

Computational Project Report

Signal Formation in a Silicon Microstrip Detector

Lorenzo Piersante
(Programming language: Python 3)

March 2022

Abstract

This report summarises the findings and the observations of a project which addresses the task of simulating the electrical signal produced in a silicon microstrip detector when energy is deposited on it by an ionizing particle. The project begins from a consideration of Ramo's theorem, which is the fundamental principle behind signal formation, and from that all the pieces required to generate an electrical signal are identified, solved for, and finally assembled together. The outcome is the prediction of a time-dependent current signal in a noise-less background which matches well the results found in the literature, especially chapter 5 of [16]. The mathematical modelling of the pre-amplification stage of the electronic processing chain is also considered, so that it was possible to calculate the voltage signal resulting from the current signal. The last section of this report investigates the effects of radiation damage on the silicon detector and a prediction of the changes in signal shape is attempted.

1 Introduction

When a charged particle is close to a conducting surface connected to ground, it will induce a certain amount of charge, with opposite sign, on the surface. If an electronic read-out system is connected to the surface, it is then possible to infer the presence of the charged particle. Let's consider the case of a parallel plate capacitor at the centre of which a positively charged particle is placed (fig. 1 (a)). As already mentioned, this particle will induce a negative charge on the capacitor plates. If a positive voltage is applied to the top plate, then the charge will move down, thus provoking more charge to be induced on the bottom electrode than it is induced on the top electrode. In other words, the charge motion generates a current through the capacitor. This current is the electrical signal that the read-out system will pick-up in order to determine the presence of the charge. The precise relationship between the charge induced on the electrodes, the motion of the charge and the geometry of the configuration of electrodes is described by Ramo's theorem, which will be discussed and proven in the next section.

Body

Figure 1: Parallel plate particle detector. (a) Parallel plate capacitor, (b) parallel plate capacitor with dielectric medium.

In Particle Physics applications, the main interest is not the determination of the presence of a particle, but the determination of particle's properties such as energy or momentum. A way to accomplish that consists in stopping the motion of the particle through the conversion of its energy into some measurable quantity. In the context of charge induction in an electrode configuration, that would be the conversion of energy into charges. If a semiconductor material, such as silicon, is interposed between the two capacitor plates (fig. 1 (b)), then a particle that impinges on the detector will excite electrons (and holes) across the

band gap. The number of electrons that are excited will depend on the energy of the incident particle, so that the signal that these charges will induce will be proportional to the energy deposited on the detector. More details concerning the operation of realistic semiconductor particle detectors will be outlined in later sections.

Silicon detector systems revolve around these basic ideas, and thanks to their flexibility they have become some of the most important particle tracking devices in modern Particle Physics. Silicon detectors generally constitute the innermost layer of detectors surrounding the beam-line, and because of that, they are traversed by very large particle fluxes. This fact translates into two major requirements for such devices: (i) quick response time, (ii) strong resistance to radiation damage. Furthermore, the launch of the High-Luminosity LHC, expected for 2027 [23], will pose the challenge of these detectors having to withstand even larger fluxes. Therefore, research into the physical processes occurring within such devices and the strive for their optimisation is a scientifically relevant endeavour.

Silicon detectors come in different designs according to the spatial resolution requirements. This project will focus on the one-sided p-on-n microstrip type, which is broadly composed of a lightly doped n-type (donor density of 10^{12}cm^{-3}) silicon wafer, $300\mu\text{m}$ thick and between 5 and 10cm in width and length, onto which, on one side only, narrow strips of highly doped p-type semiconductor (acceptor density of 10^{19}cm^{-3}) are implanted. The p-type strips (henceforth the p+ implants) considered in this project are $1\mu\text{m}$ thick, $20\mu\text{m}$ wide and are arranged with a pitch of $100\mu\text{m}$ between their centres. Fig. 2 displays a schematic view of the microstrip detector set-up considered in this project. The bottom of the wafer and the strips are metallised, generally with a $1\mu\text{m}$ thick aluminium layer, in order to create the electrical contacts needed for the application of a potential difference across the silicon slab and for the detection of the electrical signal induced by the motion of the charges liberated inside the silicon. The fact that the strips effectively constitute a segmented electrode allows a spatial resolution in the horizontal direction (fig. 2 (a)). Each strip is linked to an electronic read-out apparatus which is responsible for signal detection and amplification.

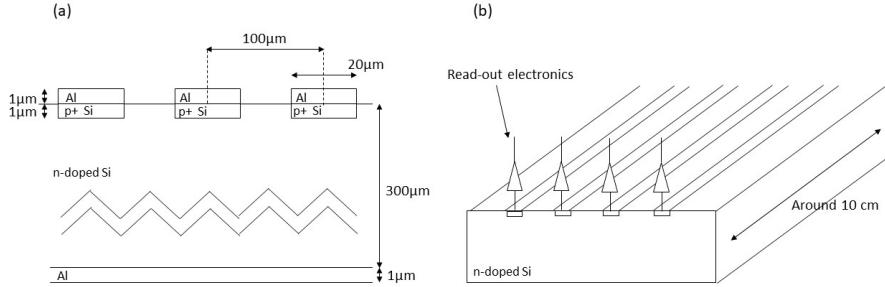


Figure 2: Silicon microstrip detector. (a) cross-section, (b) view of the longitudinal direction.

2 Principles of Signal Formation and Elements of Numerical Methods

2.1 Ramo's Theorem

Ramo's theorem is the pivotal concept upon which the entire project is developed. This theorem connects the motion of the charges within a configuration of electrodes to the charge induced on the electrodes themselves, thus enabling the calculation of the electrical signal. The proof that is provided here is based on energy arguments and follows [16]. A more thorough demonstration is given in [14], yet the argument below has the advantage of putting more emphasis on the physical processes at play. Fig. 3 shows an outer electrode (surface S_o)

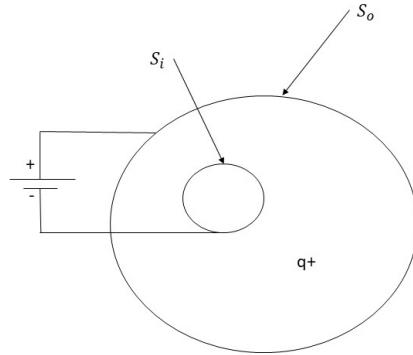


Figure 3: Electrode set-up for Ramo's theorem derivation.

at potential V , set by a power supply, and a grounded inner electrode (surface S_i). In the space between them a charge q is placed. Let's call the electric field due to the electrodes and the electric field due to the charge \vec{E}_0 and \vec{E}_q respectively. These two electric fields add up to produce some total electric field

\vec{E} with electric potential ϕ .

Given this, the energy balance equation for this system (charge and electrodes) is:

$$dW_q + dW_V + dW_E = 0 \quad (1)$$

where $dW_q = q\vec{E}_0 \cdot d\vec{r}$ is the work done on q by \vec{E}_0 when it moves an infinitesimal amount, $dW_V = d(QV) = VdQ$ is the energy provided by the power supply to keep the potential of the outer electrode at a constant value (Q is the charge that enters or leaves the power supply, it is positive if positive charges enter or negative charges leave the power supply, negative otherwise), and dW_E is the change in energy of the total electric field.

The energy of the electric field is given by:

$$U_E = \frac{1}{2}\epsilon_r\epsilon_0 \iiint_V \vec{E}^2 d\vec{r}^3 \quad (2)$$

where it is assumed that the space between the electrodes is filled with some homogeneous linear dielectric medium. If \vec{E} is substituted by $\vec{E}_0 + \vec{E}_q$, the equation becomes:

$$U_E = U_0 + U_q + \epsilon_r\epsilon_0 \iiint_V \vec{E}_0 \cdot \vec{E}_q d\vec{r}^3 \quad (3)$$

In order to evaluate the cross term, let's consider the decomposition of ϕ into the sum of the electrode potential, ϕ_0 , with boundary conditions $\phi_0(S_i) = 0$ and $\phi_0(S_o) = V$, and the charge potential, ϕ_q , with boundary conditions $\phi_q(S_i) = 0$ and $\phi_q(S_o) = 0$. Furthermore, ϕ_0 satisfies Laplace's equation, whilst ϕ_q satisfies Poisson's equation with a point source. The fact that $\vec{\nabla} \cdot \vec{E}_0 = 0$ justifies the equality $\vec{E}_0 \cdot \vec{E}_q = -\vec{\nabla} \cdot (\phi_q \vec{E}_0)$. The substitution of this expression and the application of the divergence theorem to (3) causes the cancellation of the cross term since ϕ_q vanishes at the surfaces of the volume. This implies that the energy of the electric field separates into two components associated with \vec{E}_0 and \vec{E}_q respectively. Hence $dW_E = dU_0 + dU_q = 0$ because \vec{E}_0 cannot change in response of the charge motion since the electrode potentials are fixed, while $dU_q = 0$ since \vec{E}_q cannot exchange energy with the electrodes, nor do work on q . Consequently, (1) becomes $dW_q + dW_V = 0$, which rearranging, substituting for the two differentials and dividing by V reads:

$$dQ = -q \frac{\vec{E}_0}{V} \cdot d\vec{r} \quad (4)$$

This equation describes the amount of charge that is induced on the outer electrode surface because of an infinitesimal displacement of q . The vector quantity \vec{E}_0/V (where V is just a constant) is called *weighting field*, \vec{E}_W , and it describes the coupling of a charge to a specific electrode in the configuration. The potential associated to this field is unsurprisingly called *weighting potential*, ϕ_W . Two matters must now be addressed: (i) the dependence of \vec{E}_W on the permittivity of the material between the electrodes, and (ii) the dependence of \vec{E}_W on

the presence of a fixed space charge between the electrodes. These two issues are relevant to the project because of the physical nature of silicon detectors, and the answer to both of them is negative.

Regarding (i), it is not difficult to show, using the discontinuity of the \vec{D} -field at the outer electrode surface, that the total induced charge is:

$$Q_{induced} = -\epsilon_r \epsilon_0 \iint_{S_o} \vec{E}_q \cdot d\vec{S} \quad (5)$$

And since $\vec{\nabla} \cdot \vec{E}_q = q\delta(\vec{r} - \vec{r}_q)/\epsilon_r \epsilon_0$, that is the electric field due to q depends linearly on $1/\epsilon_r$, it means that the induced charge will be independent of the permittivity of the medium. Obviously, this argument is valid so long the material can be considered linear.

Regarding (ii), still assuming the original set-up, the space charge contribution to the field can be accounted through a potential ϕ_ρ which satisfies $\vec{\nabla}^2 \phi_\rho = -\rho/\epsilon_r \epsilon_0$ with boundary conditions $\phi_\rho(S_i) = 0$ and $\phi_\rho(S_o) = 0$. Given these facts, it is not hard to realize that this field component cannot in any way contribute to (1) because it cannot exchange energy with the power supply (boundaries are set to 0), and because its energy is fixed by the assumption of a static charge density. Since the field energy is fixed, the field cannot exchange energy with q , thus it cannot do work on it. Hence, the presence of fixed space charge does not affect the induced charge.

The negative answer to points (i) and (ii) implies that (4) alone fully describes the process of charge induction, and that ϕ_W can be calculated by solving Laplace's equation for an electrode configuration where all but the electrode of interest are grounded. Because of the normalised nature of the field, the electrode of interest can simply be set to $V = 1$.

Finally, the induced charge and the induced current as a result of the motion of a charge can be written as:

$$dQ_S = -q\vec{E}_W \cdot d\vec{r} \quad (6)$$

$$i_S = -q\vec{E}_W \cdot \vec{v} \quad (7)$$

where (7) can be obtained by dividing (6) by dt , so that \vec{v} is the velocity of q , and $\vec{E}_W = -\vec{\nabla}\phi_W$. It should be clear that ϕ_W is non-dimensional, while \vec{E}_W has units of inverse length. These two equations together constitute **Ramo's theorem**.

ΔQ_S and i_S are respectively called the *signal charge* and the *signal current*, and they are the charge and current that will be collected by the read-out electronics. They are related by:

$$\Delta Q_S = \int_0^t i_S(t') dt' = q(\phi_W(\vec{r}(t)) - \phi_W(\vec{r}(0))) \quad (8)$$

In case q is actually composed of a charge distribution, appropriate volume integrals of the charge density must be evaluated.

2.2 Structure of the Report

An examination of (7) reveals that, in order to be able to compute the electrical signals in a silicon detector, we must know the weighting field of the read-out electrodes and the dynamics of charged particles inside the semiconductor material, that is the charge motion under the influence of the electric fields inside the detector. This observation justifies the order in which the subject matter is laid out in the report:

1. Calculation of the weighting potential for a segmented electrode configuration;
2. Calculation of the electric field in a silicon microstrip detector;
3. Finite volume method calculation of the flow of a charge distribution in the detector;
4. Combination of the charge flow results and the electric and weighting potentials to generate current signals according to (7)
5. Signal processing by pre-amplifier module;
6. Consideration of radiation damage effects.

The theory required to understand the points enumerated above will be presented as the different topics are discussed.

2.3 Important Concepts in Numerical Methods

Prior to beginning the discussion of the project material, it is necessary to provide the definitions of some terms and concepts that will be used throughout the report.

Most of the problems that were tackled in this project involved the solution of a linear partial differential equation. In order to solve it numerically, it is necessary to approximate the derivatives that appear in the equation with finite differences evaluated at a finite number of grid points. In this way it is possible to express the value of the argument of the differential operator at a specific grid point in terms of its values at the neighbouring grid points. Since the original continuous equation is linear, the result of this procedure is a linear algebraic equation for each grid point, which taken together constitute a linear system of equations. In symbols:

$$L(\phi) = s \rightarrow A\vec{\phi} = \vec{b} \quad (9)$$

where L is a linear partial differential operator and s is the source function. The right hand side of (9) is the resulting set of *discretised equations*: the matrix A is called the *coefficients matrix*, \vec{b} is the *source vector*, and $\vec{\phi}$ is the vector of the values of the solution function evaluated at all grid points.

When the grid points are arranged in regular square lattice (2D), we call it a *structured mesh*. Discretised linear PDEs on structured meshes generate *sparse*

matrices which are often *n-diagonal*.

The system $A\vec{\phi} = \vec{b}$ is usually very large, so exact solution methods (e.g. Gaussian elimination) are not feasible. Therefore, approximate solution methods are employed. Such solution procedures commonly start from an initial guess, and then repeatedly apply a formula to correct this initial guess until the numerical solution approaches the *exact solution* (i.e. the solution that would be obtained from matrix inversion) within some specified tolerance. This kind of approach is known as an *iterative method* and it can be expressed mathematically as:

$$\vec{\phi}^{n+1} = B\vec{\phi}^n + M\vec{b} \quad (10)$$

where $\vec{\phi}^{n+1}$ is the updated solution, and B is called amplification matrix. This equation is devised so that the exact solution is a stationary point.

The separation of the approximate solution at iteration step n from the exact solution can be quantified using the so-called *global residual*:

$$\vec{R} = \vec{b} - A\vec{\phi}^n \quad \text{Vector global residual} \quad (11)$$

$$R = \sqrt{\vec{R} \cdot \vec{R}} \quad \text{Global residual} \quad (12)$$

The process of arriving at the solution is named *convergence*, and the system is solved when R lies below the prescribed tolerance.

Lastly, given the approximate solution at iteration step $n+1$, the system of equation reads: $A\vec{\phi}^{n+1} = \vec{b}$. Subtracting off $A\vec{\phi}^n$ from each side of the equation, the equation becomes:

$$A(\vec{\phi}^{n+1} - \vec{\phi}^n) = A\vec{\delta}^{n+1} = \vec{b} - A\vec{\phi}^n = \vec{R}^n \quad (13)$$

where \vec{R}^n is the vector global residual at iteration n , while $\vec{\delta}^{n+1}$ is the correction needed to update the solution. This form of the discretised equations is known as *correction equations*.

2.4 Overview of Solution Methods for Linear Systems of Equations

In the course of the project, there were two iterative solution methods that were preferentially applied to the solution of the discretised equations: the Gauss-Seidel method (GS) and the Conjugate Gradient method (CG). The GS method is based on the following matrix decomposition:

$$A\vec{\phi} = (L + D + U)\vec{\phi} = \vec{b} \quad (14)$$

where L is a strictly lower triangular matrix, U is a strictly upper triangular matrix and D is the diagonal.

As an iterative algorithm, (14) can be arranged as follows:

$$(L + D)\vec{\phi}^{n+1} = \vec{b} - U\vec{\phi}^n \quad (15)$$

where it is easy to see that the exact solution is a stationary point. In order to improve the convergence properties of (15), a so-called *over-relaxation parameter* is added to the expression (see section 4.4 for more insight):

$$(\omega L + D)\vec{\phi}^{n+1} = \omega(\vec{b} - U\vec{\phi}^n) + (1 - \omega)D\vec{\phi}^n \quad (16)$$

which is usually written in the more familiar form:

$$\vec{\phi}^{n+1} = (1 - \omega)\vec{\phi}^n + \omega D^{-1}(\vec{b} - U\vec{\phi}^n - L\vec{\phi}^{n+1}) \quad (17)$$

this equation goes by the name of Successive Over-Relaxation (SOR) iteration scheme, and it is stable for $0 < \omega < 2$ (see [9]).

The advantages of this method are the ease of implementation, the little workload per iteration, and its point-wise nature which can be very useful when working with unstructured meshes that model complex geometries. However, the degree of implicitness (i.e. the number of entries of $\vec{\phi}$ treated as unknowns against the ones treated as known values, in other words, the number of entries that are updated against the ones retained from the previous iteration step) of (17) is not high (around 1/2), so that the convergence rate of the SOR method is much worse in comparison to the CG method.

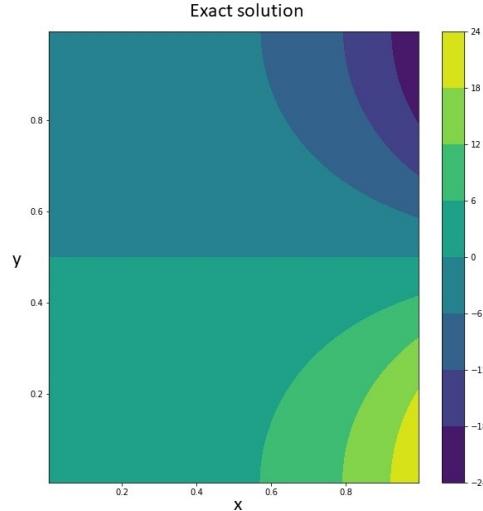


Figure 4: Exact solution.

The CG method is a fully implicit method that is based on the concept of fastest descent. The system $A\vec{\phi} = \vec{b}$ is turned into a quadratic form such that its minimum (or maximum) corresponds to the solution of the system. The solution is found by exploring this hyper-surface following the direction opposite to the maximum slope until the minimum is located. The details of the method are provided in chapter 3 of [6].

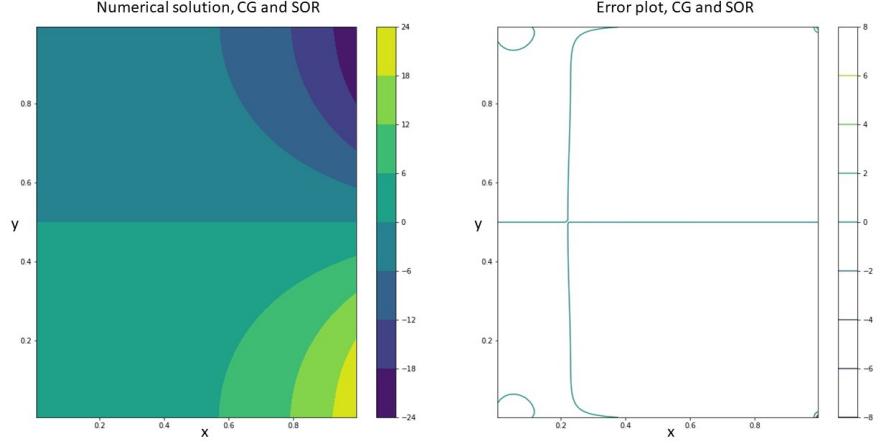


Figure 5: Numerical solution. (a) Output of numerical calculation, (b) error diagram.

This method possesses much better convergence properties in comparison to the SOR method, however, it has the downside of being applicable only to symmetric coefficients matrices. Nevertheless, for the solution of Laplace's equation for the weighting potential this is not a problem since the resulting coefficients matrix is symmetric.

In order to further study the convergence properties of these two solution methods, let's consider the solution of Laplace's equation, $\vec{\nabla}^2\phi = 0$, on a square region of space, $0 \leq x \leq 1$ and $0 \leq y \leq 1$, with boundary conditions: $\phi(x, 0) = e^{\pi x}$, $\phi(0, y) = \cos(\pi y)$, $\phi(x, 1) = -e^{\pi x}$, $\phi(1, y) = e^{\pi} \cos(\pi y)$. The exact solution to this problem is $\phi(x, y) = \cos(\pi y)e^{\pi x}$. The numerical solution was attempted with both methods on a structured mesh of size 0.01.

Fig. 4 shows the colour map of the exact solution, while fig. 5 displays the colour map and the error diagram of the numerical solutions obtained with a residual requirement of 10^{-8} . However, the most interesting graph for the purpose of understanding is fig. 6. It shows (y-axis is a log scale) the variation of the global residual with the iteration count of the two methods. The CG method is manifestly superior, reaching the solution in about 250 iterations against the approximately 1800 of the SOR algorithm. Another feature of the plot that should be underlined is the fact that the convergence rate of the CG method is not as good as the one of the SOR in the initial phases of convergence. This fact, along with the erratic nature of the CG algorithm, will prove to be very unfavourable in parallelised applications of this method.

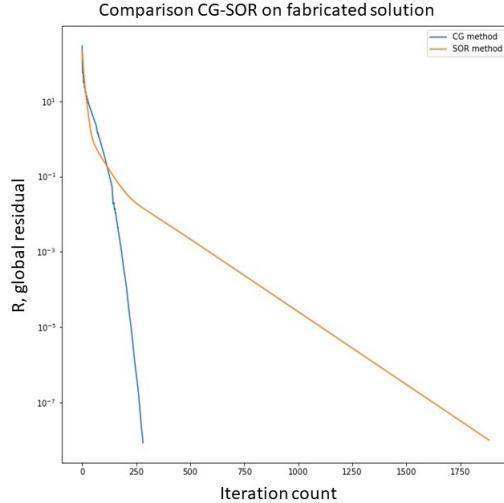


Figure 6: Residual plot, comparison of convergence of SOR and CG methods.

3 Calculation of the Weighting Potential

3.1 The Physical Set-up and the Discretised Equations

From fig. 2 in section 1, it is clear that the electrode configuration of a microstrip detector consists of a very large backplane electrode surmounted by an array of narrow metal strips. Since the detector can be regarded as infinitely long in the direction parallel to a strip, the problem of calculating the weighting potential is essentially two-dimensional in the plane perpendicular to the strips. This observation is valid for all the problems treated in this project, so from now on it will be regarded as a fact about the system.

Considering a typical strip situated far away from the extremities of the detector, the weighing potential can be calculated by adopting the following boundary conditions:

- the strip of interest is set to 1;
- all strips on its left and right are set to 0. Since the solution domain must be bounded on the left and right, only a limited number of them can be accounted for;
- the backplane electrode is set to 0;
- a surface to delimit the solution domain above the strips and far away from them is chosen and set to 0;
- the left and right bounding surfaces are set to 0 as well.

A schematic view of the solution domain considered is given in fig. 7 (a). Strictly speaking, the bounding surfaces that delimit the solution domain should be taken at infinity. Nevertheless, accurate results can be obtained by taking them sufficiently far away from the region of interest, so that their influence on the solution is limited. For this reason, these surfaces were selected at a distance that was 5 times the characteristic size in the relevant direction: horizontally 5 times the pitch, vertically 5 times the height.

At this point, the discretisation of the Laplacian operator on a structured mesh must be considered. Let's call h the mesh size, then, using the central difference approximation of the second derivative, Laplace's equation at a grid point (x, y) takes the form:

$$\frac{\phi(x+h, y) - 2\phi(x, y) + \phi(x-h, y)}{h^2} + \frac{\phi(x, y+h) - 2\phi(x, y) + \phi(x, y-h)}{h^2} = 0 \quad (18)$$

where $\phi(x, y)$ represents the weighting potential. Multiplying the equation by h^2 and indicating the point (x, y) with P , and the neighbouring grid points $(x+h, y)$, $(x-h, y)$, $(x, y+h)$, $(x, y-h)$ with E, W, N, S ; (18) reduces to:

$$\phi_S + \phi_W - 4\phi_P + \phi_E + \phi_N = 0 \quad (19)$$

which clearly gives rise to a symmetric coefficients matrix.

3.2 Implementation of the Solution Method

The solution domain shown in fig. 7 (a) is very large, hence, assuming a constant grid spacing, a sequential iteration over the entire domain would take a very long execution time. This issue calls for the parallelisation of the solution procedure: the domain is broken down into blocks and the iterative solver is applied in parallel across multiple blocks.

Initially, this kind of strategy was attempted with the CG method due to its appealing convergence properties. However, two major drawbacks were encountered: (i) a single iteration on a block was taking a lot of computing time because of the large number of long operations required in an iteration step, (ii) the communication between the blocks (i.e the fact that the boundary conditions of a block were changing after each iteration) was making the convergence behaviour of the method rather erratic, (iii) because of the slow reduction of residuals (see fig. 6) in the initial stages of convergence the method did not make any substantial progress after many iterations.

For these reasons, it was decided to switch to a parallelised SOR method. In fact, because of its pointwise nature, this method is very stable against constant changes of boundary conditions.

Fig. 7 (b) shows the division into blocks. The blocks containing a strip have a mesh size of $0.5\mu\text{m}$, while the empty blocks have a coarser mesh size of $1\mu\text{m}$. This allows to speed up the algorithm by reducing the number of operations in the regions where the potential is expected to vary slowly. The curly numbered arrows display the order in which the blocks were swept by the SOR solver.

The parallelisation of the process was handled using the functions associated

with the *Pool* object of the *multiprocessing* library in Python. The simple parallelisation that was carried out reduced the execution time of a single iteration from 160s (sequential mode) to 92s, an improvement of around 42.5%.

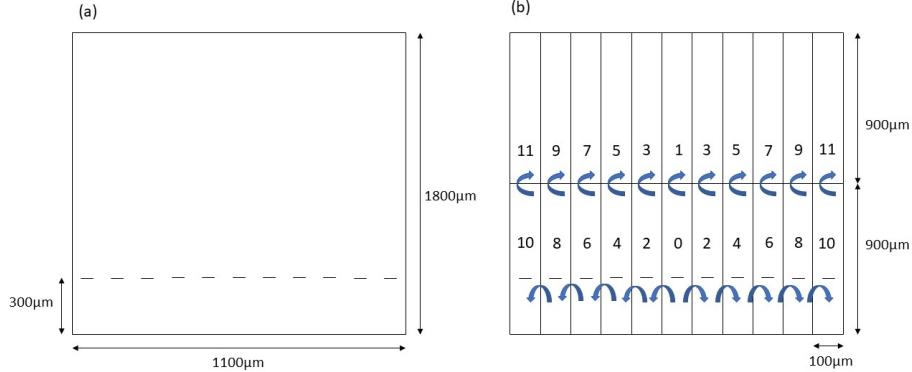


Figure 7: Solution domain. (a) Full solution domain for a segmented arrangement of electrodes, (b) division in blocks for parallelisation.

3.3 Results and Discussion

Fig. 8 shows the potential map that was obtained after 2000 iterations, $R = 10^{-3}$. The algorithm was interrupted at this point because the convergence rate had become incredibly slow, while the solution was already pretty satisfactory. In fact, it compared well with the results discussed in [19] and chapter 5 of [16]. Since the grid was swept from left to right, from bottom to top, the solution tended to being updated faster on the right region of the domain than it was on the left region. In other words, the boundary conditions propagated more promptly to the right than they propagated to the left in consequence of the adopted sweeping pattern. Therefore, when the potential was interpolated over a grid of size 0.1 μm, it was decided to reflect the potential on the right region to the left in order to obtain the final potential map. This also suggested that the introduction of a symmetry boundary condition at the symmetry axis of the set-up could have been exploited in the optimisation of the algorithm. As a matter of fact, later it was realised that the introduction of such boundary condition could have been used to restrict the solution domain to just half of the previously described set-up.

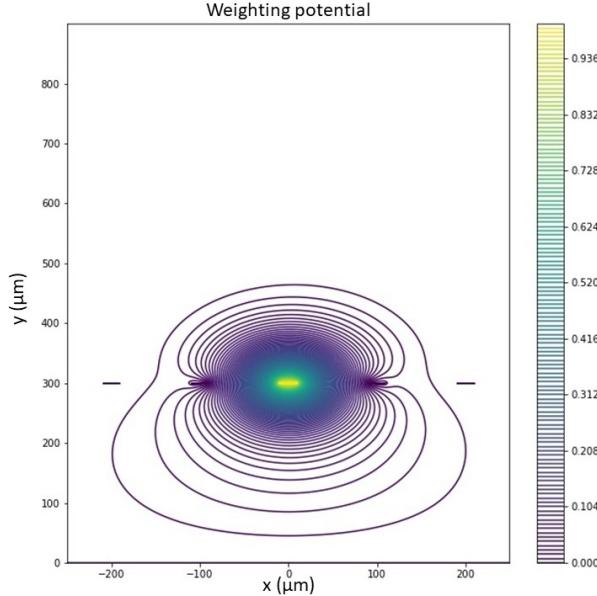


Figure 8: Numerical result for the weighting potential, 5 strips shown.

4 Calculation of the Electric Potential

4.1 Physics Background

Section 1 provided a general description of the structure of a silicon microstrip detector. For the purpose of particle detection, this device is always operated at a certain potential difference in order to facilitate charge collection and signal formation. Note, however, that a signal is formed even though the liberated electrons and holes did not actually reach the electrodes.

The mathematical modelling of the electric potential inside the detector can be understood from the analysis of a one-dimensional np-junction diode. Fig. 9 (a) shows a schematic view of the such diode, N_D denotes the donor density on the n-side, N_A the acceptor density on the p-side. If no external potential is applied to the diode, some electrons will still diffuse to the p-side in order to recombine with the holes. This diffusive motion causes the formation of a layer of positive charge density on the n-side, and a similar layer of opposite charge on the p-side. The accretion of this *depletion region* will come to a halt when the electric field originated by the space charge in the depletion region is sufficient to prevent further diffusion.

The electric field in the depletion region can be modelled using the so-called *junction approximation*, that is the assumption that the depletion layers on the p and n sides of the diode have a uniform charge density that terminates sharply. Figure 2.3 (b) in chapter 2 of [11] shows a more realistic charge distribution for the depletion region and it demonstrates the excellent validity of the junction

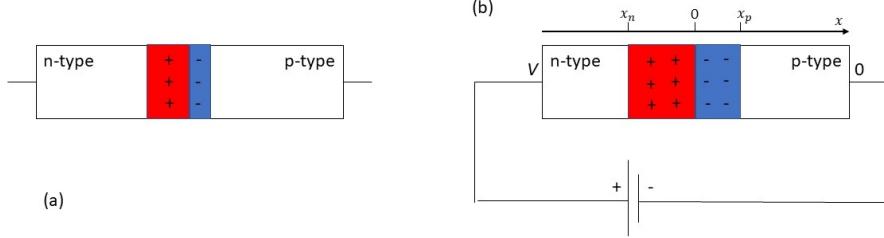


Figure 9: NP-junction. (a) No applied external potential, (b) external potential applied in reverse bias.

approximation.

If a positive voltage is applied on the n-side, while the p-side is grounded (a situation called *reverse bias*), the depletion region will increase in size. This phenomenon can be understood either in terms of energy band diagrams or in terms of drift-diffusion of the charge carriers in the semiconductor. Operating the microstrip detector in this condition is advantageous for two reasons: (i) a larger electric field inside the bulk material guarantees a faster collection time, (ii) the absence of the majority charge carriers reduces the noise in the signal. Therefore, the discussion that follows will focus on this type of arrangement.

Fig. 9 (b) shows the np-junction in reverse bias. The junction location is labelled as $x = 0$, while x_n and x_p denote the start and end of the depletion region on the n and p sides respectively. For this one-dimensional system the governing equation for the electric potential ϕ is:

$$\frac{d^2\phi}{dx^2} = -\frac{\rho}{\epsilon_r \epsilon_0} \quad (20)$$

that is the one-dimensional Poisson's equation, where the charge density distribution is given by the junction approximation with densities N_D and N_A . This equation can be solved by direct integration on the n and p sides respectively. The integration constants are defined by the boundary conditions $\phi(x_n) = V$, $\vec{E}(x_n) = 0$, $\phi(x_p) = 0$, $\vec{E}(x_p) = 0$. Then, the potential on the two sides of the junction turns out to be:

$$\phi(x) = -\frac{N_D e x^2}{2\epsilon_r \epsilon_0} + \frac{N_D e x_n x}{\epsilon_r \epsilon_0} + \frac{N_A e x_p^2}{2\epsilon_r \epsilon_0} \quad \text{n-side} \quad (21)$$

$$\phi(x) = \frac{N_A e x^2}{2\epsilon_r \epsilon_0} - \frac{N_A e x_p x}{\epsilon_r \epsilon_0} + \frac{N_A e x_p^2}{2\epsilon_r \epsilon_0} \quad \text{p-side} \quad (22)$$

using (21) with the charge neutrality requirement, $N_D|x_n| = N_A|x_p|$, allows to express the total width of the depletion region, $w = |x_n| + |x_p|$, in terms of the applied potential V . The result is:

$$w = \sqrt{\frac{2\epsilon_r \epsilon_0 V}{e} \frac{N_D + N_A}{N_D N_A}} \quad (23)$$

The potential required to totally deplete the junction (i.e. the width of the depletion region is equal to the size of the detector) is called the *bias voltage*. If the applied potential exceeds the bias voltage the potential difference $V_{applied} - V_{bias}$, only contributes as a superimposed electric field that is independent of the charge density.

Therefore, it should be evident that the electric potential calculation in the silicon detector must be divided into two parts:

- the calculation of the electric potential at the bias point under the junction approximation using zero electric field boundary conditions at the silicon-electrode interfaces;
- the calculation of the electric potential in the absence of charge density setting the backplane electrode at a potential V and the strip to ground.

4.2 The Discretised Equations and the Numerical Set-up

The set-up that was taken into consideration for the calculation of the electric potential is shown in fig. 10. Only one strip is accounted for because the presence of the other strips can be dealt with periodic boundary conditions applied to the left and right surfaces of the solution domain, which correspond to the surfaces midway between two strips. Such an arrangement is suitable for the calculation of the electric potential of a strip far removed from the extremities of the detector.

When the potential is at or exceeds the bias point, the silicon detector presents a positive space charge density throughout the n-type bulk and a negative charge density in the p+ implant. The donor density in the n-bulk defines the uniform charge density in this region of the material. On the other hand, the depletion layer in the p+ implant is very thin because of its extremely high acceptor density. Regardless, the charge density of the implant was also modelled as a uniform one given that the mesh resolution was in no way sufficient to resolve that detail.

This implies that the governing equation inside the silicon is Poisson's equation, while outside the silicon (i.e. the free space above the silicon slab) it is Laplace's equation. Expanding on the results of section 3.1, it is easy to see that the discretised version of Poisson's equation arising from a central difference approximation is:

$$\phi_S + \phi_W - 4\phi_P + \phi_E + \phi_N = -(h^2\rho)/(\epsilon_r\epsilon_0) \quad (24)$$

Equations (19) and (24) describe the bulk of the solution domain and the interfaces with the metal electrodes, yet they must be modified at the silicon-air interfaces in order to account for the discontinuity in permittivity. This is easily done by considering the \vec{D} -field at those interfaces.

The divergence equation for \vec{D} is $\vec{\nabla} \cdot \vec{D} = \rho$. Integrating over the volume of a cell (a cube in a structured mesh), applying the divergence theorem and using

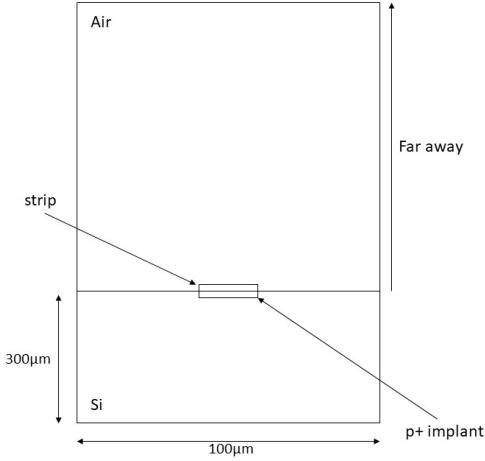


Figure 10: System considered in calculation of the electric potential.

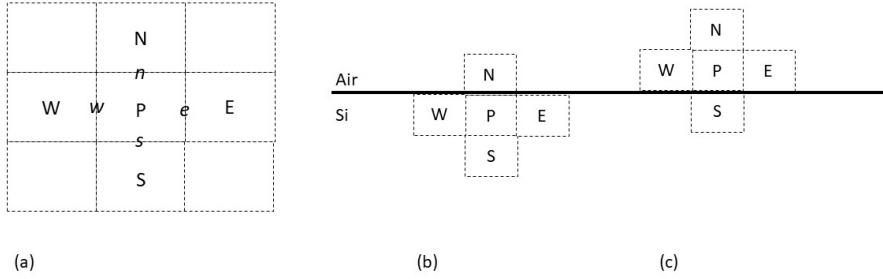


Figure 11: Closer look to the structured mesh geometry. (a) Labelling convention of cell centres and faces, (b) silicon-air interface, (c) air-silicon interface.

$\vec{D} = \epsilon_r \epsilon_0 \vec{E}$ along with $\vec{E} = -\vec{\nabla} \phi$, the equation becomes:

$$\begin{aligned} & \epsilon_n \left(-\frac{\phi_N - \phi_P}{h} \right) h^2 - \epsilon_s \left(-\frac{\phi_P - \phi_S}{h} \right) h^2 + \\ & + \epsilon_e \left(-\frac{\phi_E - \phi_P}{h} \right) h^2 - \epsilon_w \left(-\frac{\phi_P - \phi_W}{h} \right) h^2 = \rho h^3 \end{aligned} \quad (25)$$

where each term on the left side of the equation is the flux of \vec{D} across the cell faces. The gradients of ϕ , evaluated using the values at the grid points, define the relevant electric field components at the cell faces, while the small letters indexing ϵ indicate the permittivity at the cell faces (see fig. 11 (a)).

Evaluating (24) at the silicon-air interface (fig. 11 (b)) and dividing both sides

by ϵ_r , ϵ_0 and h , the expression can be rearranged as:

$$\phi_S + \phi_W - \left(3 + \frac{1/\epsilon_r + 1}{2} \right) + \phi_E + \frac{1/\epsilon_r + 1}{2} \phi_N = -\frac{h^2 \rho}{\epsilon_r \epsilon_0} \quad (26)$$

where the discontinuity of the permittivity at the northern face (ϵ_n) is approximated by its average value across the cells P and N .

The air-silicon interface (fig. 11 (c)) is handled in a similar manner by setting ρ to 0 and without carrying out the division by ϵ_r since outside the material there are no sources. This yields the following equation:

$$\frac{\epsilon_r + 1}{2} \phi_S + \phi_W - \left(3 + \frac{\epsilon_r + 1}{2} \right) + \phi_E + \phi_N = 0 \quad (27)$$

Equations (26) and (27) together characterise the discontinuity in permittivity on each side of the material boundary.

At this stage, all the equations required for the evaluation of the electric potential in the silicon detector are discretised and the remaining task consists of identifying the appropriate boundary conditions at the electrode surfaces.

4.3 The Boundary Conditions at the Electrodes

As explained in section 4.1, the calculation of the electric potential in the detector is composed of the two listed steps.

The second one is simple to carry out. The backplane electrode is set to $V = 1$, the metal strip is grounded, while a surface far away from the detector is chosen to bound the solution domain and it is set to 0. In this situation, the governing equation is effectively Laplace's equation since this component of the potential does not depend on the charge density, nevertheless equations (26) and (27) must still be included in the coefficients matrix in order to properly account for the material discontinuity. The backplane surface was set to unity because, by linearity of the equation, the potential distribution corresponding to the over-bias voltage could be obtained from the normalised map through the multiplication by the amount the applied voltage exceeded the bias point. By contrast, the calculation of the potential at the bias point requires more attention. The no- \vec{E} -field boundary condition at the backplane electrode is straightforward to implement through the requirement that the electrode surface and the cell just above it must have the same potential. This modifies (24) as follows:

$$\phi_W - 3\phi_P + \phi_E + \phi_N = -(h^2 \rho) / (\epsilon_r \epsilon_0) \quad (28)$$

where P corresponds to a cell situated above the backplane electrode.

A similar treatment should be reserved for the strip electrode, yet such a simple treatment has some unwanted implications. Firstly, the presence of highly dense negative space charge in the p+ implant translates into the fact that the potential will vary quite rapidly across this region of the detector. This implies that a condition similar to (28) applied there would result in a non-uniform potential across the electrode which is physically incorrect since the strip is made

of a conductor. This issue was resolved by setting the electrode to the average potential of the p+ implant at the expense of relaxing the strict no-field condition. Nonetheless, such an approach introduces the constraint that the zero-point potential is now not 0, but some negative value set by the charge in the p+ implant. Therefore, to generate a realistic potential map, the bounding surface at infinity must be set to the same negative potential of the strip.

Despite the fact that the requirements that have emerged thus far are physically and mathematically sound, they constitute a hindrance for the numerical simulation. The reason is that now a 0-everywhere initial guess is bound to be terribly distant from the exact solution, which in turn means that it will take a long time to the iterative solver to reach the solution.

This uninviting prospect can be prevented, on one side, through an approximation of the detector set-up, and, on the other, through the implementation of a solution procedure capable of speeding up the SOR method (CG cannot be used here because the material discontinuity introduces non-symmetric terms). The latter point will be the object of the next section. The former consists of modifying the boundary condition at the p+ implant so that $V = 0$ is again the zero-point potential.

The approximation is based on the fact that, since the negative charge determines the zero-point potential of the set-up, it is reasonable to directly set the potential of the p+ implant and the strip electrode to 0. Then, the boundary condition at infinity can be set to 0 as well. Such an approach discards the presence of the negative charge altogether, yet it is justified on the ground that the mesh resolution adopted cannot really resolve the details of the charge distribution within the p+ implant, and that the small dimension of the implant will limit the introduction of errors in the macroscopic potential distribution. Furthermore, this choice is corroborated by figure 13 (a) in [17] which shows the electrostatic potential in a one-sided microstrip detector where both the strip and the implant are at the same 0 potential.

A calculation of the electric potential with the “more realistic” boundary conditions was attempted. The results are presented in appendix H.

4.4 Von Neumann Error Analysis and Geometrical Multigrid Methods

This section constitutes a short mathematical digression needed for the introduction of the Geometrical Multigrid Method (GMG) that was developed to improve the convergence rate of the SOR algorithm.

Start from (10) and subtract from each side $\vec{\phi}^E = B\vec{\phi}^E + M\vec{b}$, where $\vec{\phi}^E$ is the exact solution. The outcome of this operation is:

$$\vec{\epsilon}^{n+1} = B\vec{\epsilon}^n \quad (29)$$

where $\vec{\epsilon}^n = \vec{\phi}^n - \vec{\phi}^E$ is the error vector at iteration n .

In continuous space the error is a continuous function. Restricting the attention to the one-dimensional case over a domain of length L , the error function can

be expanded as a complex Fourier series:

$$\epsilon(x) = \sum_{m=0}^{\infty} C_m e^{i \frac{m\pi}{L} x} \quad (30)$$

where Dirichlet boundary conditions (in this case the error at the domain boundaries is zero by definition) or periodic boundary conditions are assumed at the boundary nodes.

Since the mesh is discrete, the error function is effectively sampled at $N+1$ points. Therefore, given a mesh size h , a domain length $L = Nh$ and denoting the node locations as $x_j = jh$, the series becomes:

$$\epsilon^n(x_j) = \sum_{m=0}^N C_m^n e^{i \frac{m\pi}{N} j} \quad (31)$$

which is the *discrete Fourier series* of the error function at iteration n . Only $N+1$ modes are retained since only $N+1$ sample points are available to evaluate the amplitudes.

From the inspection of (29), it is evident that (31) evaluated at node j is the j -th entry of the error vector. It often occurs that $\vec{\epsilon}^n$ expressed as in (31) is actually an eigenvector of the amplification matrix, and even if it is not, the error expressed in this form provides many useful insights concerning the error damping mechanism of an iterative solution method.

Let's assume that B is tri-diagonal, a rather common situation for amplification matrices arising from iterative methods applied to coefficients matrices describing one-dimensional PDEs, and let's call the diagonal, upper off-diagonal and lower off-diagonal entries in the j -th row d_j , u_j and l_j respectively. Substituting (31) in (29), the j -row becomes:

$$\sum_{m=0}^N C_m^{n+1} e^{i \frac{m\pi}{N} j} = \sum_{m=0}^N (l_j e^{-i \frac{m\pi}{N}} + d_j + u_j e^{i \frac{m\pi}{N}}) C_m^n e^{i \frac{m\pi}{N} j} \quad (32)$$

which, if it is to be satisfied, leads to the following requirement on the amplitudes:

$$\frac{C_m^{n+1}}{C_m^n} = d_j + l_j e^{-i \frac{m\pi}{N}} + u_j e^{i \frac{m\pi}{N}} \quad (33)$$

which is a ratio called *amplification factor*. It can be shown that $|C_m^{n+1}/C_m^n| \leq 1$ for all wavelength components of the error is a necessary condition for the convergence of an iterative method. Furthermore, the ratio between the largest and the smallest amplification factors (the so-called *condition number*, κ) is a measure of the rate of convergence of an iterative method: the closer κ is to 1, the faster the algorithm. This statement is generally valid for amplification matrices arising from physical PDEs, where a κ close to 1 is a manifestation of a coefficients matrix very nearly diagonal, and so very easily invertible.

If we apply the same procedure to the pure GS method, adopting the same

nomenclature, and paying attention to the fact that l_j and d_j are now implicit coefficients, the amplification factor becomes:

$$\frac{C_m^{n+1}}{C_m^n} = -\frac{u_j e^{i \frac{m\pi}{N}}}{d_j + l_j e^{-i \frac{m\pi}{N}}} \quad (34)$$

where the coefficients now refer to the coefficients matrix, A , and not the amplification matrix, B . Calculating the modulus of (34) for the discretised one-dimensional Laplace's equation, the result is:

$$\left| \frac{C_m^{n+1}}{C_m^n} \right| = \sqrt{\frac{1}{5 - 4\cos(\theta_m)}} \quad \text{where } \theta_m = \frac{m\pi}{N} \quad (35)$$

hence, it is easy to see that the smallest amplification factor, $1/9$ at $\theta_m = \pi$, corresponds to the shortest wavelength component , whilst the largest amplification factor is associated to the longest wavelength component, $\theta_m = \pi/N$. Consequently, the short wavelength error components will be damped out quickly by the algorithm in contrast to the long wavelength components that require many more iterations to die out. This behaviour arises from the fact that the long wavelength components are attached to the boundary nodes only, which makes their reduction more computationally expensive.

In addition, the largest amplification factor increases as the mesh is refined, while the smallest one remains the same, implying that, as the mesh is refined, the condition number increases, thus slowing down the rate of convergence, and vice-versa if the mesh is made coarser. This observation stands at the heart of GMG methods, which, as the name suggests, employ the solution of the system over multiple grids, one coarser than the other, in order to speed up the convergence rate of the selected iterative solver.

The explanation of the GMG concept was conducted using the GS method, yet its general features are valid for all classic iterative solvers.

4.5 Implementation of the solution method

In the case of the component of the potential resulting from over-bias, the simple SOR method was applied to the domain shown in fig. 12 (a). The bottom block had a mesh size of $0.5\mu\text{m}$, while the top block a mesh size of $1\mu\text{m}$. This problem converged relatively quickly since there were no sources and the potential at the backplane electrode was already set by Dirichlet boundary conditions.

On the other hand, the calculation of the bias potential was carried out using a GMG method on two grids on the domain shown in fig. 12 (b). The fine mesh had a size of $0.5\mu\text{m}$, while the coarse mesh had a size of $1\mu\text{m}$. In the context of a GMG approach the iterative solver, in this case the SOR method, is generally referred as the *smoother*.

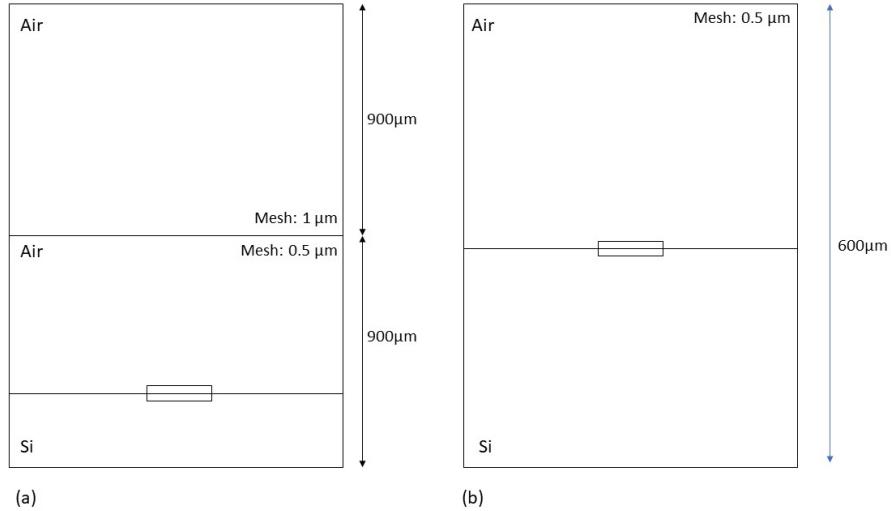


Figure 12: Solution domain for the electric potential calculation. (a) For over-bias potential, (b) for bias-point potential.

Then, the solution procedure involved the following steps:

1. the generation of the coefficients matrices for the fine, A_f , and coarse, A_c , meshes;
2. smoothing of the initial guess on the fine mesh ($\sim 30\text{-}50$ iterations), that is solving $A_f \vec{\phi}_f = \vec{b}$ to partial convergence, where $\vec{\phi}_f$ is solution vector on the fine mesh;
3. *restriction* of the vector global residual on the fine mesh, \vec{R}_f , to the coarse mesh: this consists of the interpolation of \vec{R}_f on the points corresponding to the coarse mesh in order to obtain the vector global residual on the coarse mesh, \vec{R}_c ;
4. solution to partial convergence ($\sim 60\text{-}120$ iterations) of the correction equation on the coarse mesh, $A_c \vec{\delta}_c = \vec{R}_c$;
5. *prolongation* of the correction vector from the coarse mesh to the fine mesh: $\vec{\delta}_c$ is interpolated on the points corresponding to the fine mesh to obtain the correction vector on the fine mesh, $\vec{\delta}_f$;
6. correction of solution vector on the fine mesh, $\vec{\phi}_f$, with the prolonged correction vector, $\vec{\delta}_f$;
7. further smoothing on the fine mesh

8. points 3, 4, 5, 6, 7 are repeated until the global residual requirement on the fine mesh is met.

The restriction and prolongation steps were performed using the cubic spline (local cubic polynomial interpolation) option of the `interpolate.RectBivariateSpline` object of the `interpolate` class in the `scipy` library.

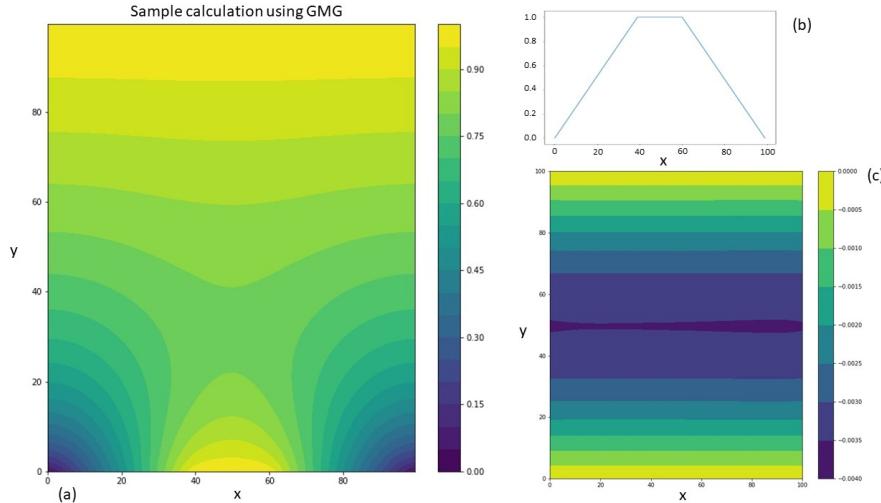


Figure 13: Sample calculation using GMG method, (a) Result, (b) boundary condition on the bottom surface, (c) difference between GMG-SOR and simple SOR solution.

The efficiency gain can be appreciated by comparing the performance of the GMG-SOR method against the one of the simple SOR method in solving the problem whose result is displayed in fig. 13 (a). The equation of interest is Laplace's equation on a rectangular region of space of size 100×100 with periodic boundary conditions on the left and right, $\phi = 1$ on top and the potential distribution in fig. 13 (b) on bottom. The residual requirement was 10^{-4} and the mesh size 0.5.

The normal SOR solver arrived at the solution after 5430 iterations, while the GMG-SOR method arrived at the solution after 962 iterations on the fine mesh, and 1900 on the coarse mesh (proportion of coarse to fine iterations is 2 : 1). A full SOR iteration over the entire mesh involves $11N$ (where N is the number of grid points) long operations (5N in the GS iteration and 6N in the residual calculation). Therefore, assuming that the the restriction and prolongation steps take $20N$ iterations each, it means that the total number of operations executed in the normal SOR procedure is $(5430 \times 11)N = 59730N$, while in the GMG-SOR procedure this number is $(11 \times 962 + 20 \times 19 + (11/4) \times 1900)N = 16187N$, which corresponds to a reduction in the operation count of $\sim 73\%$.

4.6 Results and discussion

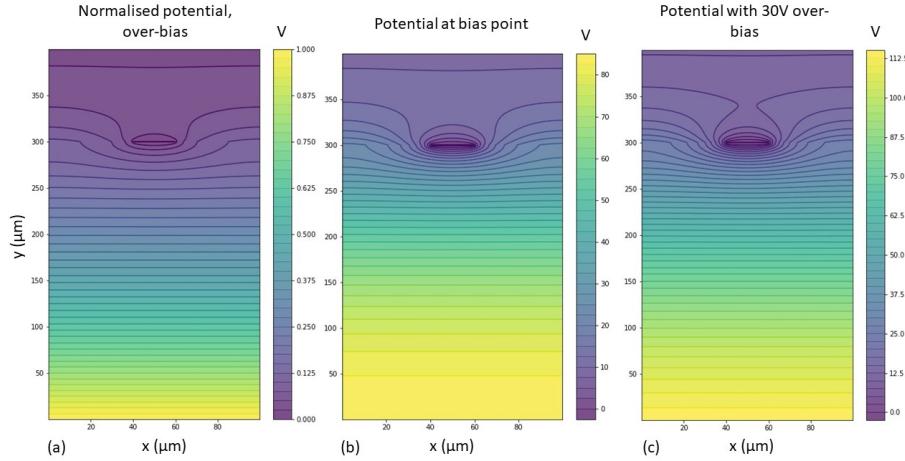


Figure 14: Electric potential results. (a) Normalised over-bias potential, (b) bias-point potential, (c) total potential with 30V over-bias.

Fig. 14 (a) displays the result obtained for the normalised electric potential arising from over-bias, while fig. 14 (b) shows the potential obtained at the bias point under the approximations discussed in section 4.3. Fig. 14 (c) superimposes the two maps to obtain the potential distribution in the detector when the applied potential exceeds the bias point by 30V.

The calculation of the over-bias potential required 5000 SOR iterations to reduce the initial residual of 2.6678 to a value of 0.0007. On the other hand, the computation of the bias-point potential was much more computationally expensive. The GMG-SOR solver performed 19000 iterations over the fine mesh and 78300 iterations on the coarse one to decrease the residual from 39.482 to a value of 0.998. It must be noted that, due to the interaction between the two meshes, the residual reached peaks of few hundreds (which were damped in \sim 2-5 iterations) during the convergence process. This fact indicates that the initially low residual was very likely a consequence of the zero-potential guess throughout the domain. Moreover, it must be pointed out that the extremely slow convergence was mainly dictated by the 0-field condition at the backplane electrode. In fact, since the potential was not fixed by a Dirichlet-type condition, the algorithm had to “build the potential up” through the iterative process, which required a very large number of iteration steps.

Fig. 15 (a) shows a slice of the potential in fig. 14 (b) through the symmetry axis ($x = 50$ line) of the microstrip detector. As predicted by (21), the potential varies quadratically within the n-type bulk, and, as expected, the p+ implant is barely resolvable, which suggests that the adoption of the above

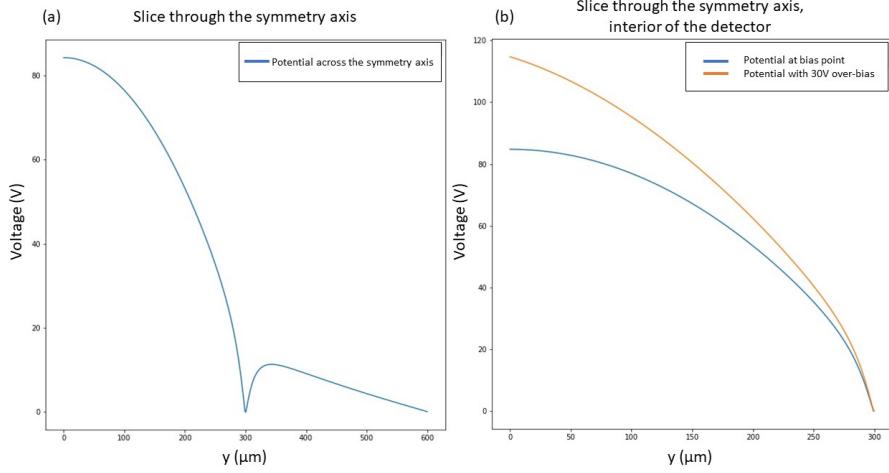


Figure 15: Slices through the symmetry axis of the potential. (a) Bias-point potential, (b) comparison of bias-point potential and total potential within the detector.

mentioned approximations did not introduce any noticeable error to the macroscopic structure of the potential. Assuming that $N_D \ll N_A$, (23) becomes $w \approx \sqrt{(2\epsilon_r \epsilon_0 V)/(eN_D)}$. Setting w equal to the detector size and solving for V , it is possible to estimate the bias potential, which turns out to be about 70V. The numerical simulation at the bias-point predicted a backplane potential of 84.2951V, which is not very different from the estimated one. The difference can be traced back to the deviation of the microstrip detector geometry from the one of a simple one-dimensional np-junction.

Lastly, fig. 15 (b) shows the comparison of the same slice in the interior of the detector for the potential maps in fig. 14 (b) and (c). From this, it is clear that the backplane potential in fig. 14 (c) is 114.2951V.

5 The Finite Volume Method Calculation of the Charge Motion

5.1 Physics Background

The electric potential that was calculated in the previous section determines the motion of the electron-hole pairs liberated by a particle incident on the detector. Nevertheless, these charges also undergo a process of diffusion as a result of the random component of their motion caused by the scattering with like particle and the lattice sites of the silicon. Therefore, their motion is properly described by a charge concentration subject to convective motion due to the electric field within the detector (also known as *drift*) and diffusive motion. This kind of problem can be efficiently treated using the Finite Volume Method (FVM) that is outlined in the following sections. Before that, some relevant physical concepts are in line.

According to Drude's model, the drift velocity of a positive charge inside a conductive medium is given by:

$$\vec{v} = \mu \vec{E} \quad (36)$$

where \vec{v} is the particle velocity, \vec{E} is the electric field and μ is a proportionality constant known as *mobility*. A negative charge will have the opposite velocity. This kind of description is particularly suitable for a semiconductor since the concentration of charge carriers is usually rather low.

On the other hand, in the absence of an electric field, a charge concentration will tend to spontaneously spread out. This spreading can be characterised by a *diffusion coefficient*, D .

The parameters μ and D are linked by the following relation, known as *Einstein relation*:

$$\frac{D}{\mu} = \frac{k_B T}{e} \quad (37)$$

where T is the temperature, k_B Boltzmann constant and e the electron charge. This expression can be derived by considering a semiconductor with non-uniform doping under equilibrium conditions. At equilibrium the chemical potential is constant throughout the material. This fact, along with the statistical mechanics of a semiconductor, allows to prove (37).

Now these facts can be put together to write down the continuity equation for the charge carriers within a semiconductor. Denoting the hole and electron concentrations as p and n , the equations are:

$$\frac{\partial p}{\partial t} = -\vec{\nabla} \cdot (\mu_h \vec{E} p - D_h \vec{\nabla} p) \quad (38)$$

$$\frac{\partial n}{\partial t} = -\vec{\nabla} \cdot (-\mu_e \vec{E} n - D_e \vec{\nabla} n) \quad (39)$$

where the terms within the divergence are flux terms. In particular, the first is the convective flux term, while the second is the diffusive flux term. Notice that

the difference in charge between electrons and holes is marked by the opposite sign of the convective terms.

The FVM revolves around the discretisation of these two equations.

5.2 The FVM discretisation of the continuity equation

The FVM describes the steady state or the time evolution of a system, in this case the charge concentration, by applying the continuity equation to many small *control volumes*, CVs, which cover the entire solution domain. The region of interest is divided up into many cells (the CVs), as we have already done in the previous sections, and then, by knowing the convective velocity field and the diffusion coefficient, the concentration flux through the faces of each CV is calculated. This allows to keep track of how the concentration will move around in the system. For the problem at hand, the velocity field is specified by the electric field according to (36), while the diffusive flux is described by the coefficient D .

The FVM discretisation procedure will be demonstrated for the holes continuity equation, but an almost identical equation, up to a minus sign, can be obtained for the electrons by following the same steps.

Assuming a structured mesh and following the same labelling convention as in fig. 11 (a), (38) is integrated over a CV. Supposing the concentration remains constant over the cell and applying the divergence theorem, the equation becomes:

$$\Delta V \frac{\partial p_P}{\partial t} = - \oint_{S_{CV}} (\mu_h \vec{E} p - D_h \vec{\nabla} p) \cdot d\vec{S} \quad (40)$$

where $\Delta V = h^3$, with h the mesh size as usual.

The only flux components that will contribute to (40) are the ones parallel to the normal vectors to the CV faces. Then, assuming that the flux remains constant across a face, the integral can be written as:

$$\begin{aligned} \Delta V \frac{\partial p_P}{\partial t} = & -\mu_h (E_n p_n - E_s p_s + E_e p_e - E_w p_w) A + \\ & + D_h ((\vec{\nabla} p)_n - (\vec{\nabla} p)_s + (\vec{\nabla} p)_e - (\vec{\nabla} p)_w) A \end{aligned} \quad (41)$$

where the first group of terms are the convective fluxes though the faces, the second group is made of the diffusive fluxes and A is the area of a CV face. E_n , E_s , E_e , E_w are the electric field components normal to the northern, southern, eastern, and western faces of the CV, p_n , p_s , p_e , p_w are the concentrations at those faces, while $(\vec{\nabla} p)_n$, $(\vec{\nabla} p)_s$, $(\vec{\nabla} p)_e$, $(\vec{\nabla} p)_w$ are the concentration gradients at the corresponding faces.

Moving on, the electric field components can be evaluated using a central difference approximation of the negative gradient of the potential as in (25). This approximation uses the potentials at the nodes bracketing a given face and it is second order accurate. The same treatment is reserved for the concentration gradients in the second group of terms. Dividing through A the equation becomes:

$$h \frac{\partial p_P}{\partial t} = \mu_h \left(\frac{\phi_N - \phi_P}{h} p_n + \frac{\phi_S - \phi_P}{h} p_s + \frac{\phi_E - \phi_P}{h} p_e + \frac{\phi_W - \phi_P}{h} p_w \right) +$$

$$+D_h \left(\frac{p_N - p_P}{h} + \frac{p_S - p_P}{h} + \frac{p_E - p_P}{h} + \frac{p_W - p_P}{h} \right) \quad (42)$$

which can be further arranged as:

$$\begin{aligned} h \frac{\partial p_P}{\partial t} &= \frac{\mu_h}{h} (a_n p_n + a_s p_s + a_e p_e + a_w p_w) + \\ &+ \frac{D_h}{h} (p_S + p_W - 4p_P + p_E + p_N) \end{aligned} \quad (43)$$

where the coefficients a_n, a_s, a_e, a_w are defined respectively as $\phi_N - \phi_P, \phi_S - \phi_P, \phi_E - \phi_P, \phi_W - \phi_P$. Note that, unsurprisingly, the diffusive terms discretise like a Laplacian operator.

Finally, the time discretisation of the equation is carried out using a Crank-Nicolson (CN) scheme, that is a mid-point rule such that the time derivative is substituted by a forward difference, whilst the other concentration terms are approximated by their average between time steps t and $t + \Delta t$. The result is:

$$\begin{aligned} &\left(\frac{\mu_h}{2h} (a_n p_n + a_s p_s + a_e p_e + a_w p_w) + \frac{D_h}{2h} (p_S + p_W - 4p_P + p_E + p_N) - \frac{h}{\Delta t} p_P \right)_{t+\Delta t} = \\ &= - \left(\frac{\mu_h}{2h} (a_n p_n + a_s p_s + a_e p_e + a_w p_w) + \frac{D_h}{2h} (p_S + p_W - 4p_P + p_E + p_N) + \frac{h}{\Delta t} p_P \right)_t \end{aligned} \quad (44)$$

where the terms on the right hand side compose a source term that is evaluated using the concentration values at the previous time step.

The CN scheme is implicit, which guarantees the stability of the method, and it is second order in the time step, which ensures accuracy.

5.3 The Discretisation of the Convective Terms

In the previous section the discretisation of the electric field, of the diffusive flux and the time discretisation were carried out. Nonetheless, the charge concentration at the CV faces was left untouched, no interpolation using the node values of the concentration was attempted. The reason for that stems from the fact that the discretisation of these terms requires more attention since the choice of a discretisation method over another has strong implications on the stability of the resulting numerical scheme.

In the following discussion, the one-dimensional convection-diffusion equation:

$$\frac{\partial \xi}{\partial t} = -\frac{\partial(u\xi)}{\partial x} + D \frac{\partial^2 \xi}{\partial x^2} \quad (45)$$

where u is the convective velocity, will be used as the model for the study of the properties of the various candidate discretisation methods. Furthermore, a structured mesh will be assumed throughout.

First of all, chapter 5 of [1] presents two important features that any convective discretisation scheme should possess, namely *conservativeness* and *transportiveness*. The former describes the fact that a good discretisation scheme returns the same flux when evaluated on the right (outflux) and the left (influx) of a

given cell face, in other words the flux leaving a CV is the same as the flux entering the adjacent CV that shares the same face. The latter indicates the fact that the discretisation scheme is upwind biased, that is the interpolation of the concentration at the cell face weights the upwind cell more than the downwind one.

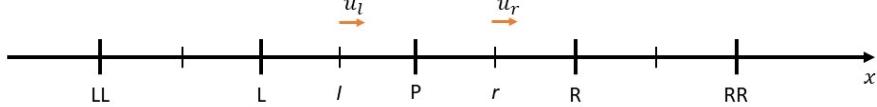


Figure 16: One dimensional structured mesh

Keeping this in mind, given that throughout the project extensive use of the central difference (CD) approximation was made, it is natural that this would be the first candidate method. It is evident that the CD interpolation at the CV face consists of a simple linear interpolation. Referring to fig. 16, the interpolation at the left and right faces of cell P reads:

$$\begin{aligned}\xi_l &= \frac{\xi_L + \xi_P}{2} \\ \xi_r &= \frac{\xi_P + \xi_R}{2}\end{aligned}\quad (46)$$

then the Euler explicit discretised convection-diffusion equation becomes:

$$\frac{\xi_P(t + \Delta t) - \xi_P}{\Delta t} = -\frac{u_r \xi_R + (u_r - u_l) \xi_P - u_l \xi_L}{2h} + D \frac{\xi_R - 2\xi_P + \xi_L}{h^2} \quad (47)$$

where u_r and u_l are the velocities at the cell faces, and, if not otherwise stated, ξ is evaluated at t .

If the diffusive terms are differentiated with respect to ξ_P , the result is:

$$\frac{\partial(\text{diffusive terms})}{\partial \xi_P} = -\frac{2D}{h^2} \quad (48)$$

This quantity describes the *sensitivity* [3] of this group of terms to small changes in the central cell. Since it is always negative, it means that small perturbations are always damped out, so that the CD scheme guarantees the numerical stability of this group of terms. Reason why this was the method of choice in section 5.2.

When the sensitivity of the convective terms is evaluated, the result is:

$$\frac{\partial(\text{convective terms})}{\partial \xi_P} = -\frac{(u_r - u_l)}{2h} \quad (49)$$

which can also be either positive or negative. Hence, numerical errors can either be damped or enhanced, which undermines the stability of a CD-based scheme.

The overall stability of the discretised equation is a combination of the effects of (49) and (50), where the presence of diffusion always has a stabilizing effect. To gain some more insight about the stability of CD, the Von Neumann stability analysis of (48) was carried out under the simplification $u_r = u_l$. The amplification factor of the error from a time step to the next was found to be:

$$\left| \frac{C_m^{t+\Delta t}}{C_m^t} \right|^2 = \left(1 - 4d \sin^2 \left(\frac{\theta_m}{2} \right) \right)^2 + c^2 \sin^2(\theta_m) \quad (50)$$

where $d = (D\Delta t)/h^2$ and $c = (u\Delta t)/(2h)$.

From this expression, it is easy to see that $0 < c < 1$ and $0 < d < 1/2$ are necessary conditions for stability. Moreover, it is not difficult (yet very algebraically tedious) to show that if $(uh)/D > 2$ (which can be identified as a condition on the Reynolds number), then there is maximum of the amplification factor in the interval $0 \leq \theta_m \leq \pi$ which is greater than 1, making the scheme unstable.

On the other hand, if $(uh)/D < 2$, the extrema occur at the boundaries of the interval and the scheme is stable for $0 < d < 1/2$. Combining the two inequalities, the condition on the time step for stability is $0 < \Delta t < (2D)/u^2$. Which implies that the CD method is unstable for inviscid flow, that is drift-only motion of the charges.

Given a detector potential of 100V and assuming a uniform field, then the CD-Euler explicit scheme applied to the motion of electrons requires a time step of 0.0034ns for stability ($\mu_e = 135 \text{ } \mu\text{m}^2\text{V}^{-1}\text{ns}^{-1}$, $T = 300 \text{ K}$, $h = 0.1 \mu\text{m}$).

The instabilities of the CD method mostly arise from the fact that the interpolation scheme is not upwind biased, therefore to overcome its limitations transportive methods must be considered.

The simplest of them is the Upwind Differencing (UD) discretisation method. Referring to the r face in fig. 16, this approach reads:

$$\begin{aligned} \xi_r &= \xi_P & \text{for } u_r > 0 \\ \xi_r &= \xi_R & \text{for } u_r < 0 \end{aligned} \quad (51)$$

which is a first order accurate interpolation that simply approximates the concentration at the CV face with the concentration of the upwind cell.

Assuming that $u_l > 0$ and $u_r > 0$, the discretised convection diffusion equation then reads:

$$\frac{\xi_P(t + \Delta t) - \xi_P}{\Delta t} = -\frac{u_r \xi_P - u_l \xi_L}{h} + D \frac{\xi_R - 2\xi_P + \xi_L}{h^2} \quad (52)$$

where the diffusive terms are left in their CD form, since, by (49), they are unconditionally stable.

The sensitivity of the convective terms of this scheme is:

$$\frac{\partial(\text{convective terms})}{\partial \xi_P} = -\frac{u_r}{h} \quad (53)$$

which is stable as long as u_r and u_l have the same sign. This ensures the stability of the method, yet this comes at the expense of accuracy.

In fact, UD is only first order accurate in the mesh size which introduces a lot of *numerical diffusion* (also called *false diffusion*) [1] [3]. This can be seen by rearranging (52) as follows:

$$\begin{aligned}\frac{\xi_P(t + \Delta t) - \xi_P}{\Delta t} &= -u \frac{\xi_R - \xi_L}{2h} + \left(\frac{uh}{2} + D \right) \frac{\xi_R - 2\xi_P + \xi_L}{h^2} \\ &= -u \frac{\xi_R - \xi_L}{2h} + (D_{num} + D) \frac{\xi_R - 2\xi_P + \xi_L}{h^2}\end{aligned}\quad (54)$$

which clearly demonstrates that the UD scheme is plagued by this kind of numerical error (the enhanced stability is in fact a consequence of the stabilising nature of the extra diffusion).

Moreover, UD is also affected by streamline-grid skewness instabilities [2] [3], which are discussed in section 5.6.

The problems connected with UD differencing can be solved through the selection of a higher order upwind biased scheme. The method of choice for this project was the Quadratic Upstream Interpolation (QUICK).

Following [3], this interpolation method applied to the r face in fig. 16 takes the form:

$$\begin{aligned}\xi_r &= \frac{\xi_P + \xi_R}{2} - \frac{1}{8}(\xi_L + \xi_R - 2\xi_P) \quad \text{for } u_r > 0 \\ \xi_r &= \frac{\xi_P + \xi_R}{2} - \frac{1}{8}(\xi_{RR} + \xi_P - 2\xi_R) \quad \text{for } u_r < 0\end{aligned}\quad (55)$$

whose proof is provided in appendix A. The equations can be interpreted as a linear interpolation corrected by a quadratic upwind biased correction that accounts for the curvature of the concentration distribution.

This scheme is third order accurate, which removes almost totally the problem of false diffusion. Further, if the QUICK expressions are used to evaluate the convective term, under the assumption of constant velocity the sensitivity is:

$$\frac{\partial(\text{convective terms})}{\partial \xi_P} = -\frac{3}{8} \frac{u}{h} \quad (56)$$

which implies that, as long as $u_r \approx u_l$ the scheme is stable.

5.3.1 Stability of the QUICK Scheme

Despite its accuracy and transportiveness, the QUICK scheme introduces some instabilities which must be discussed in order to properly implement it in a FVM solution procedure. The first problem is raised in [4]. Still considering the one-dimensional convection diffusion problem, in steady state, the QUICK scheme gives rise to a discretised equation of the form:

$$a_P \xi_P = a_{LL} \xi_{LL} + a_L \xi_L + a_R \xi_R + a_{RR} \xi_{RR} \quad (57)$$

In normal flow conditions, in the absence of sinks or sources, an increase in ξ_P should lead to an increase to ξ in all neighbouring cells. This is indeed the

case if all the coefficients on the right of (56) are positive. However, QUICK presents both positive and negative coefficients which can result in overshoots and instabilities in the iterative solution of the discretised equation.

This problem can be fought by adopting the following formulation of QUICK [4]:

$$\begin{aligned}\xi_r &= \xi_P + \frac{1}{8}(3\xi_R - \xi_L - 2\xi_P) \quad \text{for } u_r > 0 \\ \xi_r &= \xi_R + \frac{1}{8}(3\xi_P - \xi_{RR} - 2\xi_R) \quad \text{for } u_r < 0\end{aligned}\quad (58)$$

which are a UD-like formulas in contrast to (55) that are CD-like.

These expressions envision the solution of the QUICK discretized equations using a *deferred correction* iterative solution method. Such an approach is based on retaining the UD part of QUICK, which always contributes with positive coefficients, as a set of implicit coefficients, whilst moving the remainder of the QUICK interpolation coefficients to the source term of the equation and evaluating them using the results of the previous iteration step. This procedure reduces the spectral radius of the amplification matrix, thus improving the stability of the solution method. Additional details will be laid out in section 5.5. The adoption of the deferred correction method based on (57) implies that it is possible to learn a lot about the time step requirement of the QUICK discretised convection-diffusion equation by considering the UD one. The focus is on the inviscid case since it corresponds to the most restrictive flow condition, and, since QUICK is a higher order method, it is should be expected to give rise to even more restrictive stability requirements [3].

The result of Von Neumann stability analysis applied to (52) with $D = 0$ is:

$$\left| \frac{C_m^{t+\Delta t}}{C_m^t} \right|^2 = \left(1 - 2c \sin^2 \left(\frac{\theta_m}{2} \right) \right)^2 + c^2 \sin^2(\theta_m) \quad (59)$$

where $c = (\Delta t u)/h$. It is clear that $0 < c < 1$ is a necessary condition, and, since there cannot be any extrema in the interval $0 < \theta_m < \pi$, this also turns out to be a sufficient condition.

Making the same approximations of the previous section, the time step for stability of the UD-Euler explicit scheme is 0.0022ns for electrons and 0.0067ns for holes ($\mu_h = 45 \text{ } \mu\text{m}^2\text{V}^{-1}\text{ns}^{-1}$).

As discussed in section 5.2, the time discretisation scheme that was adopted is implicit and second order. This would suggest that time steps longer than the ones calculated in this and the previous section should not hinder the stability of the numerical scheme, yet, in the author's opinion, it is a good practice to select time step values that do not exceed too much the ones that ensure the stability of the explicit schemes, even if the adopted scheme is of higher order, in order to attain accurate results and as a precaution against unexpected instabilities. For example, in the case of non-uniform flow, a time step might be suitable for a region of the solution domain, but not for another. Additionally, given that the QUICK discretisation tends to be more restrictive than the UD and CD are [3], a conservative choice of the time step is a guard against the

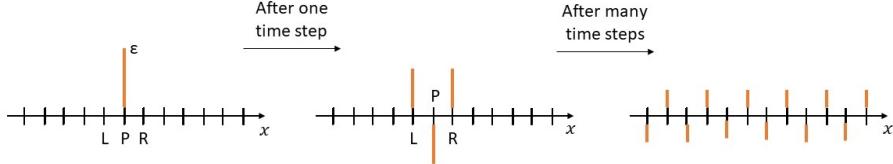


Figure 17: Emergence of wiggles. (a) Initial disturbance, (b) disturbance after one time step, (c) error distribution after many time steps

onset of unexpected instabilities.

A last feature of QUICK that needs to be discussed is the emergence of “wiggles”, that is concentration overshoots, near the perimeter of the concentration distribution, especially in the case of the inviscid flow problem. This problem was mentioned at the beginning of the section, but it is discussed at length here because choosing a small time step is beneficial to the reduction of this problem. The appearance of wiggles can be understood from the *discrete stability analysis* [2] of the diffusion-only equation. This type of analysis studies how a small localised error propagates from a time step to the next.

The starting point is the discretised diffusion equation:

$$\frac{\xi_P(t + \Delta t) - \xi_P}{\Delta t} = +D \frac{\xi_R - 2\xi_P + \xi_L}{h^2} \quad (60)$$

if a small perturbation is applied to point P , $\xi_P \rightarrow \xi_P + \epsilon$, and assuming that the unperturbed distribution exactly satisfies (59), the perturbation at the time step $t + \Delta t$ at the points P , R and L is:

$$\epsilon(t + \Delta t)_P = (1 - 2d)\epsilon \quad (61)$$

$$\epsilon(t + \Delta t)_{R,L} = d\epsilon \quad (62)$$

where $d = (D\Delta t)/h^2$. In order to ensure that the disturbance is damped out at R and L , there is the condition $0 < d < 1$, which results in the inequality $-1 < 1 - 2d < 1$.

While the last condition ensures that the magnitude of the error at P is reduced, in the case $1 - 2d$ is negative, the error at this cell will switch sign, producing the error distribution displayed at the centre of fig. 17. After many time steps, the error distribution will resemble the one on the right of fig. 17, that is wiggles will appear. Convection will only worsen the problem.

It is important to note that the emergence of wiggles is connected to the size of the time step, so, as already mentioned, taking a cautiously small time step can reduce this effect.

5.4 The Fully Discretised Equation

Now that the mathematical machinery needed to treat the convective term has been developed, it is finally possible to write down the fully discretised continuity

equation.

Looking at the electric potential in fig. 14 (c), it is apparent that in most of the detector material the largest electric field component is along the vertical direction. Therefore, in (44), the QUICK scheme was used to discretise p_s and p_n . Whereas, the simpler UD method was applied to p_w and p_e . Multiplying both sides of (44) by h , and dividing by μ_h , the result is the following equation:

$$\begin{aligned} & \frac{D_h}{\mu_h} (p_s + p_w - 4p_P + p_E + p_N)_{t+\Delta t} - \frac{2h^2}{\mu_h \Delta t} p_P(t + \Delta t) + \\ & \quad + (a_{SS} p_{SS} + a_{SP} p_S + a_{WP} p_W + a_{PP} p_P + a_{EP} p_E + a_{NP} p_N + a_{NN} p_{NN})_{t+\Delta t} = \\ & = - \frac{D_h}{\mu_h} (p_s + p_w - 4p_P + p_E + p_N)_t - \frac{2h^2}{\mu_h \Delta t} p_P(t) - \\ & \quad - (a_{SS} p_{SS} + a_{SP} p_S + a_{WP} p_W + a_{PP} p_P + a_{EP} p_E + a_{NP} p_N + a_{NN} p_{NN})_t \quad (63) \end{aligned}$$

where SS and NN have the obvious meaning. The coefficients of this equation must be evaluated at every CV in the solution domain. The diffusive and time discretisation terms will be the same for all cells, on the other hand, the convection coefficients will depend on the flow conditions at the cell faces.

Table 1 summarises the different possibilities expressed in terms of a_s , a_w , a_e and a_n .

	p_{SS}	p_S	p_W	p_E	p_N	p_{NN}	p_P
$E_s, E_n > 0$ $E_e, E_w > 0$	$-\frac{1}{8} a_s$	$a_s - \frac{1}{4} a_s - \frac{1}{8} a_n$	a_w	0	$\frac{3}{8} a_n$	0	$a_n + a_e + \frac{3}{8} a_s - \frac{1}{4} a_n$
$E_s, E_n > 0$ $E_e, E_w < 0$	$-\frac{1}{8} a_s$	$a_s - \frac{1}{4} a_s - \frac{1}{8} a_n$	0	a_e	$\frac{3}{8} a_n$	0	$a_n + a_w + \frac{3}{8} a_s - \frac{1}{4} a_n$
$E_s, E_n < 0$ $E_e, E_w > 0$	0	$\frac{3}{8} a_s$	a_w	0	$a_n - \frac{1}{4} a_n - \frac{1}{8} a_s$	$-\frac{1}{8} a_n$	$a_s + a_e + \frac{3}{8} a_n - \frac{1}{4} a_s$
$E_s, E_n < 0$ $E_e, E_w < 0$	0	$\frac{3}{8} a_s$	0	a_e	$a_n - \frac{1}{4} a_n - \frac{1}{8} a_s$	$-\frac{1}{8} a_n$	$a_s + a_w + \frac{3}{8} a_n - \frac{1}{4} a_s$

Table 1: convective coefficients of (63) for different flow conditions

In the context of the coefficients matrix, A , all terms corresponding to the same concentration value must be gathered. From this it is not difficult to see that the resulting matrix will be 7-diagonal.

Furthermore, note that the coefficients on the right hand side of the equation are just the opposite of the ones on the left except for the time discretisation term. This justifies the following expression for the source term:

$$\vec{b}(t) = A \vec{p}(t) - \frac{4h^2}{\mu_h \Delta t} \vec{p}(t) \quad (64)$$

which clearly reproduces the right hand side of (63).

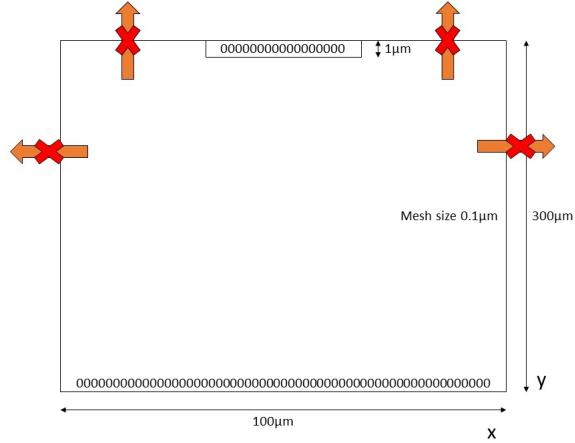


Figure 18: Solution domain for the FVM procedure

5.4.1 The Boundary Conditions

The solution domain that was considered for the solution of (62) is shown in fig. 18. It corresponds to the detector part of the domain displayed in fig. 10.

Two different types of boundary conditions were implemented:

- perfectly absorbing surface boundary condition: it consists of setting the concentration at a surface to 0, so that, when the charge concentration strikes those cells, it is effectively “sucked in”. This type of boundary condition was suitable for the backplane electrode surface and the p+ implant region since, when a charge concentration reaches either of them, it is for all practical purposes immediately collected;
- no-flux boundary condition: it entails setting the flux through a specific CV face in (41) to 0. This prevents the motion of the charge concentration through that face, thus making the surface in question impermeable. The application of this boundary condition requires the modification of the discretised equation at these special cells. In particular, all the coefficients describing the flux through the surface of interest must be set to 0. This type of boundary condition was applied to the silicon-air boundary and to the surface midway between two trips.

The no-flux condition at the material boundary makes obvious sense, yet, it requires some explanation when applied inside the detector material. The reason for this choice stems from three facts:

1. closing off the solution domain;
2. the absence of flux through those surfaces in the case of inviscid flow: from symmetry it is clear that there cannot be an electric field component in

the horizontal direction at the location of those surfaces, therefore, if there is no diffusion, the no-flux condition arises from the properties of the field;

3. the flux through those surfaces is generally very limited: it will be shown in section 5.6 that, unless the charge concentration is very close to those surfaces the diffusive flux through them is extremely limited. In fact, diffusive leakage to the neighbouring strip is in part prevented by the convective motion.

It is useful to provide an example of the modification of the discretised equation in the presence of a no-flux condition. Table 2 shows the convective and diffusive coefficients evaluated at the material boundary on the left hand side of the strip. Since the concentration flux though the northern face is forbidden, all the coefficients associated with the flux through that surface are set to 0.

(a)	p_{ss}	p_s	p_w	p_e	p_n	p_{nn}	p_p
$E_s, E_n > 0$ $E_e, E_w > 0$	$-\frac{1}{8}a_s$	$a_s - \frac{1}{4}a_s$	a_w	0	0	0	$a_e + \frac{3}{8}a_s$

(b)	p_s	p_w	p_e	p_n	p_p
	D	D	D	0	- $3D$

Table 2: coefficients of the top-left boundary. (a) Convective coefficients, (b) diffusive coefficients.

5.5 Implementation

This section discusses some of the practicalities concerning the calculation of the time evolution of the charge concentration.

First of all, the coefficients matrix was decomposed into a UD matrix, A_{UD} , which contained the UD coefficients of the QUICK scheme and the diffusive coefficients, and a correction matrix, A_{corr} , which contained the higher order part of the QUICK interpolation expression. The deferred correction iterative method then reads:

$$A_{UD} \vec{p}(t + \Delta t)^{n+1} = \vec{b}(t) - A_{corr} \vec{p}(t + \Delta t)^n \quad (65)$$

where the modification of the source term is evident. Starting from the equation written in this form, the SOR method was implemented to solve for the concentration at the advanced time.

Depending on the residual requirement (between 10^{-5} and 10^{-8}), the number of iterations required for convergence varied between 10 and 40. It was found that an over-relaxation parameter of the order of 1.2 guaranteed the fastest convergence.

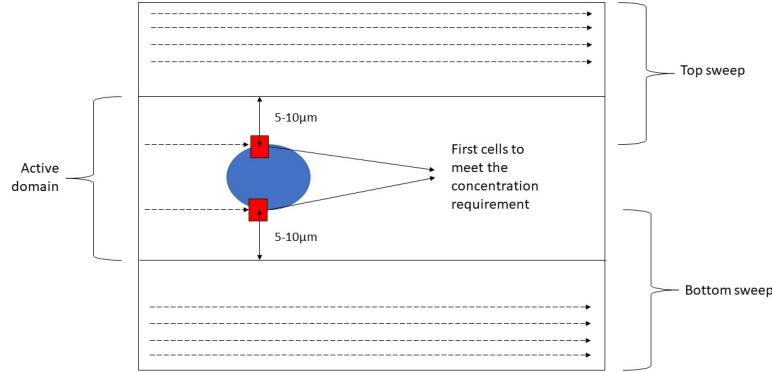


Figure 19: Active domain determination

The size of the structured mesh used in the discretisation of the domain was $0.1\mu\text{m}$. This resulted in a very large mesh of 3 million grid points, so that, even if few iterations were needed to time advance the charge concentration, the machine time taken to perform a single iteration was rather long. This issue was tackled through the introduction of an *active domain*. This idea consisted in solving for the new concentration only over a restricted region of space, discarding the rest of the domain. The approach was justified on the basis that convection and diffusion are local phenomena, that is that as long as the charge concentration is not close to the domain boundaries the effect of the boundary conditions is not felt by the charge concentration, and that the GS method on which SOR rests upon is a pointwise method, in short, the GS iteration formula applied to a grid point only accounts for the system in the immediate vicinity of the grid point. The active domain formulation of SOR significantly decreased the computation time.

The active region was identified by sweeping the solution domain from below and from above searching for the first cell that was meeting a certain minimum concentration requirement (10^{-5} of normalised concentration). Then, the active region was defined as a certain number of rows below the first cell meeting the concentration requirement in the bottom sweep, and a the same number above the first cell found in the top sweep. Fig. 19 displays a schematic view of this process. The number of rows used in the drift only problem corresponded to the distance of $5\mu\text{m}$, whilst the distance selected for the drift-diffusion problem amounted to $10\mu\text{m}$.

Despite the effectiveness of this procedure at reducing the computing time, this came at the expense of some accuracy. In fact, a tiny amount of charge (0.01% of the total) was lost in the process of moving the active domain along with the charge concentration location.

The last bit of practicalities concerns the presence of overshoots in the inviscid

flow problem whose emergence was discussed in section 5.4.1. In order to alleviate this pathologic behaviour of the numerical scheme, a simple “error correction” procedure was devised to remove the negative values of the concentration and, at the same time, ensure charge conservation (the negative overshoots are exactly balanced by the positive ones since the scheme is conservative). After the time-advanced charge concentration was found, firstly, the negative concentration values were set to 0. Next, the difference between the corrected concentration and the original concentration was calculated. Knowing the number of cells making up the charge distribution (it does not spread out in the non-diffusive case), a correction concentration per cell was computed. Finally, this correction was subtracted to all cells that exceeded the threshold of 0.1 times the maximum concentration.

This simple procedure was enough to eliminate the negative wiggles and reinstate charge conservation.

No error correction had to be performed in the drift-diffusion problem thanks to the stabilizing effect of diffusion.

Lastly, it must be mentioned that given the smaller mesh size required in the calculation of the charge motion, the electric potential calculated in section 4 had to be interpolated. A cubic spline method was used for this task.

5.6 Results and Discussion

The FVM procedure that was developed in the previous sections was applied to the motion of electron-hole pairs immersed in the potential displayed in fig. 14 (c). The time steps selected for the simulations were 0.002ns for the electrons and 0.006ns for the holes following the argument given in section 5.3.1. These time step values were adopted whenever convection was present. Despite this conservative choice, the drift-only simulations were still plagued by the presence of wiggles, reason why the error correction procedure illustrated in the previous section was devised.

Throughout the remainder of the report, it will be assumed that the charges are liberated by a soft X-ray of energy 26keV (see appendix F) impinging on the detector. The energy required to excite an electron-hole pair across the bad gap in silicon is approximately 3.736eV [14]. Then, the X-ray will produce around 7000 charge carries of the two types which are initially supposed to be distributed like a sphere of uniform density of radius 1 μm .

Using cylindrical coordinates, it is not difficult to show that the projected surface charge density on a plane passing through the centre of the sphere is:

$$\rho(\phi, r) = 2\rho_0 \sqrt{R^2 - r^2} \quad (66)$$

where ρ_0 is the uniform volume charge density and R is the radius of the charge globule. Dividing (65) by $2\rho_0$ gives the normalised charge concentration that was used in the numerical simulations of the charge motion.

For the purposes of this project, it is sufficient to account for the two-dimensional convection-diffusion of this projected surface charge density because:

1. there is no convective flux in the direction parallel to the strip length;
2. the diffusive flux in that direction will not modify the signal shape because, as already discussed, both the electric potential and the weighting potential do not vary in the strip direction.

After all, it is intuitively clear that the signal should not change if the globule of charge is translated in the direction parallel to the strip.

Fig. 20 (a) and (b) show the results that were obtained in the drift-only simulation when the liberated charges were placed at the location $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. In the case of electrons, charge collection occurred after 5.384ns, whereas for holes it occurred after 7.392ns. A theoretical calculation (see appendix B for details) involving an elementary point charge and based on an ideal np-junction potential, with the bias and applied voltages specified by fig. 14 (c), gives a collection time of 5.288ns for an electron, and of 6.553ns for a hole. The agreement is exceptionally good in the case of the electrons since the np-junction approximation works very well in the bottom half of the detector. On the other hand, the different form of the potential in the top half is responsible for a longer collection time. Either way, the good agreement with the theoretical expectation validates the results of the simulation.

Fig. 21 (a) and (b) show the results obtained for the same initial condition when diffusion was turned in.

The diagrams displaying the outcomes of the different simulations are provided in appendix C. The only pictures that deserve a short comment are fig. 37 and fig. 40. Here the charge globule was placed between two strips, that is on the boundary of the domain, and since the two-strip set-up is symmetric about that surface, only half of the domain had to be simulated. In this instance, the no-flux condition is exact also in the diffusive case as a result of symmetry considerations.

A last note regards the amount of diffusive leakage to the neighbouring strip that would occur if the simulation were performed on a two-strip domain. This kind of study was performed using a UD-based convective discretisation and the globule of holes was placed at $x = 10\mu\text{m}$, $y = 150\mu\text{m}$ (imagine there is a second strip on the left of the domain considered so far). The initial total normalised concentration amounted to 209.9, and after 6.816ns the amount of charge that had leaked to the left strip was 0.084, approximately the 0.04%. Therefore, unless the globule is placed very close to the boundary, the charge leakage is negligible. Then, the simulation with the globule of charge exactly between two strips already provides a good estimate of the appearance of the signal in these situations.

The simulation discussed in the previous paragraph was performed using UD in order to study how the behaviour of the simulation would differ from the QUICK discretised case. Initially the convergence properties of the algorithm did not deviate much from those of QUICK, however, when the globule was closer to the strip, so that there was a certain skewness between the structured grid and the field lines, the number of iterations needed for a time step increased from the usual one to around 100-120. This confirms the account given in [2].

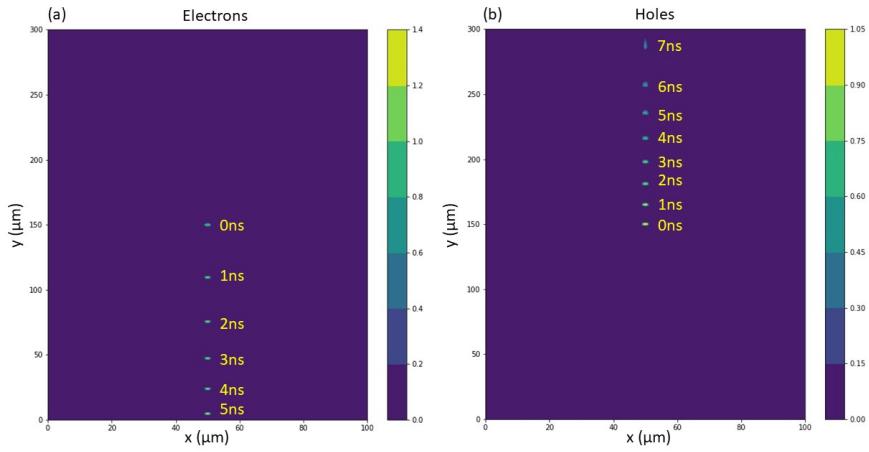


Figure 20: Results of drift-only simulation with initial condition $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. (a) Electrons, (b) holes.

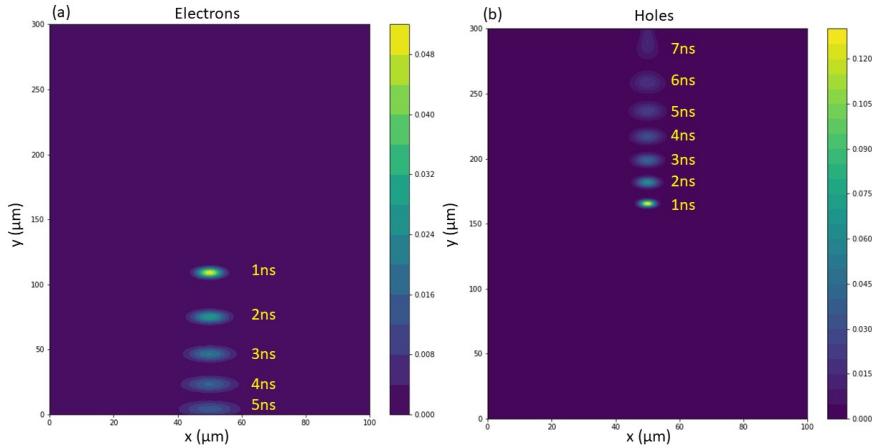


Figure 21: Results of drift-diffusion simulation with initial condition $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. (a) Electrons, (b) holes.

6 Signal Formation and Amplification

6.1 Current Signal Formation

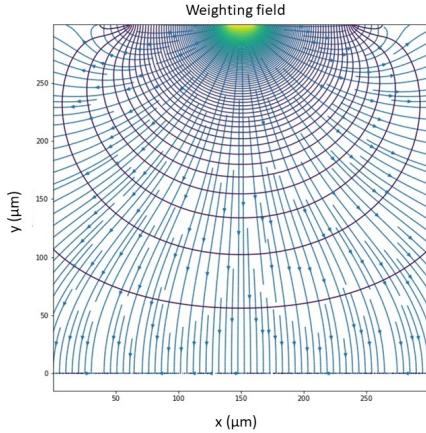


Figure 22: Weighting field

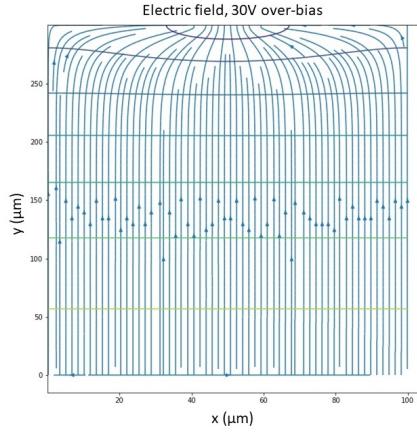


Figure 23: Electric field

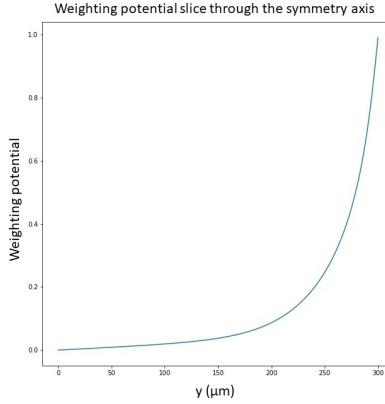


Figure 24: Slice through the symmetry axis of the weighting potential. Notice that it raises very sharply around the electrode.

Finally, all the components needed for the computation of the electric signal are available. Fig. 22 and 23 show the weighting field (3 strips shown) and the electric field (1 strip shown) inside the detector which can be combined according to (7) to produce the current signal generated by an individual CV (a parcel of charge). The contributions arising from all the CVs must be added up to produce the total current signal at a given time step. This approach corresponds to a

Riemann sum approximation of the two-dimensional volume integral, and it is only first order accurate. To the overall numerical error will also contribute the interpolation errors produced when the potentials were interpolated from a mesh of size $0.5\mu\text{m}$ to one of size $0.1\mu\text{m}$, and the differentiation errors resulting from the calculation of the fields. Afterwards, the time series of the current had to be integrated to obtain the time series of the induced charge.

6.2 Signal Amplification

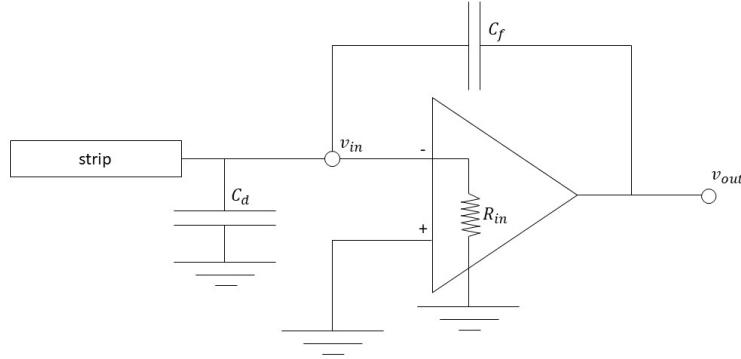


Figure 25: Circuit diagram of Charge Sensitive Amplifier

In semiconductor particle detectors the initial signal amplification is normally carried out by a charge sensitive amplifier (CSA) [14] [16]. This amplification module detects the charge induced on the strip and produces a sizeable voltage that is proportional to it. Fig. 25 shows a circuit diagram of the pre-amplification apparatus. The C_d capacitor connected to ground is needed to model the body capacitance of the strip to the backplane electrode.

Assuming that the input impedance of the op-amp is infinite, the signal current must all flow through the feedback capacitor, C_f . The input voltage, v_{in} , is then approximately equal to ground, and the voltage output, v_{out} , is then:

$$v_{out}(t) = -\frac{1}{C_f} \int_0^t i_S(t') dt' = -\frac{Q_S(t)}{C_f} \quad (67)$$

In reality R_{in} is very large, but still finite. Hence, the signal charge, $Q_S(t)$, will be divided amongst C_d and C_f . Calling A_{OP} the amplification constant of the op-amp, $v_{out} = -A_{OP}v_{in}$ so that the potential difference across C_f is $\Delta V = (A_{OP} + 1)v_{in}$. This difference is equal to Q_f/C_f , where Q_f is the charge on the feedback capacitor, which can be combined with the previous equation to yield $Q_f = (A_{OP} + 1)C_f v_{in} = C_{in} v_{in}$. The quantity C_{in} is often called the *dynamic input capacitance*, and its introduction enables to account for the effect of the amplifier through the size of this capacitance. From its defining equation, it is evident that if C_{in} is used, v_{out} can be treated as a ground. Therefore,

C_{in} and C_d are effectively in parallel implying that the fraction of charge that is transferred to C_f from the strip is $Q_f/Q_S = 1/(1 + C_d/C_{in})$. Consequently the output voltage becomes:

$$v_{out}(t) = -\frac{1}{1 + C_d/C_{in}} \frac{Q_S(t)}{C_f} \quad \text{where } C_{in} = (A_{OP} + 1)C_f \quad (68)$$

This expression is the voltage output if the amplifier were ideal, so, to make further progress, information about the frequency response must be included. In the small signal regime, an op-amp with a feedback capacitor behaves like a first-order low-pass filter. Since the frequencies of interest in a particle detector are much larger than the cut-off frequency, the amplifier responds like a pure resistance (section 2.9 of [14]). This reduces the diagram in fig. 25 to an RC circuit formed by the detector capacitance and the amplifier resistance. The impulse response of the system is then $h(t) = (e^{-t/\tau}/\tau)u(t)$ where $u(t)$ is a step function and $\tau = C_d/(\text{GBW}C_f)$, with GBW being the Gain Bandwidth Product of the amplifier.

The charge transfer from the strip to the feedback capacitor in a real amplification module, $Q_S(t)^*$, is then given by the convolution integral:

$$Q_S(t)^* = Q_S(t) * h(t) = \int_0^t Q_S(t') \frac{e^{-(t-t')/\tau}}{\tau} dt' \quad (69)$$

Some extra details about the CSA are provided in appendix E.

Lastly, the phenomenon of *cross-talk* between two adjacent strips was modelled. This effect consists of the formation of some signal charge on the strip next to the hit strip (the charge collecting strip) because of the presence of some inter-strip capacitance, C_{inter} . As charge is transferred from the strip to C_f , the residual charge left of the strip is:

$$Q(t)_{res} = Q_S(t) - \frac{1}{1 + C_d/C_{in}} Q_S(t)^* \quad (70)$$

therefore, the signal voltage of the strip is $v(t)_{res} = Q(t)_{res}/C_d$ and the cross-talk charge is $Q(t)_{cross} = (C_{inter}/C_d)Q(t)_{res}$. This charge will be processed by the amplifier on this strip to produce a voltage signal according to what was discussed previously.

6.3 Implementation and System Parameters

After the signal current time series had been obtained as explained in section 6.1, the signal charge time series was computed by performing an incremental integration using Simpson's rule. Given a function $f(x)$, the typical Simpson's formula, $(f_{i+1} + 4f_i + f_{i-1})/3$, involves two discretisation steps. Since the desired increment involved a single discretisation step (a single time step, Δt), the formula was rearranged as $(5f_{i+1} + 8f_i - f_{i-1})/12$, which corresponds to Simpson's rule applied to a single discretisation step.

Following, the charge transfer to the feedback capacitor was calculated from (69) again using an incremental integration, this time based on the trapezium rule. If (69) is evaluated from 0 to $t + \Delta t$, it is not difficult to show that it can be formulated as follows:

$$Q_S(t + \Delta t)^* = e^{-\Delta t/\tau} Q_S(t)^* + \frac{e^{-(t+\Delta t)/\tau}}{\tau} \int_t^{t+\Delta t} Q_S(t') e^{-t'/\tau} dt' \quad (71)$$

which is a very useful expression to perform the convolution integral in an incremental fashion.

The values of the capacitances relevant for the microstrip detector analysed in this project were found in the literature to be $C_d \sim 25\text{pFcm}^{-1}$ [21] and $C_{inter} \sim 1.5\text{pFcm}^{-1}$ [22]. Hence, assuming a strip length of about 5cm and taking the feedback capacitance to be 1/100 of the detector capacitance, reasonable values for the capacitances were taken to be $C_d = 100\text{pF}$, $C_{inter} = 7.5\text{pF}$ and $C_f = 1\text{pF}$.

Supposing a GBW product of 2GHz [14], τ then amounts to 50ns. Finally, A_{OP} was assumed to be 1000, so that $C_{in} = 1000\text{pF}$ which ensures a charge collection efficiency of 91%.

6.4 Results and Discussion

The following discussion will involve only the signals obtained in the case of the drift-only charge motion because the presence of diffusion has the only effect of smoothing out the signal while leaving its salient features unchanged.

All the results obtained in the project are available in appendix D.

Fig. 26 shows the current signal, the integrated charge and the amplifier voltage output produced on the hit strip by the electrons and the holes which started their motion from the initial location $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. As it can be seen, the current induced by the holes rises very sharply as the holes approach the strip since the weighting potential is almost totally concentrated around the strip (see fig. 24).

Fig. 27 displays the cross-talk voltage induced on the neighbouring strips, the signal current induced by the electrons and holes directly on these two strips (Ramo's theorem applied to the neighbouring strips, this signal will be the same on both of the adjacent strips from symmetry considerations) and the resulting integrated charge. The total integrates almost to 0, so that the steady state output voltage is very small, few microvolts. which is almost undetectable. On

the other hand, the inter-strip capacitance generates a rather sizeable voltage output on the neighbouring strips.

The signal current and charge results that have been presented so far are substantiated by fig. 5.14 of [16], which validates the signal calculation procedure that was implemented.

Fig. 28 contains the signal current that was obtained on the two hit strips when the charge globule was released at $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. The hole signal current presents two “bumps”. The first can be identified with the collision of the globule of holes with the material surface midway between the two strips. At that location, a very small horizontal drift velocity of the holes conspires with a very small horizontal component of the weighting field to produce an almost null current. As the parcel of holes separates in two, and drifts towards the strips the second bump is produced.

The integrated current has a similar shape to the one shown in fig. 25 (b), with the only difference that its magnitude is, as expected, roughly half of the previous case. The output voltage and the cross-talk signal are similar to the previous case as well.

Fig. 29 shows the signal current and charge induced on the neighbouring strips. Also in this case the signal current integrates to 0 thus resulting in a very small output voltage.

The signal results obtained for the initial condition $x = 25\mu\text{m}$, $y = 150\mu\text{m}$ are almost identical to the case of $x = 0\mu\text{m}$, with the sole difference that the hole collection time is slightly longer.

A last point to be addressed regards the combination of the cross-talk and the charge-induced signals. In the case of the hit strip, the voltage output resulting from charge induction is the largest component of the signal and it surpasses the cross-talk pulse caused by the neighbouring strips by about two orders of magnitude. By contrast, the cross-talk voltage generated by the hit strip on the neighbouring strips is often much larger than the charge-induced signal on these strips since the integrated charge usually amounts to 0. Therefore, if the relevant cross-talk and induced signals were to be added up to produce the total voltage output, one component would frequently dominate over the other producing graphs that resemble very closely the line-shape of the larger signal component.

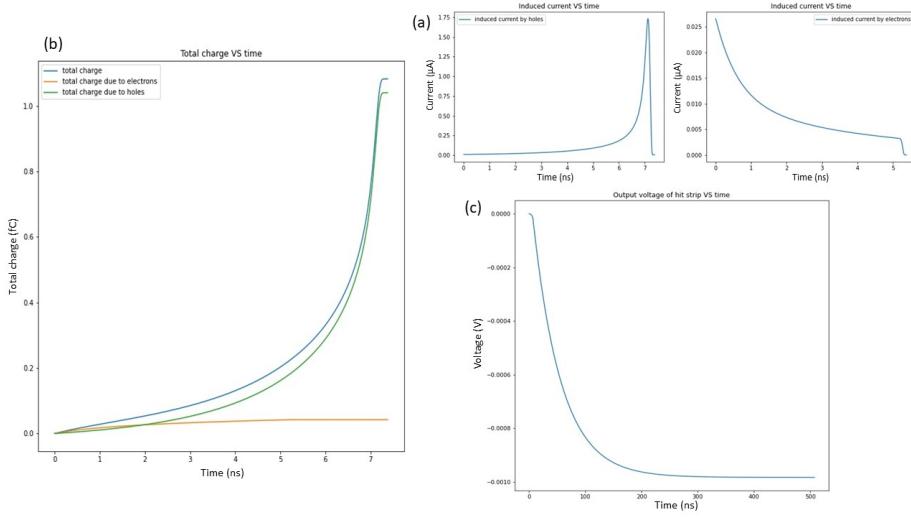


Figure 26: Signal on hit strip due to charge carriers generated at location $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. (a) Induced current, (b) integrated charge, (c) output voltage of CSA.

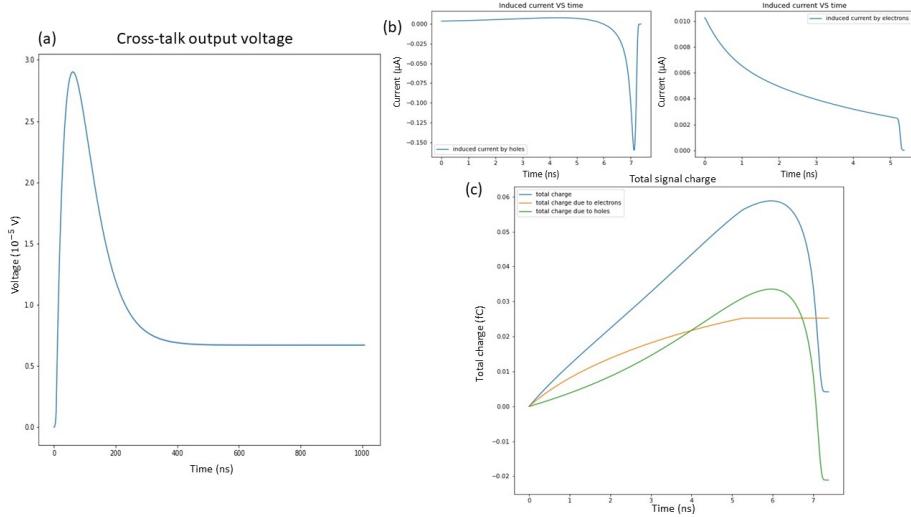


Figure 27: (a) Cross-talk voltage, (b) signal current on neighbouring strip, (c) integrated charge on neighbouring strip.

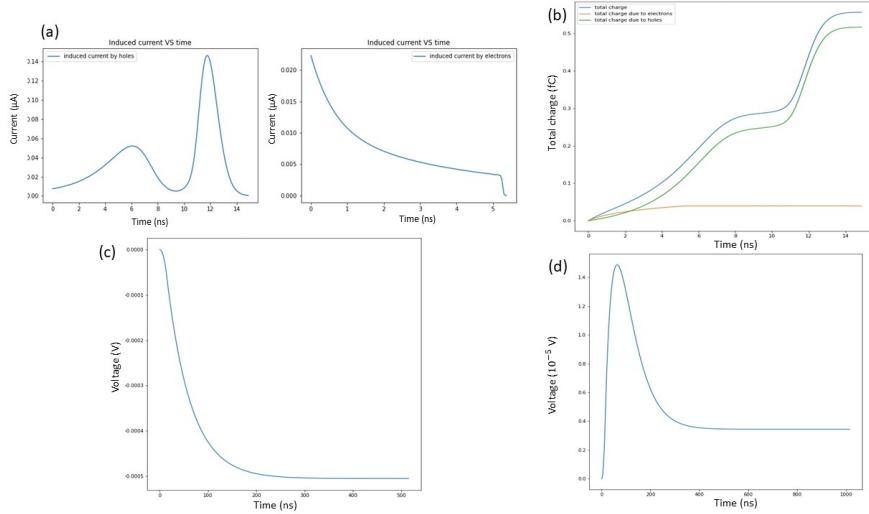


Figure 28: Signal given initial condition $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. (a) Induced current, (b) integrated charge, (c) output voltage of CSA, (d) cross-talk voltage.

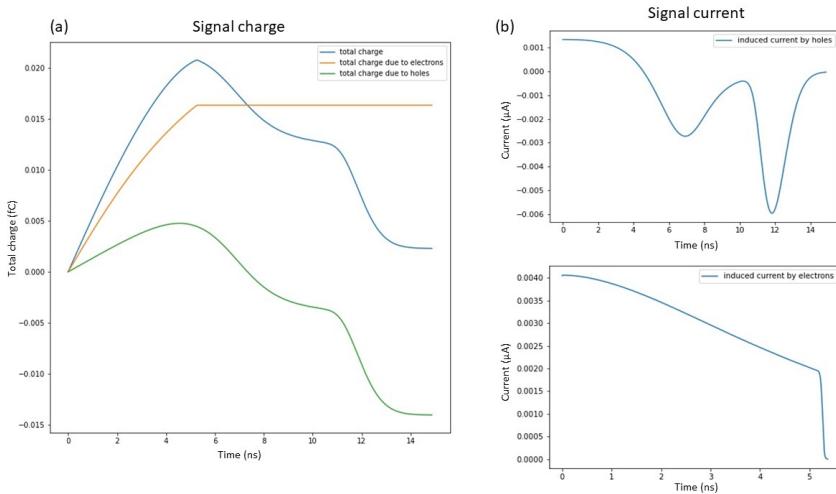


Figure 29: Signal on the neighbouring strip. (a) Integrated charge, (b) induced current by holes and electrons respectively.

7 Radiation Damage

As explained in the introduction, silicon detectors must withstand very large fluxes of particles. The primary energy loss mechanism for charged particles in a semiconductor material is ionisation, that is electron-hole pairs formation. This is a reversible process and hence it is not a cause of damage. However, incident particles can also undergo non-ionising energy loss especially via elastic and inelastic scattering off silicon nuclei. When the energy transfer to the nucleus exceeds approximately 25eV, the removal of the silicon atom from its lattice site is very likely (see chapter 8 in [16]).

Obviously, the energy of the particles taking part to collisions in accelerators is always much larger than 25eV, meaning that the knocked-off silicon atoms often have more than enough energy to cause the displacement of additional nuclei from their lattice sites. Therefore, non-ionising energy loss produces damage clusters in the silicon substrate.

The result of these collision can be the formation of a vacancy in the silicon crystal lattice, the creation of an interstitial (that is a stable complex between two existing lattice sites, this is typically favoured by the presence of impurities in the silicon) or the transmutation of the silicon to other atoms via electron, proton or neutron capture.

The effect of the formation of these lattice defects is the modification of the energy band structure of the semiconductor material. Impurity states are added between the valence band (VB) and conduction band (CB) which modify the material properties of the detector.

There are two main types of impurity states:

1. donor/acceptor centres: these levels are close to the VB or the CB and behave effectively like donors and acceptors atoms. This introduces a new inhomogeneous concentration of dopants which has the effect of altering the bias voltage required to deplete the detector. In the case of p-on-n detectors, acceptor-like centres are created in the n-bulk which in large concentrations ultimately lead to *type inversion*. These centres trap negative charge so that larger voltages are needed to fully deplete the bulk of the detector. Eventually the bias voltage exceeds the maximum voltage of the power supplies, making full depletion impossible. At that point, the detector can only be operated at partial depletion;
2. trapping centres: these are levels midway between the VB and the CB which favour the recombination of electrons in the CB with holes in the VB. This reduces the amount of minority charge carriers with time and thus degrades the quality of the induced signal.

The first point indicates that a new electric potential map is needed in order to simulate the charge motion in a damaged detector. Still using the junction approximation, this map can be obtained by repeating the calculation performed in section 4 using a larger effective charge density [16] distributed over a smaller depletion region. The numerical simulation was repeated assuming an

effective donor density of $3.5 \cdot 10^{12} \text{ cm}^{-3}$ uniformly distributed over a depletion region of thickness $200 \mu\text{m}$ from the top surface of the detector. The effective donor concentration was identified from the expression of the bias voltage of an np-junction supposing the depletion region thickness just mentioned and a bias-point voltage of 115 V . The electric potential map resulting from these specifications is shown in fig. 30 (a).

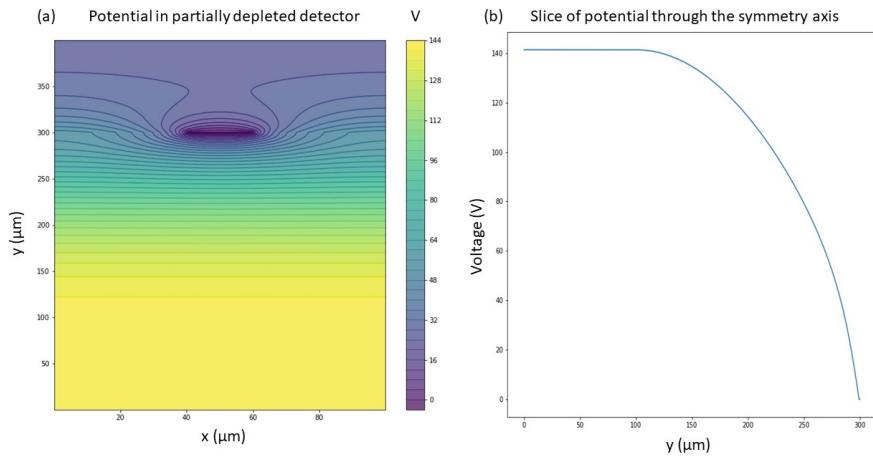


Figure 30: Potential in a partially depleted detector. (a) Potential map, (b) slice through the symmetry axis, only the interior of the detector is shown.

The second point can be accounted for by introducing a mean carrier lifetime, τ , such that $1/\tau$ describes the probability per unit time of a charged particle being trapped. The continuity equations (38) and (39) must then be modified respectively through the addition of a term $-p/\tau$ and $-n/\tau$ on their right side to describe the charge loss due to trapping. This translates into the extra term $-p_P/\tau$ in the discretised equation (44). The time constant that was assumed for the simulations was 3 ns .

Referring to fig. 30 (b), it is clear that in the region $0 \mu\text{m} < y < 100 \mu\text{m}$ there is no electric field, so the only transport process that can move charges around in that part of the detector is diffusion. Since diffusion, by its nature, has no mean velocity associated to it, it is not possible to use (7) in the computation of the electrical signal. Instead, (8) must be applied to directly calculate the total induced charge as time passes. This is accomplished by performing a simple Riemann sum across the charge distribution at every time step. Note that the distribution at $t = 0$ defines the baseline for the signal charge.

7.1 Results and Discussion

In the brief study that was conducted in the case of signal formation in a damaged detector, two types of set-up were considered:

1. motion in an undepleted detector: in this case no potential is applied to the detector, no electric field is present in the bulk of the material and the depletion region is negligibly small. The charge motion then happens solely due to diffusion under the presence of trapping. The usual initial charge distribution was considered starting from the location $x = 50\mu\text{m}$, $y = 150\mu\text{m}$.
2. motion in a partially depleted detector: the motion under the effect of the electric potential shown in fig. 30 is considered here, still under the presence of trapping. The time evolution of the same initial charge distribution was considered starting from locations $x = 50\mu\text{m}$, $y = 125\mu\text{m}$ (in the depleted region) and $x = 50\mu\text{m}$, $y = 95\mu\text{m}$ (in the undepleted region).

7.1.1 Undepleted Detector

The time steps that were selected for this simulation were 0.02ns for the electrons and 0.06ns for the holes. The longer time steps are allowed by the absence of convective motion.

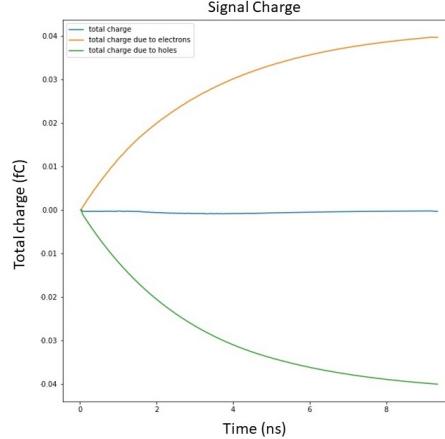


Figure 31: Induced charge in undepleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$.

The numerical simulation showed, for both electrons and holes, a Gaussian spread of the initial concentration accompanied by an exponential decay. The solution domain that was considered was the same of section 5 since the diffusion length after 5ns is around $4\mu\text{m}$ for the electrons and $1\mu\text{m}$ for the holes, which ensures that, for the time interval that was simulated, the influence on the

solution of the no-flux conditions on the left and right boundaries was very limited. The results obtained for the induced charge are shown in fig. 31. Roughly no charge is induced at the detector since the electrons and the holes spread almost identically.

The signal charge induced by the holes at the amplifier is negative, in contrast to the results displayed in section 6, because of the charge loss caused by trapping. Initially the holes induce a certain amount of negative charge on the strip, yet, as time passes, holes are trapped resulting in a reduction of the negative charge at the strip. The excess charge can leave the strip only through the power supply (i.e. the amplifier), so that, according to the convention discussed in section 2.1, the signal charge must be negative (negative charges are entering the amplifier accumulating on the feedback capacitor). A similar reasoning justifies the sign of the charge induced by the electrons.

7.1.2 Partially Depleted Detector

Firstly, let's consider the signal charge induced by the charge motion with initial condition $x = 50\mu\text{m}$, $y = 95\mu\text{m}$. The plot is very similar to the one obtained in the undepleted case. The reasons are that the electrons can move only in the undepleted region, the holes spend a significant amount of time in the this region as well, and, even if a part of the holes manages to enter the depletion region, they initially encounter only a very small electric field. Therefore, diffusion is still the dominant process. Nevertheless, if the amount of trapping were smaller, the signal charge induced in this situation would be much greater since the holes would have more time to traverse the depletion region.

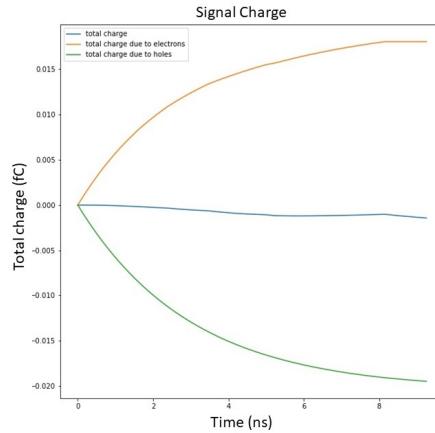


Figure 32: Induced charge in partially depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 95\mu\text{m}$.

The simulation results for the motion with initial condition $x = 50\mu\text{m}$, $y = 125\mu\text{m}$ are displayed in fig. 33. The motion of the holes is very similar to the

one shown in fig. 21 (b), with the only difference that trapping reduces the charge concentration with time.

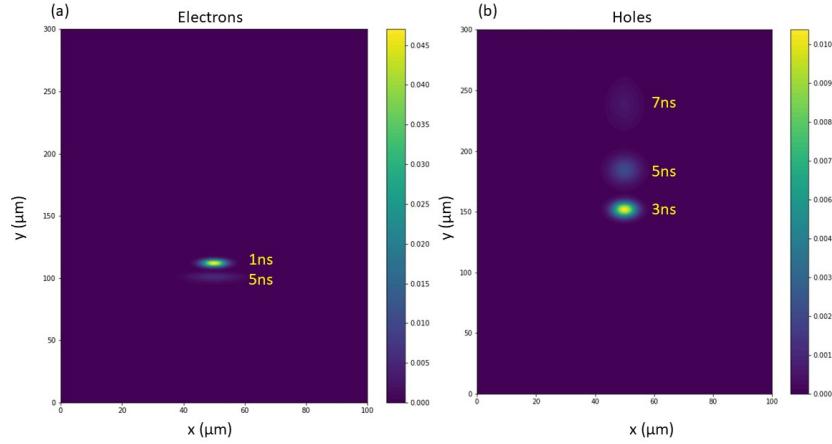


Figure 33: (a) Electron motion, (b) hole motion in partially depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 125\mu\text{m}$.

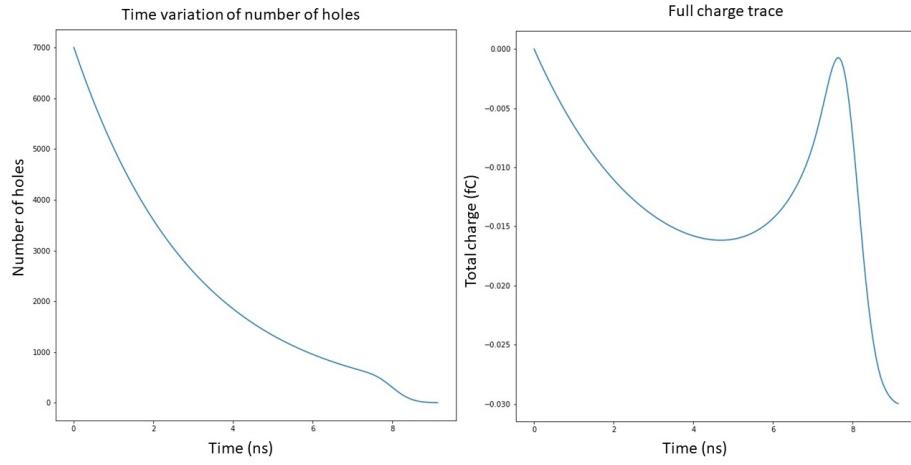


Figure 34: (a) Variation of the number of holes with time, (b) full induced charge trace.

In the charge motion situations that have been contemplated so far in this section, charge collection did not occur because the charge carriers were trapped

before they could reach the detector electrodes. This permits a carefree application of (7) in the calculation of the signal charge. However, in this last case, the collection of the holes did happen, which demands a more careful application of (7) for the purpose of signal calculation. The main issue here is concerned with identifying the time at which charge collection begins. Stopping the charge trace at that moment is necessary to ensure the preservation of the signal charge. In fact, if the signal is considered beyond that point, it will eventually drop to the baseline since the number of holes is falling to 0.

This task was accomplished through the analysis of the time variation of the number of holes (fig. 34 (a)) and of the signal charge time series ignoring charge collection (fig. 34 (b)). From the time derivative of the graph in fig. 34 (a), it was established that charge collection just begun at time 6.96ns, and it reached its maximum at 7.98ns. On the other hand, the location of the second maximum of the induced signal charge was found to occur at 7.632ns, which falls approximately between those two limits. This justified the choice of stopping the signal charge trace at that location.

Finally, fig. 35 shows the overall signal charge produced by the charge motion resulting from this initial condition. A sizeable amount of charge is induced, which generates an output voltage of about 0.025mV, the largest obtained in this part of the project.

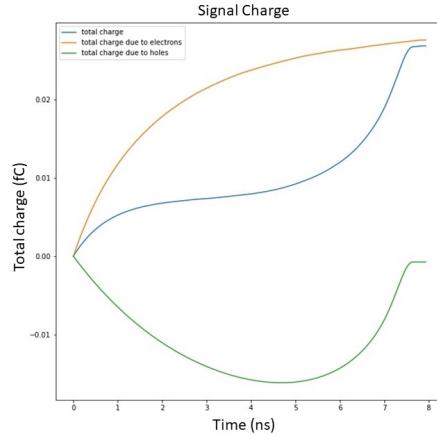


Figure 35: Induced charge in partially depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 125\mu\text{m}$.

8 Conclusion

The first part of the project (sections 3 and 4) revolved around the calculation of the two relevant scalar fields, namely the weighting potential and the electric potential. The methods that were individuated to improve these numerical simulations, either from a hardware utilisation point of view (multiprocessing) or an algorithmic point of view (GMG method), were explained and discussed. The largest part of the project was concerned with the transport dynamics of the charge carrier concentration within the semiconductor detector. The mathematical framework of the FVM procedure that was implemented was analysed at length in section 5. The only type of initial charge distribution that was contemplated in the project was the spherical one of section 5.6. Nonetheless, the code that was developed is capable of handling any type of initial distribution. It was preferred to consistently use the same one in order to simplify the comparison of different results.

The ultimate goal of the project was realised in section 6 where the charge motion results, the electric potential and the weighting potential were combined according to Ramo's theorem to generate the expected electrical signals.

The final section of the report gave a glimpse to the effects of radiation damage on the process of signal production. From this brief overview, it became particularly clear that a large applied potential is a necessary condition to operate a detector in a damaged state because, on one hand, this gives rise to a large depletion region, while on the other it ensures a rapid collection time of the liberated charges, which prevents excessive trapping. Furthermore, a quick amplifier response is useful in this situation since it would permit the detection of signal pulses when charge collection is not possible.

9 Appendix

9.1 A: Proof of the QUICK Interpolation Equations

Referring to fig. 16, it is possible to write down the following Taylor expansions of the quantities ξ_r , ξ_R and ξ_L in terms of ξ_P :

$$\begin{aligned}\xi_r &= \xi_P + \frac{\partial \xi}{\partial x} \Big|_P \frac{h}{2} + \frac{\partial^2 \xi}{\partial x^2} \Big|_P \frac{h^2}{8} + O(h^3) \\ \xi_R &= \xi_P + \frac{\partial \xi}{\partial x} \Big|_P h + \frac{\partial^2 \xi}{\partial x^2} \Big|_P \frac{h^2}{2} + O(h^3) \\ \xi_L &= \xi_P - \frac{\partial \xi}{\partial x} \Big|_P h + \frac{\partial^2 \xi}{\partial x^2} \Big|_P \frac{h^2}{2} + O(h^3)\end{aligned}$$

Substituting the first and second derivatives of ξ in the first equation using the second and third equations, an expression for ξ_r in terms of ξ_L , ξ_P and ξ_R can be obtained.

9.2 B: Charge Collection in a NP-Junction Detector

Assuming a very small donor density in comparison to the acceptor density the bias voltage of an np-junction can be written as:

$$V_b \approx \frac{w^2 e N_D}{2\epsilon_r \epsilon_0}$$

by inverting (23). Further, $w \approx x_n$ if the size of the p-doped region is much smaller than the size of the n-doped region.

In the case of over-bias, a uniform electric field, $\vec{E} = V_{applied} - V_b/w\hat{x}$, is superimposed to the one generated by (21). Referring to fig. 9 (b), the total electric field can then be expressed as:

$$\vec{E} = \left(\frac{V_{applied} + V_b}{w} + \frac{2V_b}{w^2}x \right) \hat{x} = (a + bx)\hat{x}$$

where the expression for V_b was used to simplify the equation.

The equation of motion for an electron and a hole in a semiconductor can then be written as:

$$v_e = -\mu_e E = \dot{x}_e$$

$$v_h = \mu_h E = \dot{x}_h$$

These equations can be solved using separation of variables. Calling x_0 the charge position at time $t = 0$, the solutions for the time as a function of initial and final locations are:

$$t_e = \frac{1}{\mu_e a} \ln \left(\frac{ax_0 + b}{ax + b} \right)$$

$$t_h = \frac{1}{\mu_h a} \ln \left(\frac{ax + b}{ax_0 + b} \right)$$

Noting that the coordinates on the n-side of the junction are negative according to fig. 9 (b), the initial location is $-w/2$, while the final positions are $x = 0$ for the hole and $x = -w$ for the electron.

The expressions for the collection times are then:

$$t_e = \frac{1}{\mu_e a} \ln \left(\frac{b - aw/2}{b - aw} \right)$$

$$t_h = \frac{1}{\mu_h a} \ln \left(\frac{b}{b - aw/2} \right)$$

Taking the values $V_{applied} = 114.2951\text{V}$, $V_b = 84.2951\text{V}$, $\mu_e = 135\mu\text{m}^2\text{V}^{-1}\text{ns}^{-1}$ and $\mu_h = 45\mu\text{m}^2\text{V}^{-1}\text{ns}^{-1}$ and $w = 300\mu\text{m}$. The collection times given in section 5.6 can be calculated.

9.3 C: Charge Motion Results in a Normal Detector

The next figures display the results of the FVM simulations discussed in section 5 for different initial conditions.

9.3.1 Drift-only

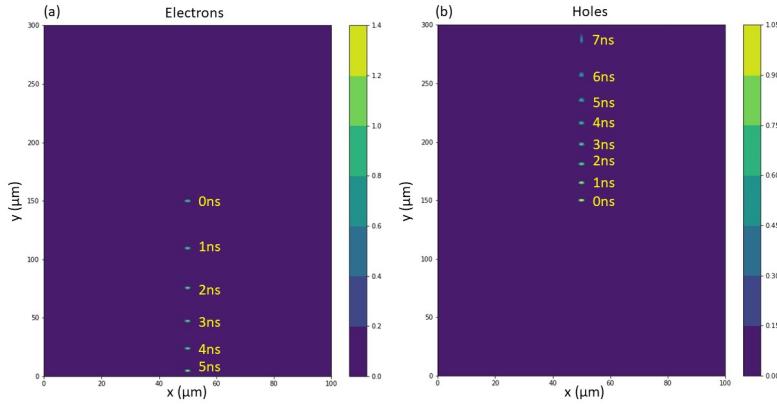


Figure 36: (a) Electron motion, (b) hole motion in fully depleted detector.
Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$.

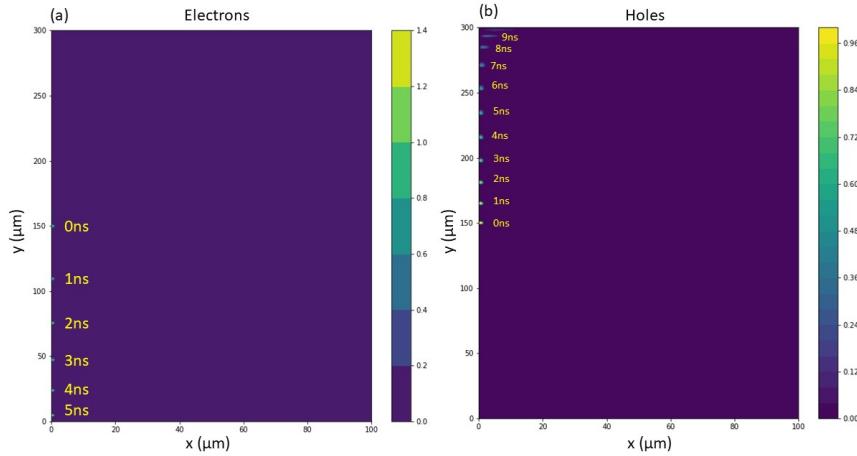


Figure 37: (a) Electron motion, (b) hole motion in fully depleted detector.
Initial condition: $x = 0\mu\text{m}$, $y = 150\mu\text{m}$.

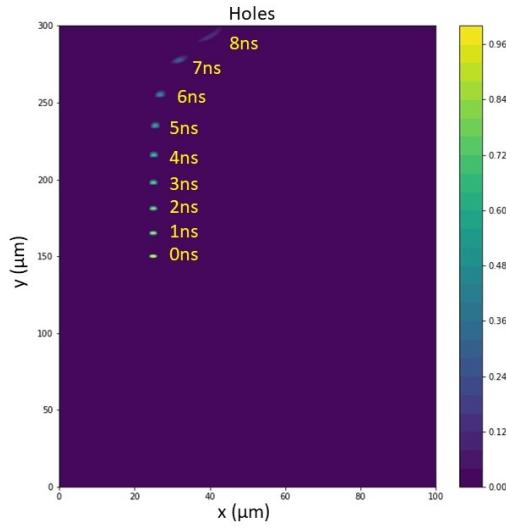


Figure 38: Hole motion in fully depleted detector. Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$.

9.3.2 Drift-diffusion

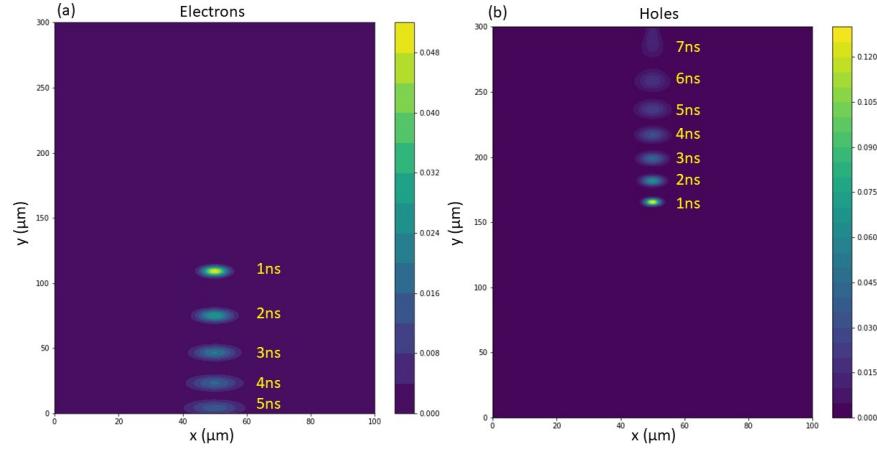


Figure 39: (a) Electron motion, (b) hole motion in fully depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$.

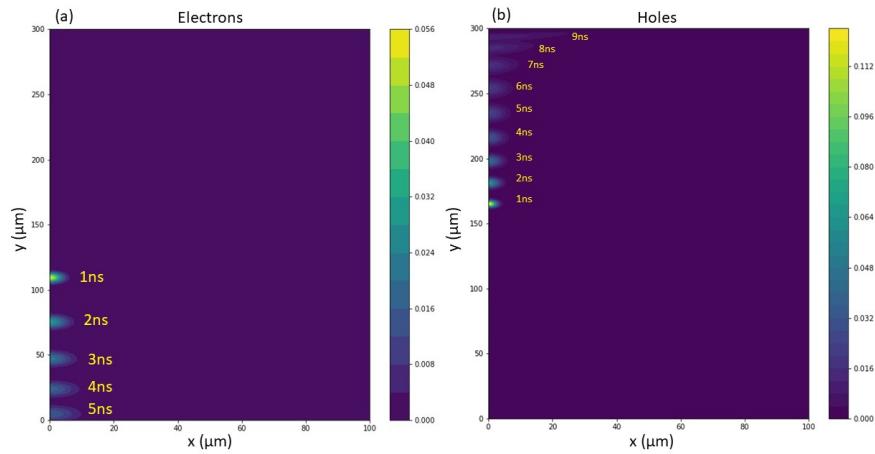


Figure 40: (a) Electron motion, (b) hole motion in fully depleted detector.
Initial condition: $x = 0 \mu\text{m}$, $y = 150 \mu\text{m}$.

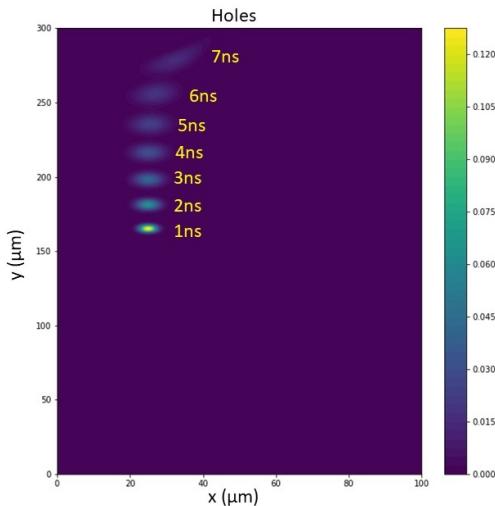


Figure 41: Hole motion in fully depleted detector. Initial condition: $x = 25 \mu\text{m}$, $y = 150 \mu\text{m}$.

9.4 D: Signal in a Normal Detector

9.4.1 Drift-only

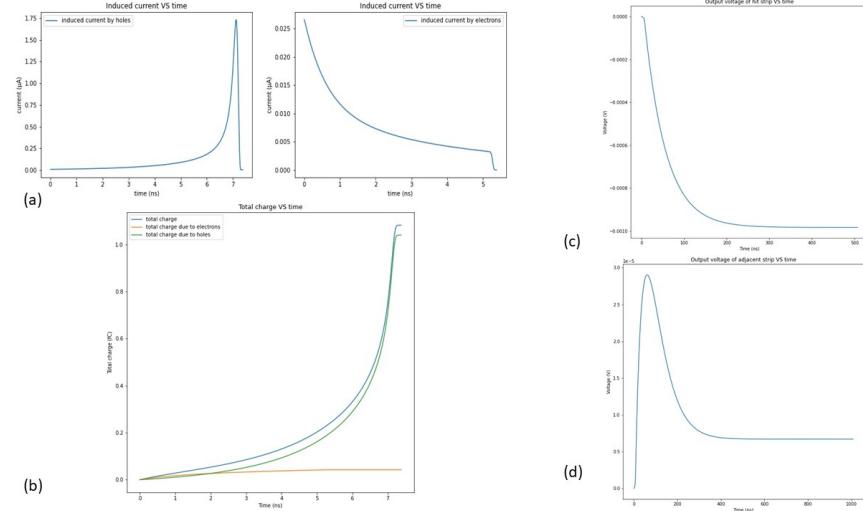


Figure 42: Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

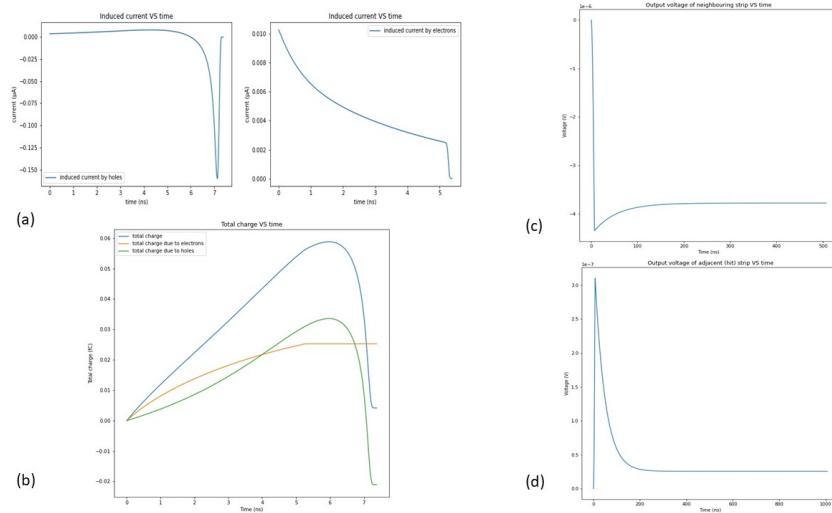


Figure 43: Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. Signal on neighbouring strips: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

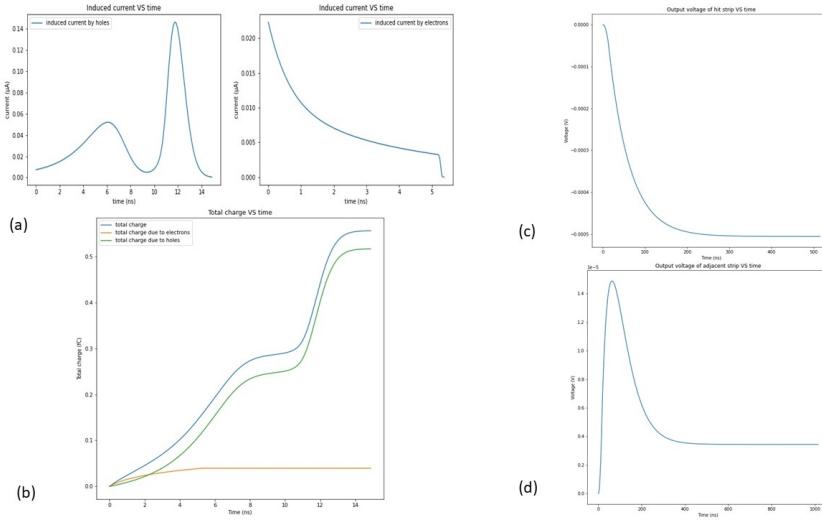


Figure 44: Initial condition: $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

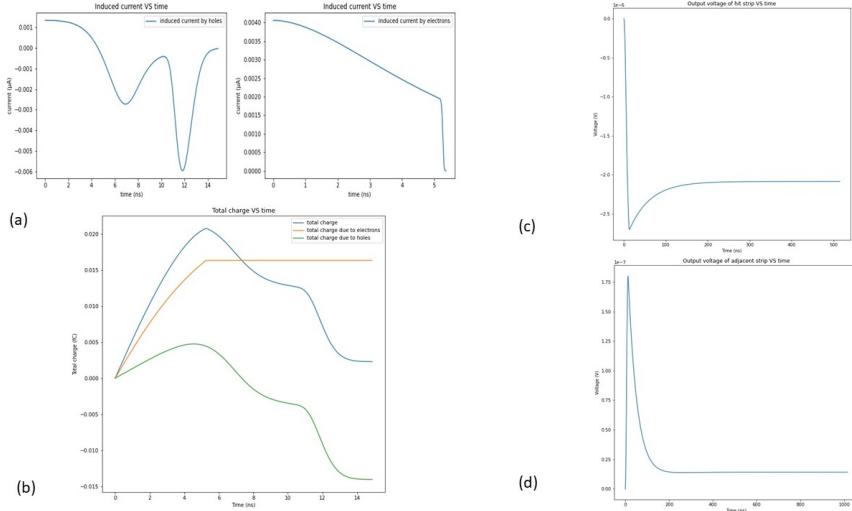


Figure 45: Initial condition: $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. Signal on neighbouring strips: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

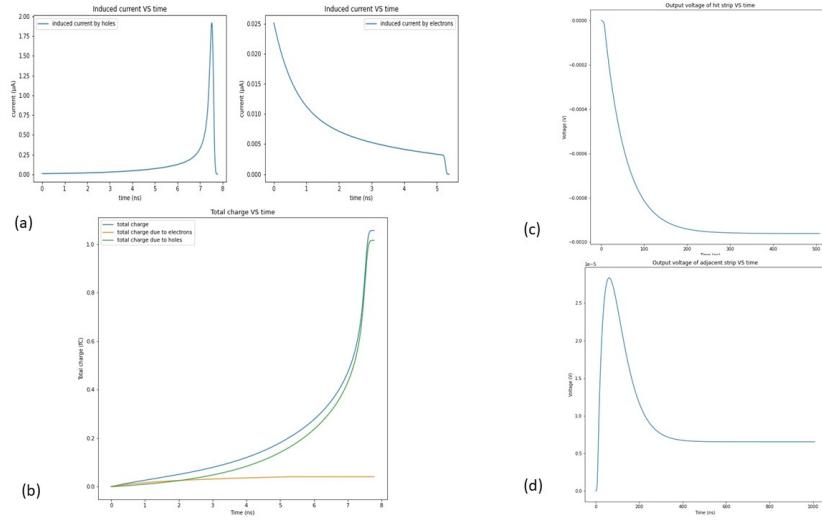


Figure 46: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

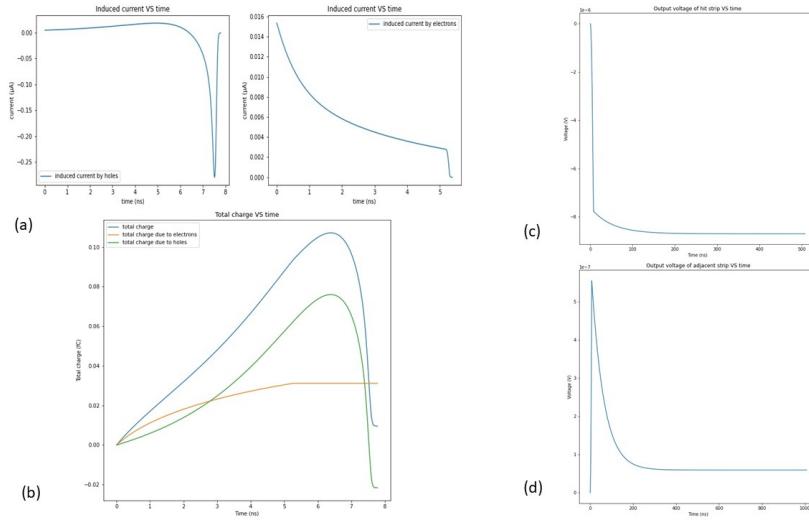


Figure 47: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on left neighbouring strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

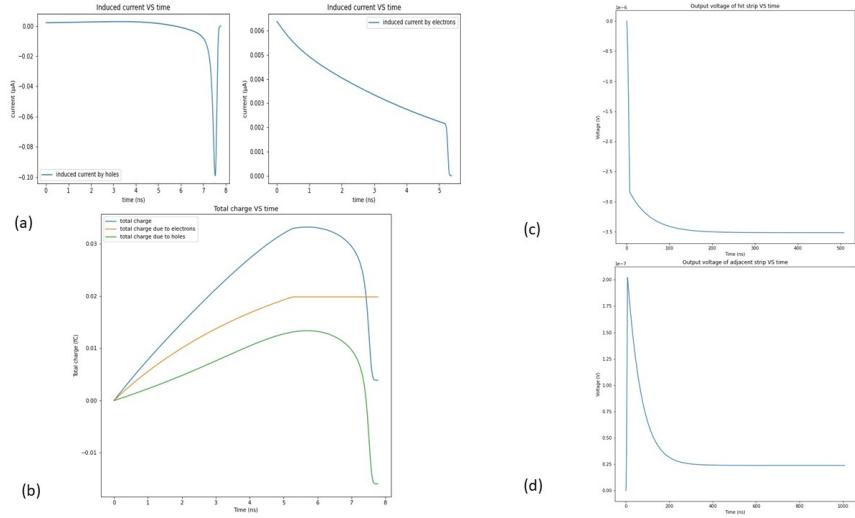


Figure 48: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on right neighbouring strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

9.4.2 Drift-diffusion

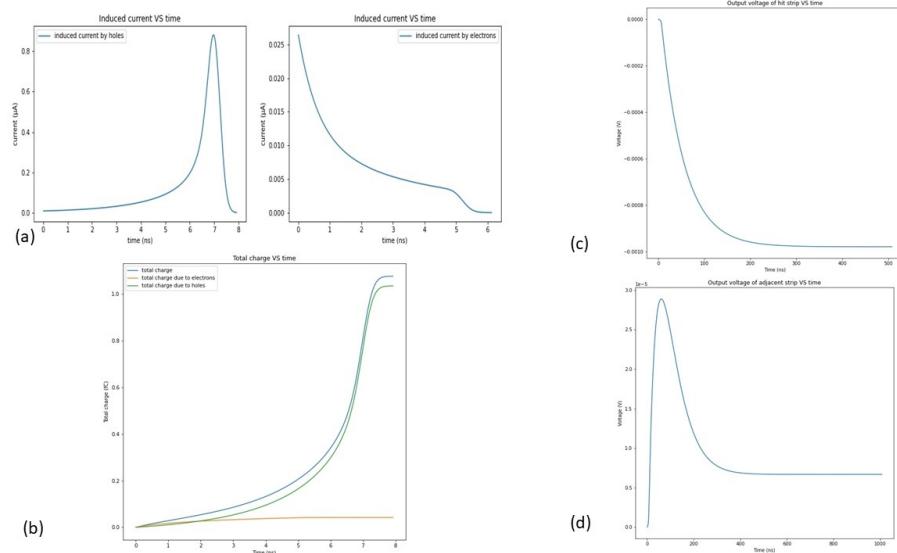


Figure 49: Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

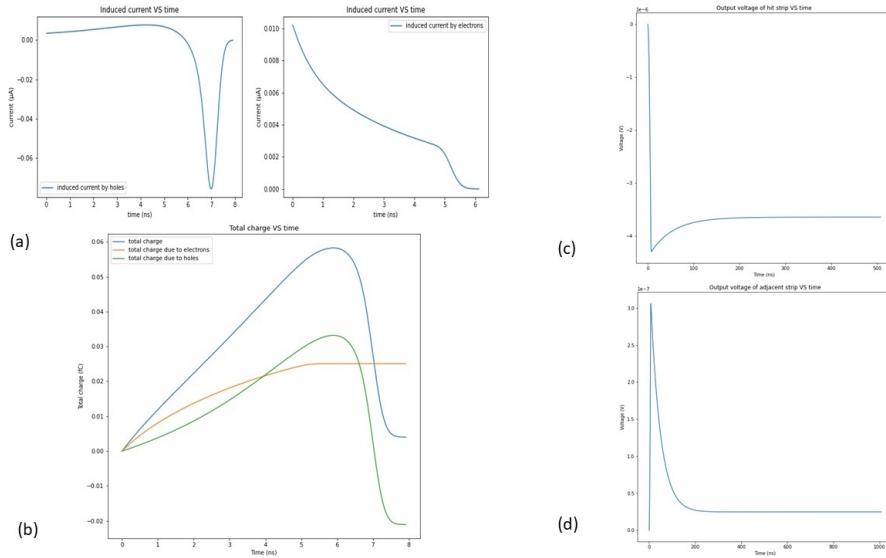


Figure 50: Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. Signal on neighbouring strips:
 (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

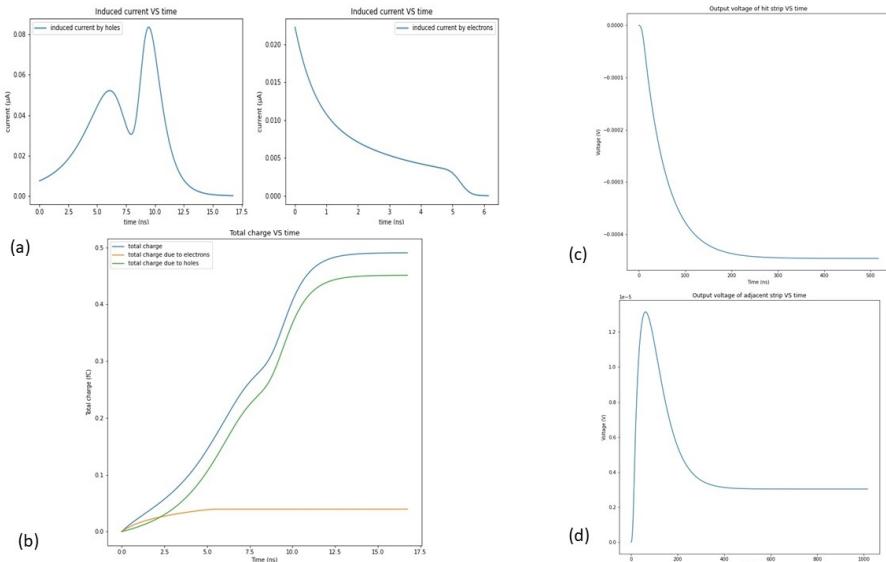


Figure 51: Initial condition: $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

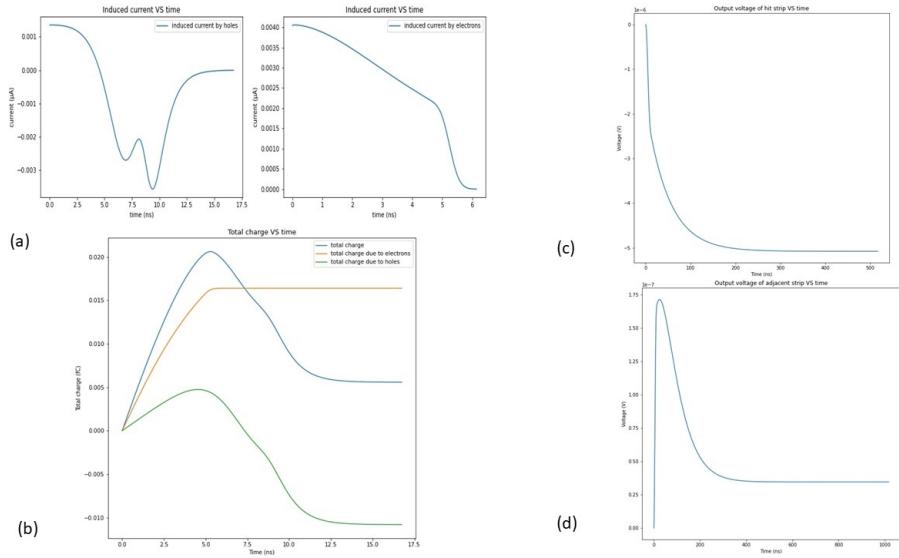


Figure 52: Initial condition: $x = 0\mu\text{m}$, $y = 150\mu\text{m}$. Signal on neighbouring strips: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

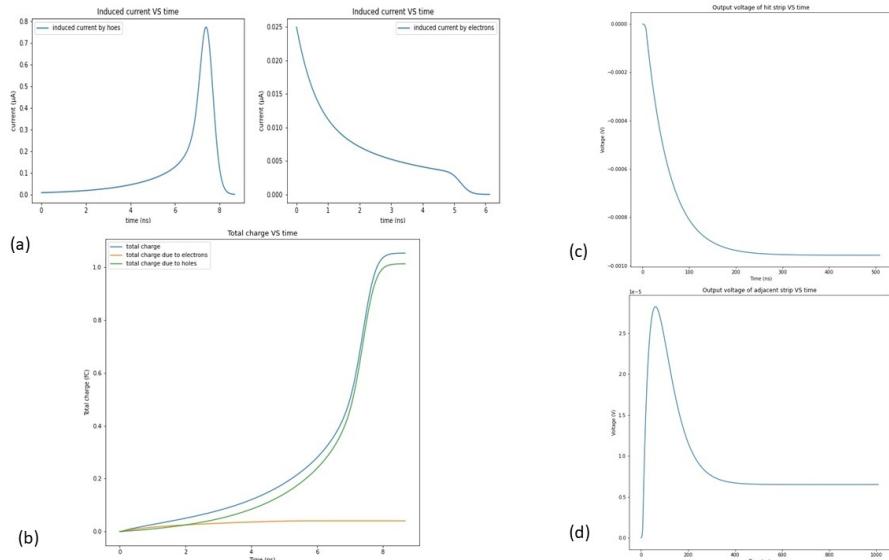


Figure 53: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on hit strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

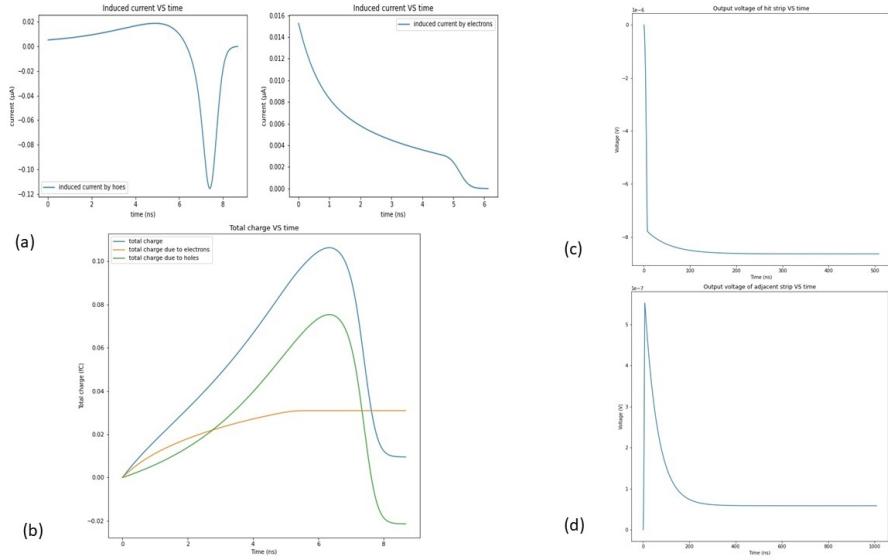


Figure 54: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on left neighbouring strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

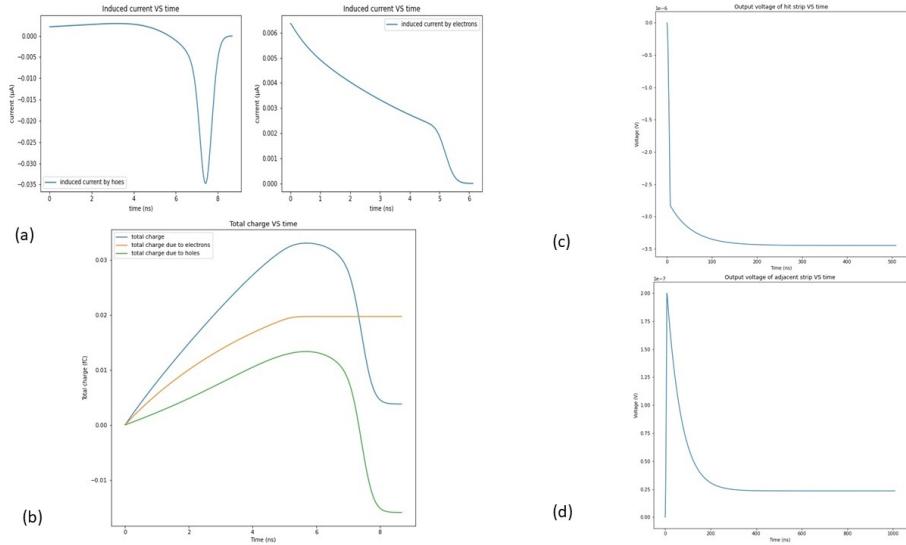


Figure 55: Initial condition: $x = 25\mu\text{m}$, $y = 150\mu\text{m}$. Signal on right neighbouring strip: (a) induced current, (b) induced charge, (c) output voltage, (d) cross-talk voltage.

9.5 E: Signal Response of a CSA

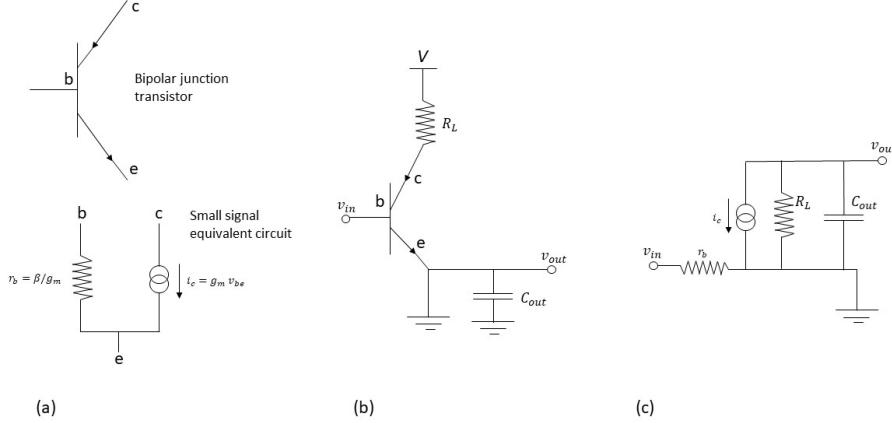


Figure 56: (a) Circuit diagram of bipolar junction transistor, (b) circuit diagram of common emitter amplifier, (c) small signal approximation of (b).

The left side of fig. 56 shows an ideal bipolar junction transistor. In the small signal model, this transistor presents only a small base resistance. The centre of the figure shows the same transistor used in a common emitter amplifier arrangement with the addition of a load collector resistance and a stray output capacitance. The small signal circuit diagram of this amplifier is shown on the right of the figure.

Combining R_L and C_{out} in a single impedance Z_L , and noting the direction of the collector current, it is not difficult to show that $v_{out} = -Z_L i_c = -Z_L g_m v_{in}$. Hence, the amplification coefficient of the amplifier is:

$$A(\omega) = -\frac{g_m R_L}{1 + j\omega R_L C_{out}} = -\frac{g_m R_L}{1 + j\omega/\omega_0}$$

where $\omega_0 = 1/(R_L C_{out})$, which is called cut-off (radial) frequency.

From this equation it is not difficult to show that the small frequency gain is just $\approx -g_m R_L$, a constant value. By contrast, the high frequency gain is frequency dependent and equal to $\approx -(g_m R_L \omega_0)/(j\omega)$. Then the product $|A|\omega$ in the high frequency regime is constant and equal to $g_m R_L \omega_0 = g_m / C_{out}$, which is called the Gain Bandwidth (GBW) product [14].

When the transistor is used in the CSA configuration with a feedback capacitor, C_f , as explained in section 6.2, charge division can be described via the introduction of the dynamic input capacitance, C_{in} . In the frequency domain, this capacitance will have a complex impedance $1/j\omega C_{in}$. Because the amplifier response is now frequency dependent, so too is C_{in} , according to $(1 - A(\omega))C_f$ (the amplifier is inverting). Assuming $|A(\omega)| \gg 1$, $C_{in} \approx -A(\omega)C_f$, and its high frequency complex impedance becomes $Z_{in} = C_{out}/(C_f g_m) = R_{in}$,

which is purely resistive. Since R_{in} is in series with the detector capacitance, C_d , the resulting circuit is effectively an RC circuit with time constant $\tau = R_{in}C_d = (C_{out}C_d)/(C_f g_m)$. Substituting for the GBW, the final expression is $\tau = C_d/(C_f GBW)$ [16].

Therefore, so long the frequencies of interest are much larger than the cut-off frequency is, which is typically the case for a particle detector, the CSA in series with the strip capacitance can be modelled as a RC circuit.

The impulse response $h(t)$ that was given in section 6.2 is actually the Green function solution of the first order ordinary differential equation:

$$\tau \frac{dy}{dt} + y = x(t)$$

where τ is the time constant of the RC circuit, and $x(t)$ is the input, that is the electrical signal induced on the strip [25].

9.6 F: Why 26keV?

The energy loss of a minimum ionising particle in silicon is $dE/dx = 1.66 \text{ MeVcm}^2\text{g}^{-1}$ [24], where the energy is normalised by the density of the material. Given the density of silicon, 2.3290 gcm^{-3} , this amounts to an energy loss per unit length of $71.275 \text{ eV}\mu\text{m}^{-1}$.

Assuming that a particle travels right past through the detector, that is approximately $300\mu\text{m}$, the total energy deposited in the detector material through ionisation is about 21.4keV . Since the energy needed to excite an electron-hole pair is 3.736eV [14], this corresponds to around 5700 liberated charge carriers. It was decided to settle for 7000 in order to have a slightly larger signal at the detector.

9.7 G: Motion and Signal in a damaged detector

9.7.1 Undepleted detector

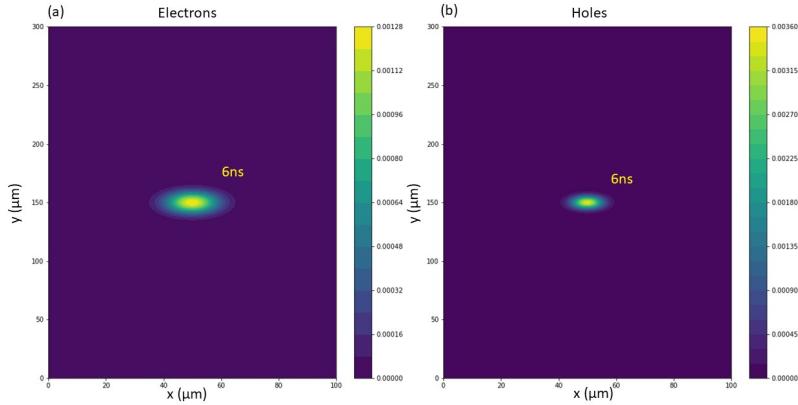


Figure 57: Diffusive spreading in undepleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 150\mu\text{m}$. (a) Electrons, (b) holes.

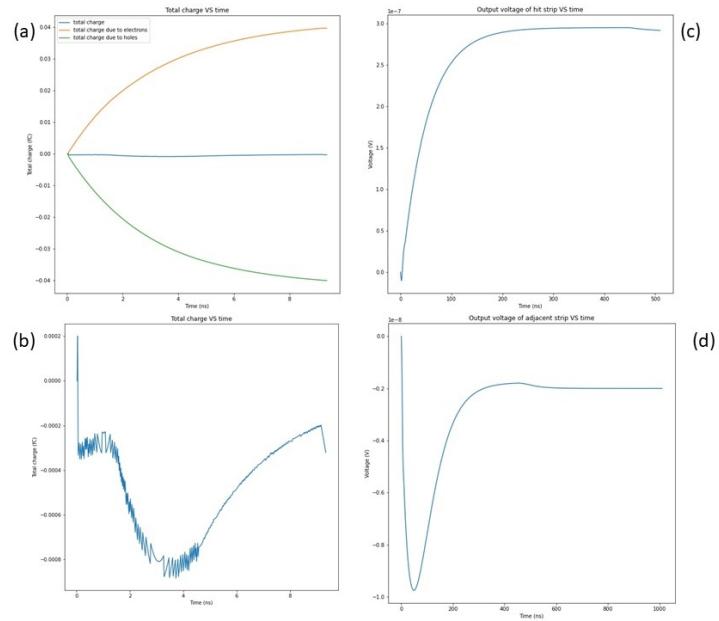


Figure 58: Signal on hit strip: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

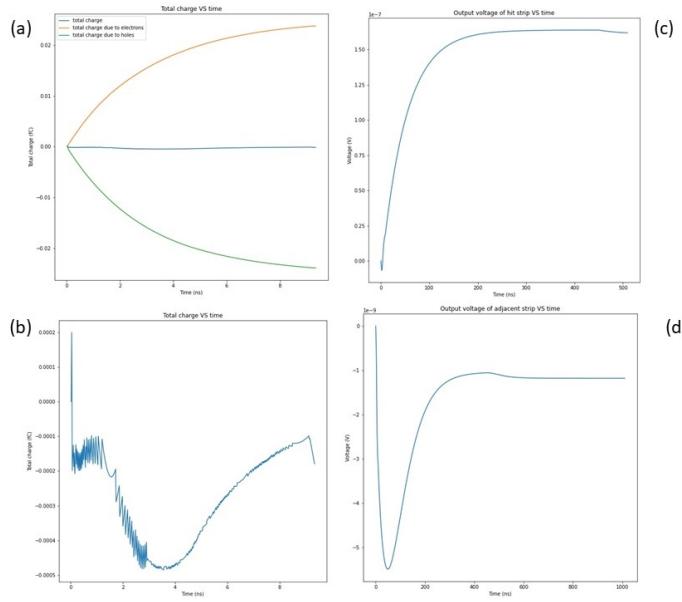


Figure 59: Signal on neighbouring strips: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

9.7.2 Partially depleted detector

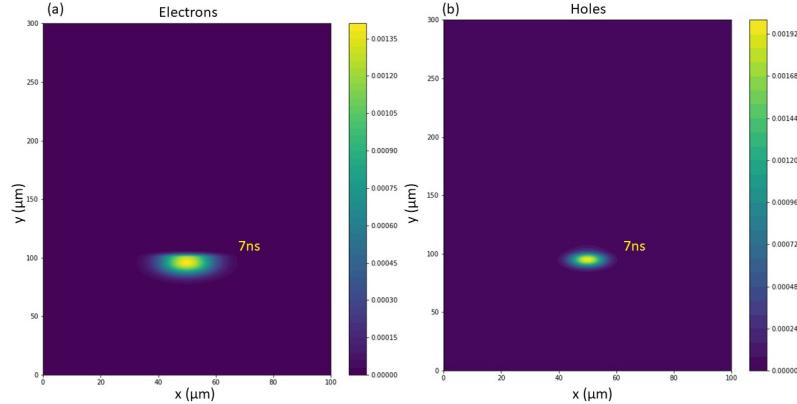


Figure 60: Charge motion in partially depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 95\mu\text{m}$. (a) Electrons, (b) holes.

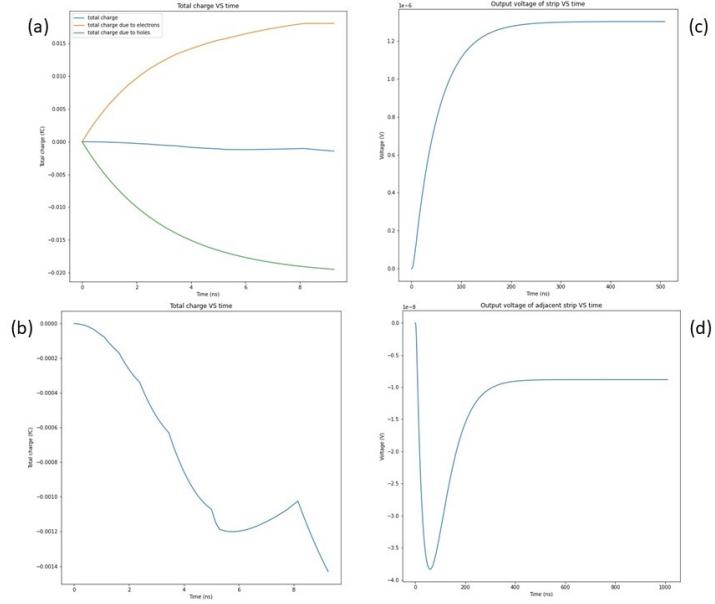


Figure 61: Initial condition: $x = 50\mu\text{m}$, $y = 95\mu\text{m}$. Signal on hit strip: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

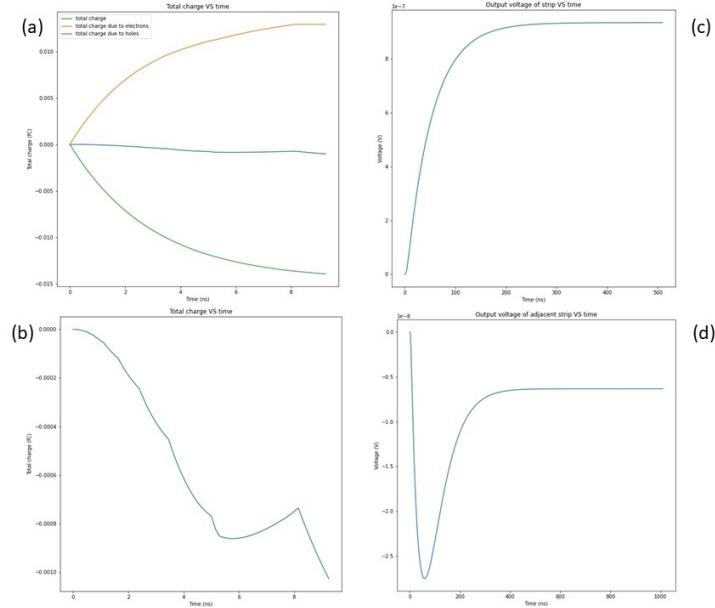


Figure 62: Initial condition: $x = 50\mu\text{m}$, $y = 95\mu\text{m}$. Signal on neighbouring strips: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

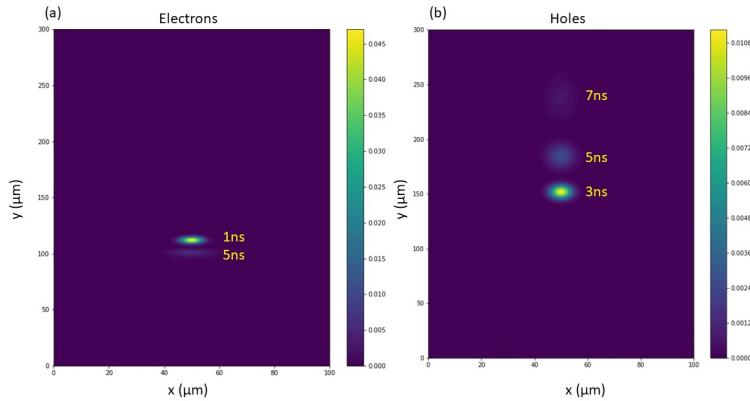


Figure 63: Charge motion in partially depleted detector. Initial condition: $x = 50\mu\text{m}$, $y = 125\mu\text{m}$. (a) Electrons, (b) holes.

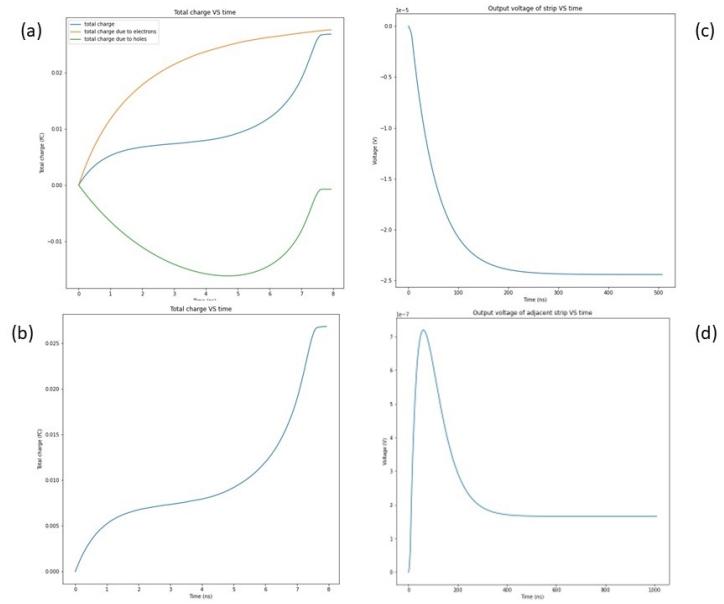


Figure 64: Initial condition: $x = 50\mu\text{m}$, $y = 125\mu\text{m}$. Signal on hit strip: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

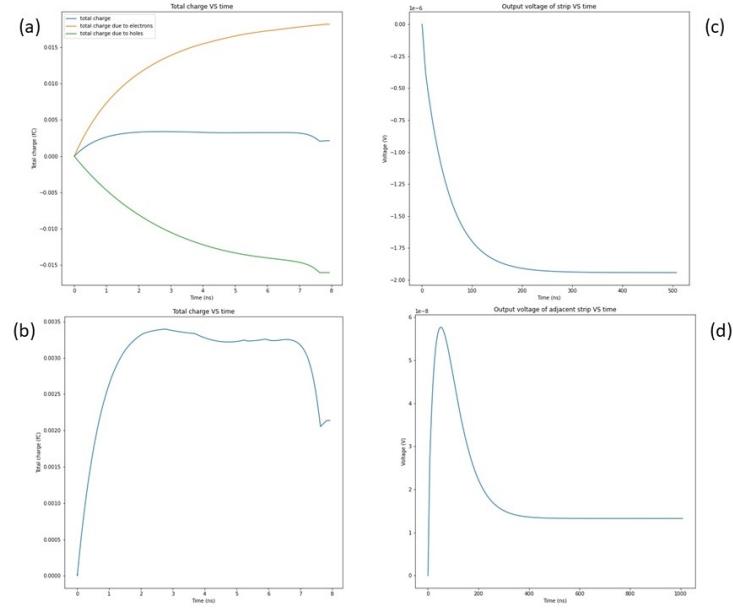


Figure 65: Initial condition: $x = 50\mu\text{m}$, $y = 125\mu\text{m}$. Signal on neighbouring strips: (a) different signal charge contribution, (b) total charge, (c) output voltage, (d) cross-talk voltage.

9.8 H: Potential at Partial Convergence

The following figure shows the bias-point potential that was obtained at partial convergence ($R \approx 100$) using a GMG method that implemented the “more realistic” boundary conditions discussed in section 4.3.

As it can be seen the maxima and minima of the potential match the ones obtained in the approximate calculation.

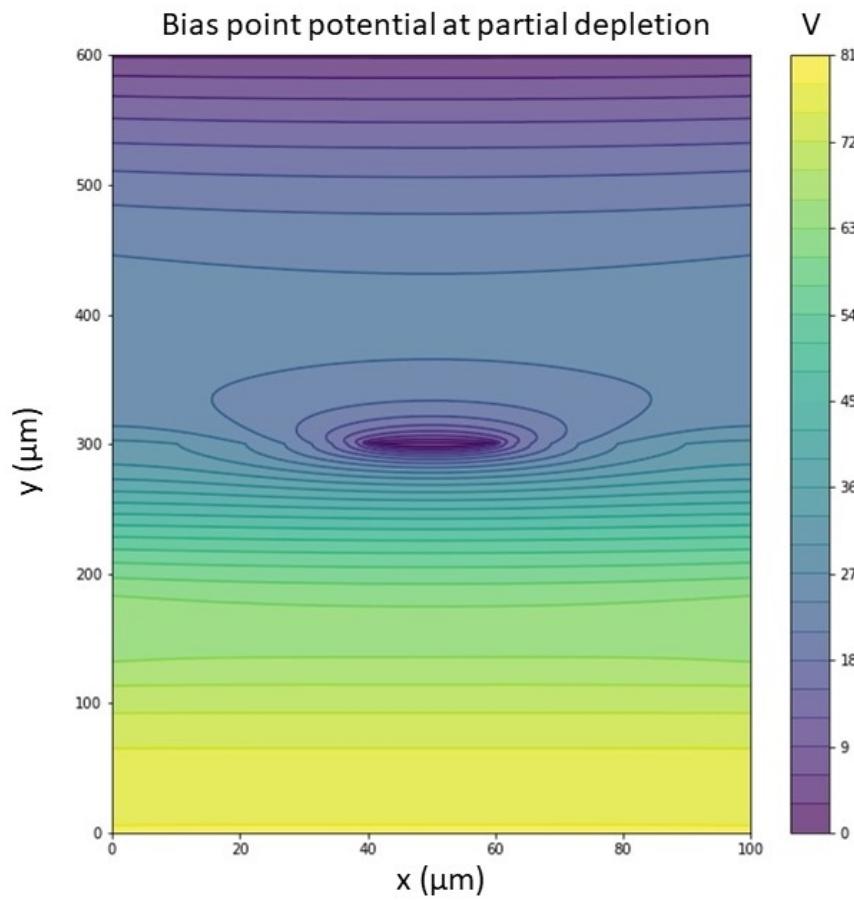


Figure 66: Bias-point potential in microstrip detector at partial convergence.

References

- [1] An Introduction to Computational Fluid Dynamics: The Finite Volume Method, H. Versteeg, W. Malasekera, Pearson Prentice Hall, 2nd edition, 2007.
- [2] Computational Fluid Dynamics, Patrick J. Roache, Hermosa Publishers, 1985.
- [3] A Stable and Accurate Convective Modelling Procedure Based on Quadratic Upstream Interpolation, B. P. Leonard, Computer Methods in Applied Mechanics and Engineering, 1979.
- [4] A Consistently Formulated QUICK Scheme for Fast and Stable Convergence Using Finite-Volume Iterative Calculation Procedures, T. Hayase, J. A. C. Humphrey, R. Greif, Journal of Computational Physics, 1992.
- [5] Numerical Analysis of Convection-Diffusion Using a Modified Upwind Approach in the Finite Volume Method, A. Hussain, Z. Zheng, E. F. Anley, 2020.
- [6] Numerical Methods for Partial Differential Equations: Finite Difference and Finite Volume Methods, Sandip Mazumder, Elsevier, 2016.
- [7] Computational Physics: Simulation of Classical and Quantum Systems, Philipp O. J. Scherer, Springer, 2010.
- [8] Fourier Analysis of the SOR iteration, R. J. LeVeque, L. N. Trefethen, IMA Journal of Numerical Analysis (1988).
- [9] Convergence rates of iterative treatments of partial differential equations, S. Frankel, Math. Comp., 1950.
- [10] Modular Series on Solid State Devices, Volume I: Semiconductor Fundamentals, Robert F. Pierret, Addison-Wesley Publishing Company, 1983.
- [11] Modular Series on Solid State Devices, Volume II: The PN Junction Diode, Gerold W. Neudeck, Robert F. Pierret, Addison-Wesley Publishing Company, 2nd edition, 1989.
- [12] Introducing a 1D numerical model for the simulation of PN junctions of varying spectral material properties and operating conditions, P. Stamatopoulos, M. Zeneli, A. Nikolopoulos, A. Bellucci, D. M. Trucchi, N. Nikolopoulos, Energy Conversion and Management, 2021.
- [13] Semiconductor Devices: Physics and Technology, S. M. Sze, M. K. Lee, John Wiley & Sons, 3rd edition, 2012.
- [14] Semiconductor Detector Systems, Helmuth Spieler, Oxford University Press, 2005.

- [15] Particle Physics in the LHC Era, G. Barr, R. Devenish, R. Walczak, T. Weidberg, Oxford University Press, 2016 .
- [16] Particle Detectors: Fundamentals and Applications, Hermann Kolanoski, Norbert Wermes, Oxford University Press, 2020.
- [17] Simulation study of signal formation in position sensitive planar p-on-n silicon detectors after short range charge injection, T. Peltola, V. Eremin, E. Verbitskaya, J. Harkonen, Journal of Instrumentation, 2017.
- [18] Induced Charge on Capacitor Plates, C. Y. Fong, C. Kittel, American Journal of Physics (1967).
- [19] Low-noise techniques in detectors, Veljko Radeka, Ann. Rev. Nucl. Part. Sci., 1988.
- [20] Detector capacitance of silicon microstrip detector. H. Ikeda, M. Takana, S. Okuno, Nuclear Instruments and Methods in Physics Research, 1991.
- [21] Capacitances in silicon microstrip detectors, E. Barberis, N. Casiglia, C. LeVier, J. Rahn, P. Rinaldi, H. F. W. Sadrozinski, R. Wichmann, T. Ohsugu, Y. Unno, H. Miyata, N. Tamura, K. Yamamoto, Nuclear Instruments and Methods in Physics Research, 1994.
- [22] Capacitance measurements in silicon microstrip detectors, S. Masciocchi, A. Peisert, C. Ronqvist, D. Vitè, R. Wheaton, IEEE Transactions on Nuclear Science, 1993.
- [23] High-Luminosity LHC, https://home.cern/resources/faqs?text=high-luminosity&field_p_faq_display_topic_target_id>All&field_p_faq_display_audience_target_id>All, visited 15/03/2022.
- [24] Passage of Particles Through Matter, S. Eidelman et al., Physics Letters, B592, 1 (2004).
- [25] Signal and Systems, A. V. Oppenheim, A. S. Willsky, Prentice-Hall signal processing series, 2nd edition, 1997.