation and enables effective association of hits with their originating bunch crossings.

Detection efficiency measurements show that more than 98% of hits are collected within the 25 ns time window required for bunch crossing association in non-irradiated sensors[10]. Even at 90% efficiency levels, the required time window reduces to approximately 8 ns, demonstrating the system's capability for high-precision timing applications.

The spatial resolution achieves approximately 10 μm, enabled by the combination of small pixel pitch (36.4 μm) and the charge sharing characteristics that provide sub-pixel position resolution through centroid calculations on cluster patterns.

3 Background: From PyMaltaSim to `Allpix`^{2} Integration

The development of fast simulation capabilities for MALTA detectors represents a fascinating example of how validated physics models can be adapted and enhanced for integration within modern simulation frameworks. To understand this evolution, let us begin by examining the foundational work established by PyMaltaSim and the specific challenges that motivated the development of Malta2FastSim.

3.1 PyMaltaSim: Foundational Physics Implementation

PyMaltaSim established the crucial first step in MALTA detector simulation through a comprehensive standalone Python implementation. The architecture of PyMaltaSim follows a hierarchical approach that mirrors the physical processes occurring within the detector: Energy depositions $\rightarrow$ Pixel hits $\rightarrow$ Words $\rightarrow$ Digits.

This progression reflects the natural flow of information in the detector system. Energy depositions from particle interactions create the fundamental physical input. These depositions are then processed through detector physics models to determine which pixels register hits above threshold. The pixel hits are subsequently organised into readout words according to the spatial grouping scheme, and finally merged into digits representing the complete detector output.

3.1.1 Validated Physics Models

The success of PyMaltaSim rested on its implementation of carefully validated physics models derived from extensive measurements on real MALTA detectors. These models capture the complex relationships between charge deposition, signal formation, and detector response.

- **Charge-dependent timing delays** form a cornerstone of the physics implementation. The relationship between deposited charge and signal formation time is implemented through look-up tables (LUTs) that relate charge magnitude to timing delays. The physical basis for this relationship lies in the signal formation process: larger charge depositions create stronger signals that saturate the front-end amplifiers more rapidly, leading to earlier threshold crossings and shorter formation times.

- **Position-dependent delays** account for the spatial variations in charge collection efficiency across each pixel. The timing delays vary with the radial distance from the pixel centre electrode according to:

$$t_{\text{pos}} = \sum_{i=0}^{2} c_i r^i \tag{6}$$

  where $r = \sqrt{p_x^2 + p_y^2}$ represents the distance from the electrode centre.

- **DAC-based threshold calculations** provide realistic modelling of the detector electronics. The threshold is calculated from the Digital-to-Analog Converter settings (IDB, ITHR, IBIAS, ICASN) using:

$$\text{threshold} = \frac{\sqrt{T_{\text{IDB}}^2 + T_{\text{ITHR}}^2 + T_{\text{IBIAS}}^2 + T_{\text{ICASN}}^2}}{2} \tag{7}$$

This approach captures the realistic noise characteristics and threshold variations present in the actual detector electronics.

3.1.2 Complete Detector Readout Chain

PyMaltaSim implements the complete signal processing chain from initial charge deposition through final digital output. The spatial grouping follows the detector's 2×8 pixel organisation, with appropriate time-based merging that accounts for the 25 ns timing resolution and 1.6 ns column merging characteristics.

The readout chain implementation includes the realistic timing constraints imposed by the detector architecture, ensuring that the simulation accurately represents the temporal characteristics observed in real detector operation.

3.2 Integration Challenge: Limitations of Standalone Implementation

While PyMaltaSim provided excellent physics validation and demonstrated the feasibility of fast MALTA simulation, its standalone nature created significant limitations for systematic detector studies. Understanding these limitations is crucial for appreciating the motivation behind Malta2FastSim development.

3.2.1 Isolation from Simulation Frameworks

The most fundamental limitation arose from PyMaltaSim's isolation from standard high-energy physics simulation workflows. Modern detector development relies heavily on integrated simulation frameworks like Allpix^2 that provide standardised particle generation, geometry handling, and analysis tools. PyMaltaSim's standalone architecture required manual particle trajectory handling that was incompatible with these standard workflows.

This isolation meant that detector studies using PyMaltaSim required custom particle generation, manual geometry definitions, and bespoke analysis tools. Such an approach is acceptable for proof-of-concept studies but becomes prohibitively complex for systematic parameter space exploration or large-scale detector optimisation studies.

3.2.2 Scalability and Performance Constraints

The Python implementation, while excellent for model development and validation, presented scalability challenges for production-level simulations. High-throughput studies required for detector optimisation demand simulation speeds of $10^4 - 10^5$ particles per second on standard computing hardware. The interpretive nature of Python, combined with the overhead of standalone operation, limited the practical scale of studies that could be undertaken.

Additionally, the lack of integration with established high-energy physics analysis frameworks meant that results required custom analysis pipelines, further complicating large-scale studies.

3.2.3 Framework Dependencies and Ecosystem Integration

Modern detector development increasingly relies on integrated tool ecosystems that span simulation, reconstruction, and analysis. PyMaltaSim's isolation from frameworks like Allpix^2 meant that studies could not take advantage of the extensive validation, cross-checks, and analysis tools available within these ecosystems.

The absence of integration also complicated collaborative work, as researchers required separate tool chains for PyMaltaSim studies versus standard framework-based analyses.

3.3 Solution Architecture: Malta2FastSim Development Strategy

The development of Malta2FastSim addressed these limitations through a comprehensive solution architecture that preserved the validated physics of PyMaltaSim while enabling full integration within the Allpix^2 framework ecosystem.

3.3.1 Native C++ Framework Integration

The transition to a native C++ implementation within Allpix^2[9] addressed the scalability limitations while enabling seamless integration with existing simulation workflows. The C++ implementation provides the computational performance required for large-scale studies while maintaining full compatibility with Allpix^2 message passing, configuration systems, and analysis tools.

This integration eliminates the need for custom particle generation or geometry handling, as Malta2FastSim operates directly on Monte Carlo particle tracks generated by the standard Allpix^2 infrastructure.

3.3.2 Enhanced Physics Models

While preserving all validated PyMaltaSim physics models, Malta2FastSim incorporates enhanced physics capabilities that address specific limitations identified during PyMaltaSim operation. The most significant enhancement involves the incorporation of multiple scattering effects for improved cluster modelling.

3.3.3 Algorithmic Innovations

The framework integration enabled algorithmic innovations that solve fundamental efficiency problems encountered in high-resolution trajectory sampling. These innovations, particularly the pixel-grouped accumulation algorithm and selective Highland scattering implementation, represent significant advances in simulation capability while maintaining full physics model validation.

The solution architecture thus represents a comprehensive approach to fast simulation development that balances physics accuracy, computational efficiency, and framework integration requirements necessary for modern detector development programs.

4 Malta2FastSim Module: Technical Implementation

The Malta2FastSim module represents the culmination of careful physics model preservation combined with innovative algorithmic solutions designed to address the specific challenges encountered in high-resolution detector simulation. Let us examine the technical implementation systematically, beginning with the physics model integration and progressing through the algorithmic innovations that enable production-ready simulation capabilities.

4.1 Physics Model Integration

The cornerstone of Malta2FastSim lies in its faithful preservation of the validated physics models established by PyMaltaSim, while adapting these models for operation within the Allpix^2 framework. This preservation ensures that all the careful calibration and validation work performed with PyMaltaSim remains directly applicable to the new implementation.

4.1.1 Look-Up Table (LUT) Framework

The module maintains PyMaltaSim's LUT-based approach for computational efficiency. Look-up tables store pre-calculated detector response parameters, enabling rapid interpolation rather than expensive real-time calculations. This approach proves particularly important for large-scale studies where computational efficiency directly impacts the feasibility of parameter space exploration.

The charge-dependent timing relationship captures the fundamental physics of signal formation in MALTA's small-electrode design. Larger charge depositions create stronger signals that saturate the front-end amplifiers more rapidly, leading to earlier threshold crossings and shorter signal formation times. The LUT implementation preserves this relationship through discrete calibration points:

$$t_{\text{signal}} = f_{\text{LUT}}(Q_{\text{deposited}}) \tag{8}$$

where the function f_{LUT} represents linear interpolation between measured calibration points spanning the range from 0.20 ke$^-$ to several ke$^-$.

Position-dependent delays account for the spatial variations in charge collection efficiency across each pixel. The timing delays increase with radial distance from the pixel centre electrode according to a polynomial relationship:

$$t_{\text{pos}} = \sum_{i=0}^{2} c_i r^i \tag{9}$$

where $r = \sqrt{p_x^2 + p_y^2}$ represents the distance from the electrode centre, and the coefficients c_i are determined from detailed measurements of real MALTA detectors.

The DAC-based threshold calculation provides realistic modelling of the detector electronics. The threshold combines contributions from multiple Digital-to-Analog Converter settings:

$$\text{threshold} = \frac{\sqrt{T_{\text{IDB}}^2 + T_{\text{ITHR}}^2 + T_{\text{IBIAS}}^2 + T_{\text{ICASN}}^2}}{2} \tag{10}$$

This approach captures both the systematic threshold variations and the realistic noise characteristics present in actual detector electronics.

4.1.2 Trajectory Partitioning and Charge Conservation

Malta2FastSim implements sophisticated trajectory partitioning that addresses a fundamental challenge in detector simulation: balancing spatial resolution with computational efficiency. Complex particle trajectories are discretised into N configurable steps, where the number of steps determines the trade-off between simulation accuracy and computational cost.

The temporal spacing between steps follows relativistic kinematics:

$$\Delta t = \frac{m|\Delta\vec{r}|}{|\vec{p}|N} \tag{11}$$

where m is the particle rest mass, $|\Delta\vec{r}|$ is the path length through the detector, $|\vec{p}|$ is the momentum magnitude, and N is the number of trajectory steps.

Charge conservation across the trajectory represents a critical physics requirement. The total deposited charge Q must be distributed among the N trajectory steps while maintaining statistical realism. The algorithm employs random partitioning with normalised weights:

$$q_i = Q \cdot s_i \quad \text{where} \quad \sum_{i=1}^{N} s_i = 1 \tag{12}$$

The random weights $\{s_i\}$ are generated from a uniform distribution and normalised to ensure perfect charge conservation while providing realistic statistical fluctuations in local energy deposition.

4.2 Algorithmic Innovations

While preserving the validated physics models represents a necessary foundation, Malta2FastSim introduces two key algorithmic innovations that address specific limitations encountered in high-resolution simulation. These innovations - pixel-grouped accumulation and selective Highland scattering - solve distinct physics problems while maintaining computational efficiency.

4.2.1 Pixel-Grouped Accumulation Algorithm

The pixel-grouped accumulation algorithm addresses a fundamental efficiency problem that emerges in high-resolution trajectory sampling. Understanding this problem requires examining the relationship between trajectory resolution and detection efficiency.

Problem Identification: As we increase the number of trajectory steps N to improve spatial accuracy, the individual step charges become $q_i \approx Q/N$. For large N, these individual charges can become small relative to detection thresholds, creating a risk of false negatives where $q_i < Q_{\text{threshold}}$ despite $Q_{\text{total}} > Q_{\text{threshold}}$.

This problem represents a fundamental trade-off between spatial accuracy and detection sensitivity that would severely limit the utility of high-resolution simulation.

Solution Strategy: Rather than comparing individual step charges to threshold, the pixel-grouped accumulation algorithm groups trajectory steps by their pixel location and accumulates charge within each pixel before applying threshold tests.

The algorithm operates through five systematic steps:

1. **Process each trajectory step**: Calculate the charge q_i and timing delays t_i for each step, then determine the pixel coordinates (col, row).

2. **Pixel grouping logic**: Implement conditional logic based on pixel location:

   - First step: Initialise pixel group
   - Same pixel as previous step: Accumulate charge and delays
   - New pixel: Finalise previous pixel group and start new group

3. **Timing calculation**: For each pixel group, calculate the charge-weighted timing:

$$\exists \min k \text{ such that } \sum_{i<k} q_i > Q_{\text{threshold}} \implies \tau = \sum_{i<k} t_k \tag{13}$$

4. **Threshold comparison**: Apply the threshold test to the accumulated pixel charge:

$$Q_{\text{pixel}} = \sum_{j \in \text{pixel}} q_j \overset{?}{>} Q_{\text{threshold}} \tag{14}$$

5. **Hit creation**: Generate PixelHit objects only for pixels passing the threshold test.

Physical Impact: This approach maintains detection efficiency while enabling arbitrary trajectory resolution, effectively solving the fundamental trade-off between spatial accuracy and detection sensitivity. The algorithm ensures that multiple small depositions within the same pixel are properly combined before threshold comparison, preventing the efficiency losses that would otherwise occur in high-resolution sampling.

4.2.2 Selective Highland Multiple Scattering

The Highland scattering implementation addresses a separate physics problem related to cluster modelling in test beam configurations. When particles traverse the detector perpendicularly - a common scenario in test beam studies - they tend to activate only single pixels, missing the lateral charge sharing that occurs with realistic particle trajectories.

Physical Motivation: Perpendicular particle trajectories present specific challenges for simulation accuracy:

- Direct perpendicular hits activate only single pixels
- Missing lateral charge sharing produces unrealistic cluster sizes
- Underestimated detection efficiency due to incomplete physics modelling

Highland Formula Implementation: For perpendicular trajectories, multiple scattering is applied using the Highland formula for the characteristic scattering angle[6]

$$\theta_0 = \frac{13.6 \text{ MeV}}{\beta p} z \sqrt{\frac{t}{X_0}} \left[1 + 0.038 \ln \left(\frac{t}{X_0} \right) \right] \tag{15}$$

where β and p represent particle velocity and momentum, z is the particle charge, t is the material thickness, and X_0 is the radiation length.

Selective Application: The scattering is applied conditionally based on trajectory analysis:

The condition `initial_pixel == final_pixel` identifies perpendicular trajectories that would otherwise produce unrealistic single-pixel hits. When this condition is met, the algorithm applies Highland scattering to deflect the trajectory until multi-pixel cluster formation occurs.

Physical Result: This selective application ensures realistic cluster size distributions and without lateral charge sharing, improving efficiency modelling accuracy specifically for perpendicular track geometries while avoiding unnecessary computational overhead for trajectories that already exhibit realistic charge sharing patterns.

4.3 Technical Implementation Details

The technical implementation of Malta2FastSim demonstrates careful attention to both computational efficiency and software engineering best practices. The module integrates seamlessly with the Allpix^2 framework while providing the performance characteristics necessary for large-scale detector studies.

4.3.1 `Allpix`^{2} Integration

Malta2FastSim operates as a native Allpix^2 module, utilising the framework's established patterns for configuration management, message passing, and event processing. The integration eliminates the manual particle generation and geometry handling required by standalone implementations, while providing access to the extensive validation and analysis tools available within the Allpix^2 ecosystem.

The module processes Monte Carlo particle tracks generated by the standard Allpix^2 infrastructure, ensuring compatibility with established particle generators and beam modelling tools. This integration enables direct comparison with other detector simulation approaches and facilitates systematic studies across multiple detector technologies.

4.3.2 Configuration Management

The module supports comprehensive configuration through standard Allpix^2 mechanisms. Key configurable parameters include DAC settings (IDB, ITHR, IBIAS, ICASN), trajectory discretisation parameters, timing resolution settings, and Highland scattering parameters specific to the detector material properties.

This configuration flexibility enables the module to adapt to different detector variants, experimental conditions, and simulation requirements without requiring code modifications.

4.3.3 Memory Management and Performance Optimisation

The LUT interpolation algorithms achieve O(log n) complexity through binary search methods, ensuring that lookup operations remain efficient even for large parameter spaces. The overall simulation complexity scales linearly with trajectory step count N and logarithmically with LUT complexity.

A Real-World Validation of MALTA Sensor Performance

The development of fast simulation capabilities for MALTA detectors must ultimately be validated against real experimental data. This appendix presents an examination of Device Under Test (DUT) characterisation methodology and results, demonstrating how systematic testing of MALTA samples informs both detector optimisation and simulation validation.

A.1 DUT Classification Framework

To enable systematic studies of MALTA detector performance across the parameter space relevant for HL-LHC operation, samples are classified according to multiple design and processing variables. This classification framework ensures comprehensive coverage of the design space while enabling statistical analysis of performance dependencies.

A.1.1 Substrate Technology Categories

The choice of substrate technology represents one of the most fundamental design decisions affecting MALTA detector performance:

- **Epitaxial Substrates**: High-resistivity epitaxial layers grown on heavily-doped substrates. These substrates provide excellent uniformity and process control, generating approximately 1500 e^- signal charge for minimum ionising particles. The epitaxial approach offers mature technology with well-understood electrical characteristics.

- **Czochralski Substrates**: Thick (100-300 μm), high-resistivity p-type silicon substrates produced by the Czochralski crystal growth method. The increased active thickness results in significantly enhanced signal generation. This enhanced signal amplitude provides improved robustness against radiation-induced charge trapping effects.

A.1.2 Thickness Variations

The active sensor thickness directly impacts both signal amplitude and charge collection efficiency. MALTA samples are produced with thicknesses ranging from 50 μm to 300 μm, enabling systematic studies of the trade-offs between signal strength, depletion voltage requirements, and radiation tolerance. Thicker substrates provide enhanced signal amplitude but require higher depletion voltages and may exhibit increased susceptibility to bulk radiation damage.

A.1.3 Process Modification Categories

Standard CMOS imaging processes require modification to achieve the radiation hardness and charge collection efficiency needed for particle physics applications. MALTA incorporates two key process modifications that can be implemented individually or in combination:

- **N-layer Gap (NGAP)**: This modification creates a gap in the low-dose n-type blanket implant at pixel edges. TCAD simulations demonstrate that this approach enhances lateral electric fields in pixel cornersâĂŤregions that are typically problematic for charge collection efficiency. The NGAP modification is particularly effective for improving efficiency in the challenging corner regions where charge carriers must drift the maximum distance to reach collection electrodes.

- **Extra Deep P-Well (XDPW)**: An additional deep p-well implant is introduced around pixel edges to further enhance charge collection. This 4 μm wide implant creates stronger lateral electric fields that improve the drift of ionisation charge toward collection electrodes. The XDPW modification is particularly important for maintaining efficiency after radiation damage when charge trapping becomes significant.

A.1.4 Doping Level Classification

The doping concentration in the substrate affects both the depletion characteristics and the radiation tolerance of MALTA sensors. Samples are classified according to their doping levels. The exact information concerning the doping levels is classified. However, the recognised nomenclature is:

1. Low
2. Medium
3. High
4. Very High

A.1.5 Irradiation Level Classification

Since MALTA sensors must operate in the intense radiation environment of HL-LHC experiments, systematic irradiation studies are essential for validating radiation tolerance. Samples are categorised according to their neutron fluence exposure:

$$\text{Irradiation Level } i \equiv i \times 10^{15} \text{ 1-MeV } \text{n}_{\text{eq}}/\text{cm}^2 \tag{16}$$

where i ranges from 1 to 5, corresponding to fluences from 1×10^{15} to 5×10^{15} 1-MeV $\text{n}_{\text{eq}}/\text{cm}^2$. This classification encompasses the range expected for the innermost tracking layers at HL-LHC, where the most demanding radiation tolerance requirements must be met.

A.2 Experimental Methodology: Test Beam Characterisation

The systematic characterisation of MALTA DUTs requires controlled experimental conditions that replicate the essential features of particle detection while enabling precise measurement of efficiency and timing performance. Test beam facilities provide the necessary infrastructure for such measurements through well-characterized particle beams and precision tracking systems.

A.2.1 Test Beam Infrastructure

The measurements described here were performed using the MALTA telescope installed at the H6A beam line of the CERN Super Proton Synchrotron (SPS). This facility provides 180 GeV/c hadron beams with excellent stability and well-understood particle composition. The telescope configuration includes six MALTA planes providing precision track reconstruction with spatial resolution of approximately 10 μm, enabling accurate association of reconstructed tracks with DUT hits.

A.2.2 DUT Integration and Measurement Protocol

The DUT integration protocol ensures that measurements accurately reflect intrinsic detector performance while minimising systematic uncertainties:

1. **Mechanical Integration**: The DUT is positioned at the telescope center with precision alignment ensuring perpendicular beam incidence. Environmental controls maintain stable temperature and humidity conditions throughout data collection.

2. **Electrical Configuration**: Bias voltage and threshold settings are optimized for each DUT based on its specific characteristics. The DAC settings (IDB, ITHR, IBIAS, ICASN) are systematically varied to map the efficiency and noise performance across the operational parameter space.

3. **Data Collection**: For each configuration, sufficient statistics are collected to achieve sub-percent precision on efficiency measurements. Typical data sets include $10^5 - 10^6$ triggered events, ensuring robust statistical analysis.

4. **Systematic Studies**: Operating parameters are varied systematically to understand performance dependencies and optimize operational settings for maximum efficiency and timing resolution.

A.3 Case Study: Very High Doping, Level 5 Irradiated Sample

To illustrate the DUT characterisation methodology and demonstrate the challenges posed by radiation damage, we present detailed results from a particularly demanding test case: a very high doped MALTA sample subjected to Level 5 irradiation (5×10^{15} 1-MeV n_{eq}/cm^2). This sample represents the most challenging operational conditions expected at HL-LHC and provides crucial validation data for radiation hardness projections.

A.3.1 Sample Characteristics

The test sample incorporated the following design features:

- **Substrate**: Czochralski, 300 μm thickness
- **Doping**: Very high
- **Process modifications**: XDPW
- **Irradiation level**: Level 5 (5×10^{15} 1-MeV n_{eq}/cm^2)

This configuration represents the most radiation-hard design available, incorporating all known techniques for maintaining efficiency under extreme radiation conditions.

A.3.2 Performance Results and Analysis

The systematic measurements on the Level 5 irradiated sample yielded several findings:

Maximum Achievable Efficiency

Despite implementing all available optimisation techniques, the maximum attainable efficiency reached only 56% for this heavily irradiated sample. This result demonstrates the fundamental challenge posed by radiation-induced charge trapping at extreme fluence levels. The efficiency limitation arises from several physical mechanisms:

- **Charge Trapping**: Radiation-induced defects create energy levels within the silicon bandgap that act as charge traps, removing charge carriers from the collection process.
- **Reduced Active Volume**: High trap concentrations effectively reduce the sensitive volume of each pixel, as charge generated in regions with high trap density has low probability of successful collection.
- **Signal Amplitude Reduction**: Incomplete charge collection reduces signal amplitudes below detection thresholds, even for minimum ionising particles that would normally produce easily detectable signals.

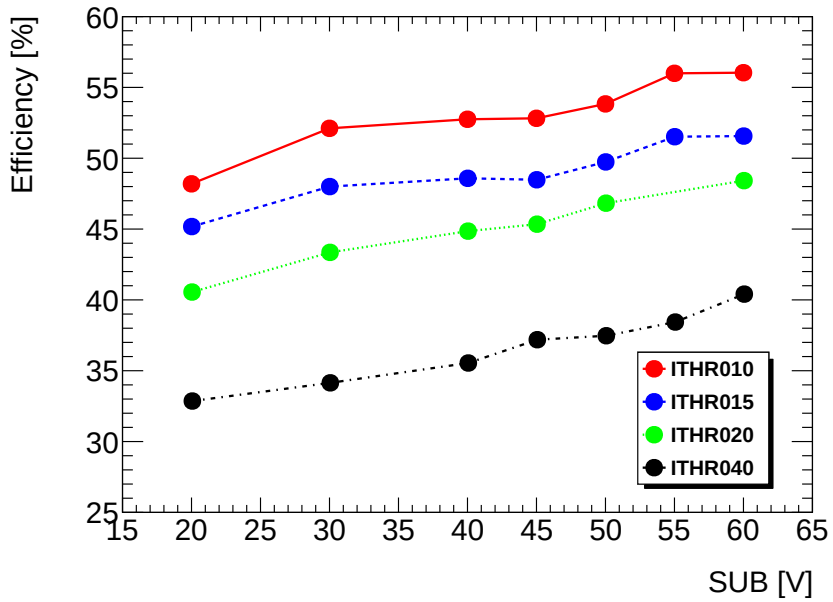

Figure 1: The efficiency vs. BIAS voltage applied to a Very High doped (vh-doped) sample with irradiance level 5. Note here to reliance of the efficiency on BIAS volatage and threshold current.

Irradiation Level Dependence

The strong dependence of efficiency on irradiation level provides crucial validation of radiation damage models. Samples subjected to lower irradiation levels demonstrated significantly higher efficiencies:

- **Unirradiated samples**: Efficiencies exceeding 98%
- **Level 1 irradiation** (1×10^{15} n_{eq}/cm^2): Efficiencies of 85-90%
- **Level 3 irradiation** (3×10^{15} n_{eq}/cm^2): Efficiencies of 70-75%
- **Level 5 irradiation** (5×10^{15} n_{eq}/cm^2): Maximum efficiency of 56%

This systematic degradation confirms the expected scaling of radiation damage effects with fluence and validates the simulation models used for radiation damage prediction.

A.3.3 Efficiency Optimisation Strategies

The systematic characterisation revealed two primary mechanisms for efficiency enhancement in radiation-damaged sensors:

Bias Voltage Optimisation

Increasing the applied bias voltage enhances the electric field strength throughout the depleted region according to:

$$E \propto \sqrt{V_{\text{bias}} e} \tag{17}$$

Higher electric fields accelerate charge carrier drift velocities, reducing the probability of charge trapping during collection. The relationship between drift velocity and electric field follows:

$$v_{\text{drift}} = \mu E \tag{18}$$

where μ is the carrier mobility. In radiation-damaged silicon, many charge carriers are trapped during their drift to collection electrodes. By increasing drift velocities through higher electric fields, the collection time is reduced, thereby decreasing the probability of trapping and improving overall efficiency.

Threshold Current Reduction

Lowering the operating threshold current (ITHR DAC setting) directly increases detection efficiency by accepting weaker signals that would otherwise fall below the discrimination threshold. This approach is particularly effective for radiation-damaged sensors where signal amplitudes are reduced due to incomplete charge collection. However, threshold reduction must be balanced against noise considerations to maintain acceptable signal-to-noise ratios.

A.3.4 Implications for Future Detector Development

The results from this challenging test case provide several important insights for future MALTA development:

Design Strategy Validation

The fact that any measurable efficiency could be achieved at Level 5 irradiation validates the design strategy of combining multiple radiation hardness techniques. The XDPW and NGAP process modifications, combined with very high doping and thick Czochralski substrates, provide measurable benefits even under extreme radiation conditions.

Operational Parameter Optimisation

The systematic efficiency optimisation demonstrates that careful tuning of operational parameters can significantly improve performance in radiation-damaged sensors. This finding emphasizes the importance of adaptive operational strategies that adjust bias voltages and thresholds based on accumulated radiation dose.

Technology Development Direction

While the 56% efficiency achieved represents a significant accomplishment given the extreme irradiation level, it falls short of the requirements for innermost detector layers at HL-LHC. This result motivates continued development of more aggressive radiation hardness techniques.

A.4 Conclusions and Future Outlook

The systematic characterisation of MALTA DUTs provides essential validation of detector performance and radiation tolerance, informing both current operational strategies and future development directions. The case study of the Level 5 irradiated sample demonstrates both the challenges posed by extreme radiation environments and the effectiveness of systematic optimisation approaches.

Key findings from this work include:

1. **Radiation Hardness Validation**: The systematic efficiency degradation with increasing irradiation level confirms radiation damage models and validates design strategies for harsh radiation environments.

2. **Optimisation Strategy Effectiveness**: The demonstrated ability to improve efficiency through bias voltage and threshold optimisation provides practical operational guidance for radiation-damaged sensors.

3. **Technology Development Requirements**: The performance limitations at extreme irradiation levels motivate continued development of more aggressive radiation hardness techniques.

B Mathematical Foundations and Derivations

This appendix provides a comprehensive mathematical exploration of the physics principles that govern MALTA detector operation and simulation. Rather than simply presenting final formulas, we will build understanding from first principles, carefully developing each concept through systematic derivation. This pedagogical approach illuminates not just what the mathematics tells us, but why these particular equations emerge from the underlying physics.

Understanding these mathematical foundations serves multiple purposes in our exploration of MALTA detectors. First, it provides the theoretical framework necessary to appreciate why certain design choices prove optimal. Second, it reveals the physical limitations that constrain detector performance. Finally, it establishes the mathematical rigor underlying the simulation approaches we have developed.

B.1 Semiconductor Physics: The Foundation of Detection

To understand how MALTA detectors convert particle interactions into measurable electrical signals, we must first establish the mathematical framework governing charge transport in semiconductors. This journey begins with the fundamental equations that describe how charge carriers move through silicon under the influence of electric fields and concentration gradients.

B.1.1 Building the Drift-Diffusion Framework

The behavior of charge carriers in semiconductors emerges from the interplay between two fundamental transport mechanisms: drift under electric fields and diffusion due to concentration gradients. Let us develop this understanding systematically, beginning with the basic physics and building toward the complete mathematical description.

Consider first a volume element of silicon containing mobile charge carriers. The rate at which the carrier concentration changes in this volume depends on four physical processes: carriers entering due to drift, carriers entering due to diffusion, carriers generated by external processes, and carriers lost to recombination. This physical reasoning leads directly to the continuity equation.

For electrons, we can write the continuity equation as:

$$\frac{\partial n}{\partial t} = -\frac{1}{q}\nabla \cdot \vec{J}_n + G_n - R_n \tag{19}$$

Let us examine each term to understand its physical significance. The partial derivative $\partial n/\partial t$ represents the local rate of change of electron concentration. The divergence term $\nabla \cdot \vec{J}_n$ describes the net current flow out of our volume element, which decreases the local concentration when positive. The generation term G_n accounts for processes that create electron-hole pairs, such as particle interactions or thermal generation. The recombination term R_n represents processes that remove electrons from the conduction band.

For holes, the continuity equation takes a similar form, but with a crucial sign difference:

$$\frac{\partial p}{\partial t} = \frac{1}{q}\nabla \cdot \vec{J}_p + G_p - R_p \tag{20}$$

The sign difference in the divergence term reflects the fact that conventional current flow represents positive charge motion, while electron current flows opposite to electron motion.

Now we must determine the current densities $\vec{J}_n$ and $\vec{J}_p$. Physical intuition suggests that carriers should move in response to both electric fields and concentration gradients. Electric fields exert forces on charged particles, causing them to drift with velocity proportional to the field strength. Concentration gradients, meanwhile, drive diffusion as carriers move from regions of high concentration to regions of low concentration.

For electrons, the current density combines these two contributions:

$$\vec{J}_n = q\mu_n n\vec{E} + qD_n\nabla n \tag{21}$$

The first term represents drift current, where μ_n is the electron mobility and $\vec{E}$ is the electric field. The mobility quantifies how quickly electrons respond to electric fields, with typical values in silicon of approximately 1400 $\text{cm}^2/(\text{V}\cdot\text{s})$ at room temperature. The second term represents diffusion current, where D_n is the diffusion coefficient.

For holes, the current expression becomes:

$$\vec{J}_p = q\mu_p p\vec{E} - qD_p\nabla p \tag{22}$$

Notice the sign difference in the diffusion term. This reflects the fact that hole diffusion down a concentration gradient produces current flow in the same direction as the concentration decrease, while electron diffusion produces current flow opposite to the concentration decrease.

The relationship between mobility and diffusion coefficient is not arbitrary but emerges from fundamental statistical mechanics. In thermal equilibrium, the drift and diffusion currents must exactly balance to prevent net current flow. This balance condition leads to the Einstein relation:

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{k_B T}{q} = V_T \tag{23}$$

where $V_T \approx 26$ mV at room temperature is the thermal voltage.

Physical Insight for MALTA Operation: In MALTA's depleted regions, strong electric fields ensure that drift dominates over diffusion. This mathematical framework explains why depletion is crucial for fast charge collection. The drift velocity $v = \mu E$ can reach several centimeters per microsecond in MALTA's electric fields, enabling charge collection within the 25 ns bunch crossing interval required at the LHC.

B.1.2 Electric Field Distributions in Small-Electrode Geometries

The unique characteristics of MALTA detectors emerge largely from their small collection electrodes, which create complex electric field patterns that strongly influence charge collection efficiency. To understand these effects quantitatively, we must solve for the electric field distribution in the detector geometry.

The electric field derives from the electrostatic potential through $\vec{E} = -\nabla\phi$. The potential itself satisfies Poisson's equation:

$$\nabla^2\phi = -\frac{\rho(\vec{r})}{\epsilon_s} \tag{24}$$

where $\rho(\vec{r})$ represents the space charge density and $\epsilon_s \approx 11.7$ is the relative permittivity of silicon.

In MALTA's depleted regions, the space charge density is dominated by ionised dopant atoms. For a uniformly doped p-type substrate with acceptor concentration N_A, we have $\rho = qN_A$ in fully depleted regions. However, the boundary conditions imposed by the electrode geometry create the most significant effects on the field distribution.

Consider a simplified two-dimensional model where a circular electrode of radius a sits at the center of a square pixel of dimension $L \times L$. This geometry captures the essential physics while remaining mathematically tractable. The boundary conditions are:

$$\phi(r = a) = V_{\text{electrode}} \tag{25}$$

$$\phi(r = L/2) = V_{\text{substrate}} \tag{26}$$

For the case of uniform space charge, we can solve this problem analytically in cylindrical coordinates. The solution takes the form:

$$\phi(r) = V_{\text{substrate}} + (V_{\text{electrode}} - V_{\text{substrate}})\frac{\ln(L/2r)}{\ln(L/2a)} + \frac{qN_A}{4\epsilon_s}(r^2 - (L/2)^2) \tag{27}$$

This solution reveals crucial insights about MALTA's operation. The logarithmic term dominates near the electrode, creating the strong radial fields necessary for charge collection. The quadratic term represents the parabolic potential arising from the uniform space charge.

The electric field components are:

$$E_r(r) = -\frac{\partial\phi}{\partial r} = \frac{V_{\text{electrode}} - V_{\text{substrate}}}{\ln(L/2a)}\frac{1}{r} - \frac{qN_A r}{2\epsilon_s} \tag{28}$$

The radial field exhibits the characteristic $1/r$ dependence that creates strong collection near the electrode center but weak fields in pixel corners. This mathematical result directly explains the efficiency problems that motivated the NGAP and XDPW process modifications described in our main text.

Let us examine the numbers to understand the practical implications. For a typical MALTA pixel with $a = 1.5$ μm and $L = 36.4$ μm, the factor $\ln(L/2a) \approx \ln(12) \approx 2.5$. With an electrode-substrate voltage difference of 6 V, the radial field at the electrode edge becomes:

$$E_r(a) = \frac{6 \text{ V}}{2.5 \times 1.5 \times 10^{-4} \text{ cm}} \approx 1.6 \times 10^4 \text{ V/cm} \tag{29}$$

This strong field ensures rapid charge collection near the electrode. However, at the pixel corner where $r \approx L/2$, the radial field drops to:

$$E_r(L/2) = \frac{6 \text{ V}}{2.5 \times 1.8 \times 10^{-3} \text{ cm}} \approx 1.3 \times 10^3 \text{ V/cm} \tag{30}$$

This factor of twelve reduction in field strength significantly slows charge collection from pixel corners, explaining why process modifications focus on enhancing corner collection efficiency.

B.1.3 Ramo's Theorem: The Mathematics of Signal Induction

Understanding how moving charges generate measurable electrical signals requires careful analysis of the electromagnetic induction process. Ramo's theorem provides the mathematical framework that connects microscopic charge motion to macroscopic current signals, forming the foundation for quantitative detector simulation.

The theorem emerges from Maxwell's equations but can be understood through physical reasoning about the relationship between charge motion and electric field changes. When a charge moves within a detector, it alters the electric field configuration throughout the device. This field change induces currents in all conductors, not just the one toward which the charge moves.

Ramo's theorem quantifies this induction process through the concept of weighting fields. For a charge q moving with velocity $\vec{v}$ at position $\vec{r}$, the current induced in electrode i is:

$$i_i(t) = q\vec{E}_{w,i}(\vec{r}(t)) \cdot \vec{v}(t) \tag{31}$$

The weighting field $\vec{E}_{w,i}$ is a purely geometric quantity, independent of the actual potentials applied to the electrodes. It represents the electric field that would exist if electrode i were at unit potential and all other conductors were grounded.

To understand why this mathematical form emerges, consider the energy stored in the electric field configuration. When the charge moves, the change in electrostatic energy must equal the work done by the induced currents. This energy conservation principle, combined with the linearity of Maxwell's equations, leads directly to Ramo's theorem.

For our cylindrical electrode geometry, the weighting field can be calculated by solving Laplace's equation:

$$\nabla^2 \phi_w = 0 \tag{32}$$

with boundary conditions $\phi_w(r = a) = 1$ and $\phi_w(r = L/2) = 0$. The solution is:

$$\phi_w(r) = \frac{\ln(L/2r)}{\ln(L/2a)} \tag{33}$$

The weighting field becomes:

$$\vec{E}_w(r) = -\nabla\phi_w = \frac{1}{\ln(L/2a)} \frac{\hat{r}}{r} \tag{34}$$

This result reveals a profound insight: the weighting field has exactly the same $1/r$ spatial dependence as the actual radial field, but with a different amplitude factor. This

mathematical relationship explains why charges near the electrode contribute much more strongly to the signal than charges far from the electrode.

The total charge collected can be calculated by integrating the induced current over time:

$$Q_{\text{collected}} = \int_0^{t_{\text{final}}} i(t')dt' = q[\phi_w(\vec{r}_{\text{final}}) - \phi_w(\vec{r}_{\text{initial}})] \tag{35}$$

This elegant result shows that the collected charge depends only on the starting and ending positions of the moving charge, not on the details of its trajectory. For complete charge collection, carriers must reach the electrode surface where $\phi_w = 1$. Carriers trapped before reaching the electrode contribute only the difference in weighting potential, which can be significantly less than the full charge.

B.2 Energy Loss and the Statistics of Particle Interactions

When charged particles traverse matter, they lose energy through electromagnetic interactions with atomic electrons and nuclei. Understanding this energy loss process requires careful consideration of both the average energy loss rate and the statistical fluctuations around this average. These mathematical descriptions form the foundation for accurate simulation of energy deposition in MALTA detectors.

B.2.1 Deriving the Bethe-Bloch Formula from First Principles

The Bethe-Bloch formula describes the average rate of energy loss for charged particles traversing matter. Rather than simply presenting this famous result, let us build understanding by deriving it from fundamental quantum mechanical principles.

The starting point is the quantum mechanical cross section for energy transfer from a charged particle to an atomic electron[3, 4].. When a charged particle with charge ze and velocity v passes an electron at impact parameter b, it transfers energy through the long-range Coulomb interaction.

The maximum energy transfer in a single collision is limited by conservation of energy and momentum. For a particle of mass M and energy E colliding with an electron at rest, the maximum energy transfer is:

$$T_{\max} = \frac{2m_e c^2 \beta^2 \gamma^2}{1 + 2\gamma m_e/M + (m_e/M)^2} \tag{36}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$ are the usual relativistic factors.

The differential cross section for transferring energy T to an atomic electron is given by the Møller scattering formula (for electrons) or the corresponding formula for other particles. For energy transfers much less than the particle's kinetic energy, this cross section takes the form:

$$\frac{d\sigma}{dT} = 2\pi r_e^2 m_e c^2 \frac{z^2}{T^2}\frac{1}{\beta^2}\left[1 - \frac{\beta^2 T}{T_{\max}}\right] \tag{37}$$

where $r_e = e^2/(4\pi\epsilon_0 m_e c^2)$ is the classical electron radius.

To find the average energy loss per unit path length, we multiply this cross section by the energy transfer T, the number density of electrons $n_e = \rho N_A Z/A$, and integrate over all possible energy transfers:

$$\left\langle \frac{dE}{dx} \right\rangle = n_e \int_{T_{\min}}^{T_{\max}} T \frac{d\sigma}{dT} dT \tag{38}$$

The minimum energy transfer $T_{\min}$ is set by atomic binding effects and is typically taken as the mean excitation energy I of the atom. Evaluating this integral:

$$\left\langle \frac{dE}{dx} \right\rangle = 2\pi r_e^2 m_e c^2 n_e \frac{z^2}{\beta^2} \int_I^{T_{\max}} \frac{1}{T} \left[1 - \frac{\beta^2 T}{T_{\max}} \right] dT \tag{39}$$

$$= 2\pi r_e^2 m_e c^2 n_e \frac{z^2}{\beta^2} \left[\ln \frac{T_{\max}}{I} - \frac{\beta^2}{2} \right] \tag{40}$$

Substituting the expressions for n_e and collecting constants, we obtain the Bethe-Bloch formula:

$$-\left\langle \frac{dE}{dx} \right\rangle = K z^2 \frac{Z}{A} \frac{1}{\beta^2} \left[\frac{1}{2} \ln \frac{2 m_e c^2 \beta^2 \gamma^2 T_{\max}}{I^2} - \beta^2 - \frac{\delta(\beta\gamma)}{2} \right] \tag{41}$$

where $K = 4\pi N_A r_e^2 m_e c^2 = 0.307$ MeV g^{-1}cm^2 and $\delta(\beta\gamma)$ is the density effect correction.

Let us understand the physics behind each term. The $1/\beta^2$ factor reflects the longer interaction time when particles move slowly, allowing more energy transfer. The logarithmic term arises from the long-range nature of the Coulomb interaction, with contributions from distant collisions that transfer small amounts of energy. The β^2 subtraction comes from relativistic corrections that reduce the energy transfer at high particle velocities.

For minimum ionising particles in silicon, this formula yields an energy loss of approximately 1.66 MeV g^{-1}cm^2 or about 3.9 MeV/cm. With silicon density of 2.33 g/cmÂş and an average energy of 3.6 eV per electron-hole pair, this corresponds to roughly 1100 electron-hole pairs per micrometer of silicon traversed.

B.2.2 Understanding Statistical Fluctuations: The Landau Distribution

While the Bethe-Bloch formula describes the average energy loss, individual particles exhibit significant fluctuations around this average. For thin absorbers, these fluctuations follow the Landau distribution, which emerges from the statistical nature of the collision process.

The mathematical development begins with recognising that energy loss results from many independent collision events, each governed by the cross section we derived above. However, unlike the central limit theorem scenario where all events contribute equally, energy loss is dominated by rare, high-energy collisions with large energy transfers.

Let us consider the probability of transferring energy T in a single collision. From our cross section formula, the probability of a collision transferring energy between T and $T + dT$ in path length dx is:

$$dP(T) = n_e \frac{d\sigma}{dT} dT\, dx \tag{42}$$

For thin absorbers where the number of collisions is relatively small, we cannot simply use Gaussian statistics. Instead, we must account for the proper Poisson statistics of the collision process combined with the $1/T^2$ energy dependence of the cross section.

The characteristic parameter that emerges from this analysis is:

$$\xi = Kz^2 \frac{Z}{A} \frac{t}{\beta^2} \tag{43}$$

where t is the absorber thickness in g/cmÂš. This parameter represents the most probable energy loss and sets the scale for fluctuations.

The full Landau distribution involves a complex mathematical function that cannot be expressed in terms of elementary functions. However, we can understand its key features through asymptotic analysis. The probability density function has the form:

$$f(\Delta E) = \frac{1}{\xi} \phi \left(\frac{\Delta E - \Delta E_{\text{mp}}}{\xi} \right) \tag{44}$$

where $\phi(\lambda)$ is the Landau function and ΔE_{mp} is the most probable energy loss.

The most important characteristic of the Landau distribution is its asymmetric shape with a long tail toward high energy losses. This tail reflects the finite probability of rare, high-energy collisions that can deposit much more energy than the average. For detector simulation, proper sampling of this tail is crucial for accurately representing the full range of energy deposition patterns.

The mean energy loss differs from the most probable energy loss by:

$$\langle \Delta E \rangle - \Delta E_{\text{mp}} = \xi[\gamma + 1 - \beta^2/2] \approx 0.42\xi \tag{45}$$

where $\gamma = 0.5772...$ is the Euler-Mascheroni constant. This shift quantifies the contribution of the high-energy tail to the average energy loss.

B.3 Multiple Scattering: A Complete Derivation of the Highland Formula

The Highland formula represents one of the most important results in experimental particle physics, describing the angular distribution of charged particles after traversing matter. Rather than presenting this formula as an empirical result, we will derive it systematically from first principles, building a complete understanding of the underlying physics.

B.3.1 Starting with Single Scattering: The Rutherford Cross Section

Our journey toward the Highland formula begins with Ernest Rutherford's famous analysis of particle scattering from individual atomic nuclei. When a charged particle approaches a nucleus, it experiences the Coulomb force and deflects according to classical mechanics.

Consider a particle with charge ze, mass m, and velocity v approaching a nucleus with charge Ze along a trajectory with impact parameter b. The equation of motion in the center-of-mass frame follows from conservation of energy and angular momentum.

The impact parameter relates to the scattering angle through the classical scattering formula:

$$\tan\left(\frac{\theta}{2}\right) = \frac{Ze^2}{4\pi\epsilon_0} \frac{1}{2T_{\text{cm}} b} \tag{46}$$

where T_{cm} is the kinetic energy in the center-of-mass frame.

To find the differential cross section, we use the fact that all particles with impact parameters between b and $b+db$ scatter into angles between θ and $\theta+d\theta$. The area of the annular ring is $2\pi b|db|$, so:

$$\frac{d\sigma}{d\Omega} = \frac{2\pi b|db|}{2\pi \sin\theta d\theta} = \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right| \tag{47}$$

Differentiating our impact parameter relation and substituting back, we obtain the famous Rutherford scattering formula:

$$\frac{d\sigma}{d\Omega} = \left(\frac{Ze^2}{4\pi\epsilon_0}\right)^2 \frac{1}{16T_{\text{cm}}^2 \sin^4(\theta/2)} \tag{48}$$

However, this classical result has a serious problem: it diverges as $\theta \to 0$, predicting infinite cross sections for small-angle scattering. This divergence arises because the Coulomb force has infinite range, leading to significant deflections even at very large impact parameters.

In real atoms, the nuclear Coulomb potential is screened by the electron cloud, effectively cutting off the interaction at large distances. This screening can be modeled using the Thomas-Fermi approximation, which gives an effective screening radius:

$$a_{\text{TF}} = \frac{0.885 a_0}{Z^{1/3}} \tag{49}$$

where $a_0 = 0.529 \times 10^{-8}$ cm is the Bohr radius.

Including screening effects, the differential cross section becomes:

$$\frac{d\sigma}{d\Omega} = \left(\frac{Ze^2}{4\pi\epsilon_0}\right)^2 \frac{1}{(2p)^2} \frac{1}{(\sin^2(\theta/2) + \theta_s^2/4)^2} \tag{50}$$

where p is the particle momentum and the screening angle is:

$$\theta_s = \frac{m_e c^2}{p} \frac{Z^{1/3}}{137} \tag{51}$$

This screened cross section now has finite total cross section and forms the foundation for multiple scattering theory.

B.3.2 Building Multiple Scattering Theory: The Statistical Approach

When a particle traverses a macroscopic thickness of matter, it undergoes many individual scattering events. For thick targets, we must develop a statistical description of the cumulative scattering effect.

The number of scattering events in thickness dx is given by:

$$dN = n\sigma_{\text{total}} dx \tag{52}$$

where n is the atomic number density and σ_{total} is the total scattering cross section.

The total cross section can be found by integrating our screened Rutherford formula over all solid angles:

$$\sigma_{\text{total}} = 2\pi \int_0^\pi \frac{d\sigma}{d\Omega} \sin\theta \, d\theta \tag{53}$$

$$= 4\pi \left(\frac{Ze^2}{4\pi\epsilon_0} \right)^2 \frac{1}{(2p)^2} \frac{1}{\theta_s^2} \tag{54}$$

For typical particle energies and materials, the number of scattering events in a thin absorber can be very large. For example, a 1 GeV particle traversing 1 cm of silicon undergoes approximately 10^5 individual scattering events.

When the number of scattering events becomes large, we can apply the central limit theorem to describe the cumulative scattering distribution. However, we must be careful about how we apply this theorem, because not all scattering events contribute equally to the final angular distribution.

The key insight, first developed by Molière, is that the multiple scattering distribution is determined primarily by scattering angles near the screening angle θ_s. Very small angles contribute little to the final distribution, while very large angles are too rare to be statistically significant.

B.3.3 Molière Theory: The Systematic Development

Hans Molière developed a systematic approach to multiple scattering by dividing the scattering events into categories based on their individual scattering angles. This categorisation allows us to apply the central limit theorem appropriately while accounting for the different statistical weights of different scattering processes.

Let us define the mean square scattering angle for single scattering events. For our screened Rutherford cross section, this quantity is:

$$\langle \theta^2_{\text{single}} \rangle = \int_0^\pi \theta^2 \frac{d\sigma}{d\Omega} \frac{2\pi \sin\theta}{\sigma_{\text{total}}} d\theta \tag{55}$$

The integrals involved in this calculation require careful treatment because of the logarithmic divergences that arise from the Coulomb interaction. The key insight is that the integral is dominated by scattering angles between the screening angle θ_s and some maximum angle θ_{max}.

The lower limit is set by atomic screening effects, while the upper limit is typically taken as the angle where single scattering and multiple scattering contributions become comparable. For most practical purposes, $\theta_{\text{max}} \sim \pi$.

Evaluating the integral with these limits:

$$\langle \theta^2_{\text{single}} \rangle = \frac{1}{\sigma_{\text{total}}} \int_{\theta_s}^\pi \theta^2 \frac{4\pi \left(\frac{Ze^2}{4\pi\epsilon_0} \right)^2}{(2p)^2} \frac{2\pi \sin\theta \, d\theta}{(\sin^2(\theta/2) + \theta_s^2/4)^2} \tag{56}$$

$$\approx \theta_s^2 \ln\left(\frac{204}{Z^{1/3}} \right) \tag{57}$$

The logarithmic factor arises from the integration over the range of scattering angles and depends on the atomic number through the screening effects.

For N independent scattering events, the central limit theorem tells us that the cumulative scattering distribution approaches a Gaussian with variance:

$$\langle \theta^2_{\text{multiple}} \rangle = N \langle \theta^2_{\text{single}} \rangle \tag{58}$$

After traversing thickness x, the total number of scattering events is $N = n\sigma_{\text{total}} x$. Substituting our expressions for N and $\langle \theta^2_{\text{single}} \rangle$:

$$\langle \theta^2 \rangle = n\sigma_{\text{total}} x \, \theta_s^2 \ln\left(\frac{204}{Z^{1/3}}\right) \tag{59}$$

Now we must express this result in terms of more convenient variables. The atomic number density is $n = \rho N_A / A$, where ρ is the material density, N_A is Avogadro's number, and A is the atomic mass. The total cross section involves factors that can be related to the radiation length X_0.

The radiation length emerges naturally from electromagnetic theory as the characteristic distance for high-energy electromagnetic processes. For our purposes, the key relationship is:

$$X_0 = \frac{716.4 \text{ g/cm}^2 \cdot A}{Z(Z+1)\ln\left(287/\sqrt{Z}\right)} \tag{60}$$

Through a series of algebraic manipulations involving the relationships between cross sections, atomic properties, and the radiation length, we can show that:

$$\langle \theta^2 \rangle = \left(\frac{13.6 \text{ MeV}}{\beta pc}\right)^2 z^2 \frac{x}{X_0} \tag{61}$$

This is the essential Highland formula, but we are not yet complete. Real multiple scattering distributions exhibit small deviations from the simple Gaussian prediction, particularly for thick absorbers where higher-order effects become important.

B.3.4 Higher-Order Corrections and the Complete Highland Formula

The Gaussian approximation provided by the central limit theorem works well for most practical purposes, but careful experiments reveal systematic deviations that become significant for thick absorbers. These deviations arise from several physical effects that our simple analysis has neglected.

First, we have assumed that all scattering events are independent and contribute equally to the final distribution. In reality, there are correlations between successive scattering events, and rare large-angle scatters can have disproportionate effects on the final distribution.

Second, we have neglected the fact that particles lose energy as they traverse matter, which gradually changes their momentum and therefore their scattering properties. This energy loss effect becomes important for thick absorbers.

Third, we have not accounted for the finite size effects that arise when the scattering angles become comparable to unity. Our small-angle approximations break

References

[1] ATLAS collaboration. Technical Design Report for the ATLAS Inner Tracker Pixel Detector. Technical Report CERN-LHCC-2017-021, CERN, 2017.

[2] I. Berdalovic et al. MALTA: a CMOS pixel sensor with asynchronous readout for the ATLAS High-Luminosity upgrade. In *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2018.

[3] H. Bethe. Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie. *Ann. Phys.*, 397:325–400, 1930.

[4] F. Bloch. Zur Bremsung rasch bewegter Teilchen beim Durchgang durch Materie. *Ann. Phys.*, 408:285–320, 1933.

[5] M. Dyndal et al. Mini-MALTA: radiation hard pixel designs for small-electrode monolithic CMOS sensors for the high luminosity LHC. *JINST*, 15:P02005, 2020.

[6] V. L. Highland. Some practical remarks on multiple scattering. *Nucl. Instrum. Methods*, 129:497–499, 1975.

[7] M. Munker et al. Simulations of CMOS pixel sensors with a small collection electrode, improved for a faster charge collection and increased radiation tolerance. *JINST*, 14:C05013, 2019.

[8] H. Pernegger et al. MALTA-Cz: a radiation hard full-size monolithic CMOS sensor with small electrodes on high-resistivity Czochralski substrate. *JINST*, 18:P09018, 2023.

[9] S. Spannagel, K. Wolters, D. Hynds, N. Alipour Tehrani, M. Benoit, D. Dannheim, N. Gauvin, A. NÃijrnberg, P. SchÃijtze, and M. Vicente. Allpix^2: A modular simulation framework for silicon detectors. *Nucl. Instrum. Methods Phys. Res. A*, 901:164–172, 2018.

[10] M. van Rijnbach et al. Performance of the MALTA telescope. *Eur. Phys. J. C*, 83:581, 2023.